VIVEKANANDA GLOBAL UNIVERSITY, JAIPUR

(Established by Act 11/2012 of Rajasthan Govt. Covered u/s 2(f) of UGC Act, 1956)

Bachelor of Computer Application

(BCA)

Basic Electronics

SEMESTER I

Author: Ms. Jyoti Shekhawat

Approval Director CIQA: 27th July 2023

Approval of Academic Council: 28th July 2023

For Vivekananda Global University

Registrar

Table of Content

Learning Objectives

After studying this unit the student will be able to:

1 Explain the concept of Electricity
2 Describe the phenomena of EMF and Potential Difference
3 Discuss a Circuit and its components
4 Explain Faraday's laws and their applications

Introduction

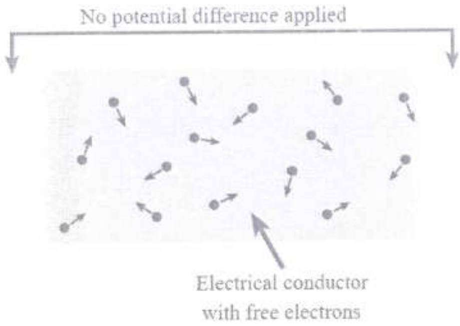
Every circuit has some basic elements like conductive source, voltage and some load. The capacitor can be useful for the voltage which allows the current to flow within the circuit.

1.1 Electric Current – The Basics

The Science of Electricity is vast and diverse. An Electric current is the very most basic concept that is at its core.

The general molecular structure of any material has electrons revolving around. The basic premise of a current is known by 'the movement of electrons within a given material'. Electrons are negatively charged particles. When the electrons move, along with them there is an amount of charge that moves and this is called Electric current.

However, in general, the motion of electrons within a material is random and haphazard. Therefore, as a result there is no movement of charge. And there is a need for an external stimulation to make these electrons move in a specific direction to create a movement of charge and hence a current.

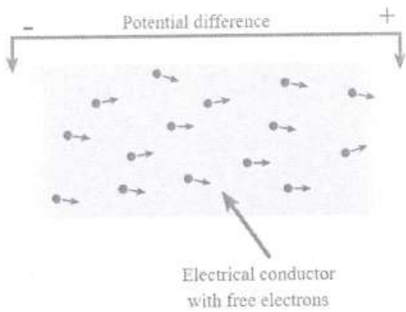


Random electron movement in a conductor with free electrons

Fig 1.1 Electron movement

A negatively charged particle will get attracted towards a positive charge and as such if that much force acts on electrons to move them towards the positive charge, then they all move in that direction. The external force that's applied is called Electromotive force (EMF) and its quantity, 'voltage' is measured in Volts. Fig 1.1

An analogy of a water pipe will help in understanding this better. If water is placed in a pipe and air pressure is applied from one side of the pipe, the movement of the water (speed and quantity) towards the other side of the pipe depends on the quantum of pressure applied. Same goes with current.



Electron flow under the action of applied electro-motive force

Fig 1.2 Flow of electron

Some materials allow free and faster flow of electrons (Conductors) while some resist (Resistors) and some completely block (Insulators).

Flow of Current Vs Flow of Electrons

The particles that carry charge along conductors are free electrons. And they move from towards a positive charge.

By convention / definition the direction of current is a direction in the circuit towards which the positive test charges are pushed, ie., negatively charged electrons move in the direction opposite to the electric field. It is illustrated in the diagram. Fig 1.3.

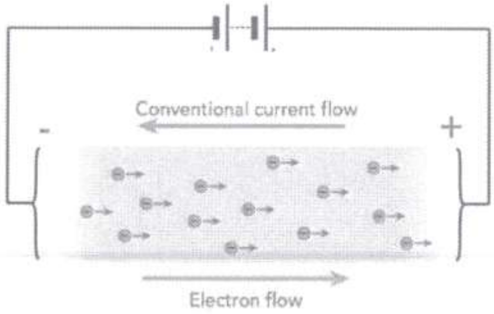


Fig 1.3 Particles carry charge

Current is basically defined as "rate of flow of charge".

Mathematically it is expressed as

$$I=Qt$$

where Q is the charge and t is the time. The SI unit of charge is 'coulomb' and that of time is 'second'.

Just to give an estimate on the flow of Electrons, it takes 6.24 billion billion electrons per second for one ampere of current.

Current in a circuit is normally designated by the letter 'I', and this letter is used in equations like Ohm's law,

$$V= I*R$$

Where V=Voltage Applied and R= Resistance, I = Current

Quantifying the amount of current flowing in a circuit is very important to design the circuit and operate it. The SI unit of current is coulomb per second, generally known as ampere, denoted with an abbreviation 'A'. Example: A current of 10 Amps is noted as 10A.

A multimeter can be used for measuring Current in a circuit.

The following table gives as indicative consumption of current in different household gadgets:

Device	Description	Typical Current Needed
Laptop	While being charged	0.25-0.35A
Home Printer	While printing	0.45-0.5 A
Television	while switched on, a 50" LED TV	0.35 A
Heating Kettle	2.0 KW Heating Kettle at 240 Volts	10 A

Table 1.1

Effects of current

Current flowing through a circuit shows a number of signs:

- Dissipation of Heat: Current passing through a circuit generates heat. This heat is dependent on the Current, Voltage and resistance of the conductor. If the current is small, the heat generated could be very small. If its larger heating may occur. This is put to a number of uses like home heating, boilers, iron box etc.
- Magnetic effect: Another effect of the current is the build-up of a magnetic field around the conductor. Placing a compass close to the conductor carrying a current indicates the magnetic field built. The magnetic field can be put to a number of uses, Electro-magnets, relays, loudspeakers, trains using levitation etc.

1.2 ElectroMotive Force (EMF) and Potential Difference (Voltage)

Electromotive force is defined "as the characteristic of any energy source capable of driving electric charge around a circuit - it is the force within a voltage source that drives the current around a circuit". The electromotive force is the very basic electrical force that drives the current around the circuits.

There is a possibility of some confusion between the phrases - Electromotive force (EMF) and the Voltage or potential difference (PD). Both are measured in volts, but the two are very different in what they are.

The volt is the standard unit of the measurement of potential difference and EMF. The voltage at a point, usually, refers to the potential difference between that point and the ground or zero volt line of that circuit.

Potential difference

As a definition "The potential difference between two points in an electrical or electronic circuit represents the work involved or the energy released in the transfer of a unit quantity of electricity from one point to the other".

When a charge moves from a point of higher potential to a point of lower potential energy is released and often this is in the form of heat. Example: Resistors. Potential is the voltage at a given point in a circuit, not the source of the force to move it around the circuit.

Hence the key difference between EMF and Potential Difference is that - EMF is the cause, i.e. it is the driving force, while Potential difference is the result of that EMF.

The key differences between the two are given below:

ELECTROMOTIVE FORCE (EMF)	POTENTIAL DIFFERENCE (PD)
It is the driving electric force from a cell or generator.	Potential difference output came from the current passing through a resistance within a circuit.
It is the cause or reason.	Potential difference is the effect or final output
The EMF is present even when their is no current drawn through the battery.	Potential difference across the conductor is zero when there is no current current.
Volt is the unit of EMF	The unit of potential difference is the volt.
EMF remains constant.	Potential difference does not remain constant - it totally depends upon the circuit conditions.
Its symbol is E.	Its symbol is V.

It does not depend on circuit resistance.	It depends on the resistance between two points of measurement.

Table 1.2

1.3 Electric power

The standard definition of power is the rate of doing work. Electric power is defined as the rate of transfer of electrical energy per unit time.

From the definition it can be seen that:

$$W = VQ / t$$

But as: $Q/t = I$

Substituting: $W = V I$

Where:

- W = power in watts
- V = potential in volts
- I = current in amps
- Q = charge in coulombs
- t = time in seconds

Definition of watt

Watt is the SI unit of power and it defines the rate of energy conversion.

1 Watt = 1 joule / 1 second.

One watt is the rate of the measurement at which work is done when a current of one ampere (I) flows through a circuit with an electrical potential difference of one volt (V).

$$W = V I$$

Some examples of the typical power needs are given in the table below:

DEVICE	DETAILS
Desktop computer	normally less than 100W

Kettle	Typical 2.0 kW
42 inch LED Television	~100W
Domestic incandescent light bulb	Up to 150 W

Table 1.3

Calculation of Power

The amount of power dissipated in a circuit can be determined as the product of the potential difference or voltage across the particular element, and the current flowing through it. In other words an electrical kettle running from a 240 volt supply, and consuming 10 amps of current will dissipate 240 x 5 = 1200 watts.

1.4 Ohm's Law

Ohm's Law is a fundamental law governing electrical and electronic circuits. It relates resistance, current and voltage in a mathematical expression.

Ohm's Law

Ohm's Law states that "The flowing of current in a circuit is directly proportional to applied potential difference and inversely proportional to the resistance in a circuit "

Ohm's Law is used across all branches of electrical and electronic sciences, Infact anywhere Electric current flows. It can be used for calculating the resistance needed in circuits, calculating voltages across a resistor, and so on.

Ohm's law can be expressed in mathematical form as:

$$V=IR$$

Where:

V = voltage expressed in Volts

I = current expressed in Amps

R = resistance expressed in Ohms

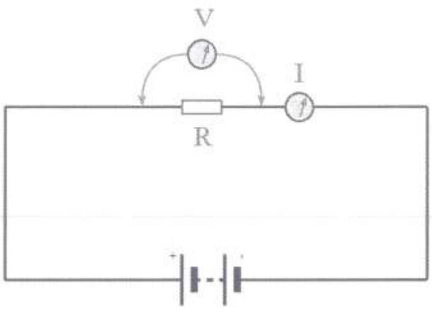


Fig 1.4 Circuit utilises Ohm's law

The formula can be manipulated to calculate any of the three, V, I and R, if two of them are known:

I=VR or

R=VI

Example: If a voltage of 6 volts is placed across a 140 Ohm resistor determine the amount of current in the circuit.

$$I = V/R = 6/140 = 0.0429\ A$$

Example: If a circuit has to maintain a voltage of 230 volts, and consuming a of current 15 A what is the resistance of the circuit?

$$R = V/I = 230/15 = 15.33\ Ohm$$

Ohm's Law is the most basic concept in electrical engineering. Design of most electrical or electronic circuits and components utilises Ohm's law.

1.5 Basic Circuit Components

An Electric Circuit is defined as the closest path for the transmitting of an electric current through a medium of electrical and magnetic field".

An electric circuit may consist of an energy source (such as a battery), conducting wires like wires made of copper or aluminium as conducting medium and a load (such as a LED light bulb). The battery provides required energy for flow of electrons, to the light bulb.

The flow of electrons through the path induced by an EMF created by an external source creates the electric current. The flowing electrons leave the path through a load, to the ground-earth, completing a closed path. Here the load refers to any electrical gadget like a laptop, heater, light bulb etc,

1.5.1 Basic Circuit Elements

As seen in the example above, a circuit is made of interconnected elements. Such elements can be either active or passive elements, depending on their ability to generate energy.

- **Active Circuit Elements**

Elements can generate energy like batteries, generators. In a typical circuit, significant active elements include the source elements. The source elements provide energy and are further classified as Independent or Dependent sources. A battery source is seen as an Independent source as it provides a constant voltage to the circuit, irrespective of the current flowing. On the other hand a transistor provides current depending on the voltage applied.

- **Passive circuit Elements**

Passive Elements can control the flow of electrons through them. They can either increase or decrease the voltage.

Resistor is a good example of a passive element. A resistor "opposes or resists the flow of current through it". When Ohm's law is applied, we see that for a constant Resistance, the current supplied can be changed by varying the voltage and vice versa.

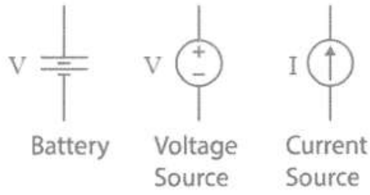


Fig 1.5 Passive circuit elements

An Inductor "stores energy in the form of the electromagnetic field". Here the voltage across is proportional to the rate of change of current flowing through it.

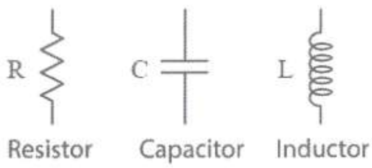


Fig 1.6 Passive circuit elements

A capacitor on the other hand "stores energy in the form of an electrostatic field". The voltage across is proportional to the charge available.

- **Circuit**

A circuit is an arrangement of physical components of power sources, and signal sources in such a way that current can flow in a complete circle. On paper a circuit is drawn as a Schematic, showing the circuit elements in symbols and their connections in lines

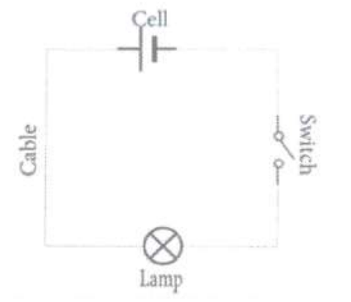


Fig 1.7 Symbols and connections

i) Closed circuit - A circuit is closed if the loop is complete, if all currents have a path back to where they came from.

ii) Open circuit - An Open circuit refers to one where the loop is not complete, if there is a gap in the path.

iii) Make or Break - Closing the path of the current makes a circuit complete while breaking the path is the opposite. Example: An electric switch

The circuit elements are connected and the connections have the following notations:

- Node - A junction where two *or more* elements connect is called a *node*. Generally marked as a black dot. (Fig 1.8)

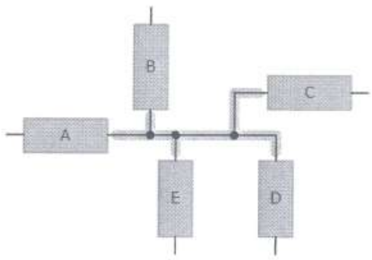


Fig 1.8 Node

- Branch - Connections in between nodes are referred to as *Branches*. Generally a branch is an element (resistor, capacitor, source, etc.) and there are as many branches in a circuit as there are elements. Example: There are six branches in the below example (Fig 1.9).

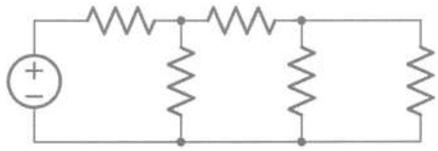


Fig 1.9 Branches

- Loop – A *loop* is closed path going through circuit elements. A loop can pass through a node only once.

- Mesh – A loop that has no other loops inside it is referred to as a Mesh.

1.5.1.1 Reference Node – When a circuit is analysed, any one of the nodes acts as a reference point for the analysis. Parameters across all other nodes are checked with respect to this reference point or Reference Node. Voltages at other nodes are measured relative to the reference node.

Generally two common choices as reference nodes are, either –

- The Negative terminal of the voltage or current source powering the circuit or the node connected to the greatest number of branches.
  OR
- Ground – The concept of ground has three important meanings. Ground is
  - The reference point from which voltages can be measured.
  - The return path back to the source for electric current
  - A direct physical connection to the Earth for safety reasons

1.6 Electromagnetic Induction

In 1831, Michael Faraday published on electromagnetic induction. He was famous for his experiments with magnetic fields and creating a flow of current.

Electromagnetic Induction is defined as “A current produced because of voltage production (electromotive force) due to a changing magnetic field”. This happens with relative movement of a conductor and a magnetic field - either the magnetic force moving with the conductor staying stationary or the conductor moving within the stationary magnetic field.

1.6.1 Faraday’s law of Electromagnetic Induction

First law: “EMF induces only when a conductor is placed in a varying magnetic field and this emf is called an induced emf and if the conductor is a closed induction

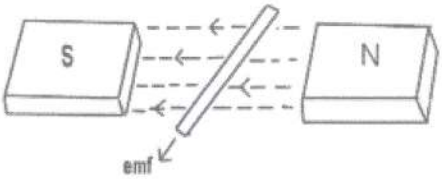


Fig 2.0 Electromagnetic

(Source: Electricaleasy)

circuit then the induced current flows through it”.

1.6.2 Fleming's Right hand thumb rule

It states that “if we arrange our thumb, forefinger and middle finger of the right-hand perpendicular to each other, then the thumb points towards the direction of the motion of the conductor relative to the magnetic field, the forefinger points towards the direction of the magnetic field and the middle finger points towards the direction of the induced current.”

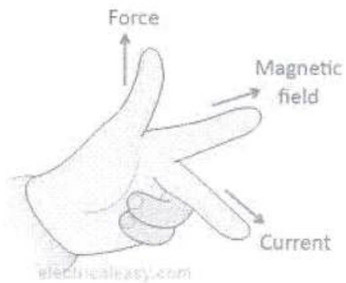


Fig 2.1 Fleming's Right hand thumb rule
(source: Electriceasy)

Faraday’s Second law: “The induced EMF have the magnitude which contain equal rate of change of flux linkages. “

According to Faraday's law the voltage induced within a coil is proportional to the number of turns and the changing magnetic field of the coil.

So now, the induced voltage is as follows:

$$e = N \times d\Phi dt$$

where,

e = induced voltage

N = number of turns in the coil

Φ = magnetic flux

t = time

1.6.3 Lenz's law of Electromagnetic Induction:

It states that, "an induced electric current flows in a direction such that it opposes the change that induced it".

According to Lenz's law

$$E = -N\,(d\Phi/\,dt) \quad \text{(volts)}$$

Eddy currents

When a conductor is placed in rapidly changing magnetic field, the induced current in the conductor is termed as Eddy current. Going by

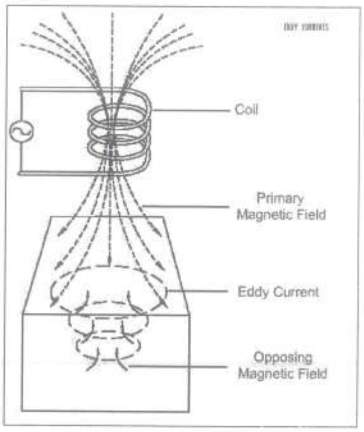


Fig 2.2 Eddy current

(Source: Geocities)

Lenz's law of electromagnetic induction, the current moves in such a way as to create a magnetic field opposing the change. Generally, the loss of each useful energy isn't desirable, but there are some practical applications. Like braking, Induction furnace etc.

□

1.7 Summary

- Electronics involves working with electricity to accomplish a particular task. Essentially it is about manipulation of charges, currents, and voltages
- Ohm's law connects the three important aspects of electricity – Voltage, Current and Resistance.
- Electromotive force is the driving force for the electrons to move
- Depending on the ease of movement of electrons, a material is classified as a conductor, resistor or insulator
- Electromagnetic induction as described by Faraday creates a current in a conductor when it moves within a magnetic field or when a magnetic field moves around a stationary conductor. This forms the fundamental principle of many electric applications.
- Circuit design is all about utilization of electricity for a specific task and involves closing the path of current with the necessary components like resistors, inductors, capacitors, load and source of energy.

1.8 Keywords

- *Electricity:* Can be defined as a Flow of charge in a direction
- *Voltage or Potential difference:* Potential difference is the work done in moving a unit of positive electric charge from one point to another.
- *Resistance:* Resistance is a measure of the opposition to current flow in an electrical circuit. Resistance is measured in Ohm
- *Electromotive Force:* a difference in potential that tends to give rise to an electric current.
- *Electromagnetic induction:* Electromagnetic Induction is the Current produced in a conductor because of creation of an electromotive force due to a changing magnetic field around the conductor
- Circuit: A path for current designed to achieve a specific task using current

1.9 Review Questions

a. What is Electric current?

b. Discuss the features of Electric current using Ohm's law

c. Describe how EMF is related to Current

d. What factors influence selection of Circuit components ?

e. Discuss the two laws of Faraday on Electromagnetic induction

f. What is a resistor ?

g. What is the resistance required in a circuit to maintain a voltage of 230 V with the current in the circuit being 5 A?

h. What are the uses of Electromagnetic induction?

i. What are the different types of Circuits?

2.0 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education
4. Millman, Electronics Devices and Circuits, TMH
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India

Online References

7. https://www.electronics-notes.com/articles/basic_concepts/voltage/electromotive-force-emf-potenti
8. https://www.nagwa.com/en/videos/614129427189/
9. https://india.oup.com/productPage/5591038/7421214/9780199479368
10. https://www.vedantu.com/question-answer/define-coefficient-of-mutual-induction-if-in-the-class-12
11. https://collegedunia.com/exams/induced-voltage-formula-physics-articleid-7291

Learning Objectives

After studying this unit 2 the student will be able to:

1 Describe Kirchoff's laws and its applications
2 Apply the concept of serial and parallel circuits and solve problems
3 Discuss a resistive, inductive and capacitive ciruits
4 Explain Lenz's law

Introduction

Every circuit has some basic elements like conductive source, voltage and some load. The capacitor can be useful for the voltage which allows the current to flow within the circuit.

2.1 Kirchhoff's Laws

Kirchoff formulated two laws – one known as Kirchoff's Current Law (KCL), and the second, known as Kirchoff's voltage law.

Kirchoff's Current Law

This states that "The current flowing into a node, i.e. a junction in a circuit is equal to the current flowing out of it".

Kirchoff's voltage law

This law states that "In any complete loop within a circuit, the sum of all voltages from components that supply electrical energy, i.e. cells or generators, will equal the sum of all voltage drops across the other components in the same loop".

The current law is important as it can solve many circuit analysis problems easily while looking at nodes within a circuit. The basis of Kirchoff's current law is that it uses the concept of the conservation of charge.

Accordingly the algebraic sum of all current flowing into and out of a junction must add up to zero.

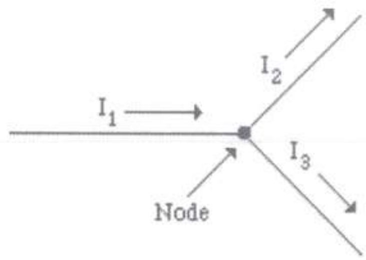


Fig 1.0 – An illustration of Kirchoff's First Law

According to the KCL, for the circuit in figure 1,

we have $-I_1 + I_2 + I_3 = 0$

or

$I_1 = I_2 + I_3$

For the situation in figure 2, we have $+V_1 - V_2 - V_3 = 0$ or $V_1 = V_2 + V_3$

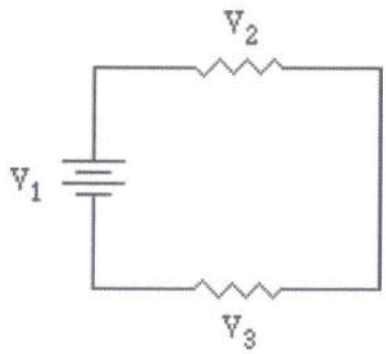


Fig1.1 An Illustration of Kirchoff's Second Law

Basic rules for Applying Kirchoff's Laws:

- It is required to maintain either a Clockwise or Counter-clockwise direction through the circuit, Else voltage will not be zero
- Series Circuits can also be analysed using Kirchoff's voltage Law

- Kirchhoff's Laws can be used to assess AC circuits by altering them to account for electromotive forces, resistors, capacitors, and other components.
- These guidelines are only relevant for circuits that cannot be reduced by mixing elements in series and parallel.
- Kirchhoff's laws have one flaw of assuming that the closed-loop magnetic field is constant.

Voltage Current relationship over changing resistance

Ohm's Law is a simple tool which helps us calculate the parameters for circuits. But it has its limitations.

Resistance is rather stable in most conductors and is unaffected by voltage or current.

However when there is a change in resistance of the conductor due to a rise in temperature due to the resistance/friction, Ohm's law does not represent the reality. In such a scenario, as it progresses to the right we see the line flattening out, the circuit requiring greater and greater increases in voltage to achieve equal increases in current.

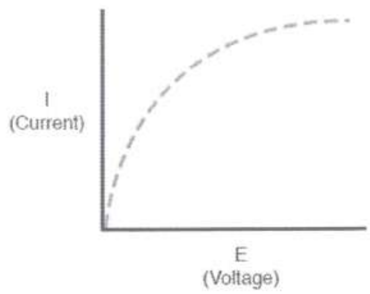


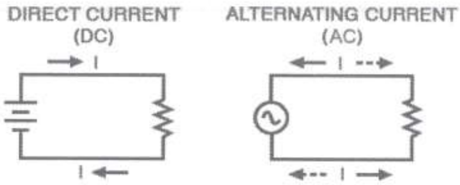
Fig-1.2 Voltage-Current Non linearity

While this change in resistance due to temperature is true for most conductors, the quantum of change could be negligible; however in some conductors like metal filaments, the change is quite significant.

Alternating Current vs. Direct Current

Certain sources of electricity (like rotating electro-mechanical power generators) naturally produce voltages alternating in polarity, reversing polarity over time (many times a second!), Either as a voltage switching polarity or as a current witching direction back and forth, this "kind" of electricity is known as Alternating Current (AC).

DC Voltage sources use the standard and familiar battery symbol. AC voltage sources use a circle with a sine wave with in it.



Direct vs alternating current

Fig 1.3 – AC and DC Symbols

2.2 Network Sources

Electrical circuits in practice are never simple single loop paths but can have multiple branches, multiple components and energy sources depending on the usage needs. The following types of electrical networks are generally recognised:

i) Linear Network

A circuit which has its elements like resistances, inductances and capacitances constant always with respect to time, voltage, temperature etc. is known as a linear network. The Ohm's law can be applied to such networks. The principle of superposition can be applied to obtain the mathematical equations for such networks.

ii) Non-linear Network

A circuit whose parameters change their values with change in time, temperature, voltage etc. is known as a non-linear network. The Ohm's law may not be applied to such network.

iii) Bilateral Network :

A circuit which exhibits same behaviour irrespective of the direction of current is called bilateral network. Example, A Network consisting only of resistances.

iv) Unilateral Network:

When the direction of current changes the behaviour of a circuit, it is called a unilateral network. Example, a Circuit consisting of diodes.

v) Active Network :

When a circuit has a source of energy in it, it is called an active network. The source could be a voltage or current.

vi) Passive Network:

A circuit without a source of energy in it is called passive. Depending on the type of voltage applied, it could be Direct current or Alternating current. In the ac circuits voltages vary periodically and as such current also varies. In the direct current circuits (D.C.) it is a fixed voltage source with polarities +ve and – ve.

vii) Lumped Network :

When in a network all the network elements are physically separable, it is known as a lumped network. Most networks are into this category.

viii) Distributed Network :

When in a network the circuit elements like resistance, inductance etc. cannot be physically separable for analysis, it is called a distributed network. The best example of such a network is a power transmission line where resistance, inductance and capacitance are distributed all along its length and cannot be shown as a separate element, anywhere in the circuit.

2.3 Network Analysis

Network Analysis helps us to calculate different electrical parameters of circuit elements in any electrical network. The circuit elements may be a combination of resistors, capacitors, inductors, voltage sources, current sources etc. The parameters may include current, voltage, resistance, impedance, reactance, inductance, capacitance, frequency, electric power, electrical energy etc.

The circuit elements in a network may be connected and can be connected in series or in parallel. Often we see networks which are fairly complicated and may have to apply different methods to simplify the network for determining the parameters.

A simple circuit is shown in Fig 1 below:

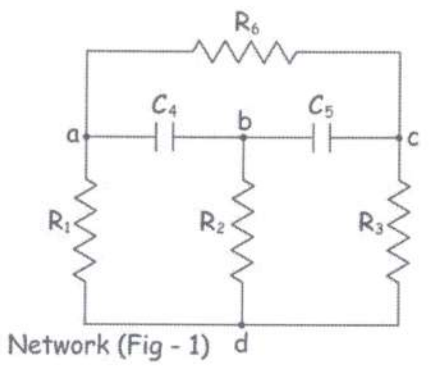


Fig 1.4: Circuit Diagram

(Source: electrical4u)

When all the voltages, as well as currents in different arms, in a circuit are determined, the circuit is considered to be analyzed.

There are two most common approaches which are used to solve a network:

Direct Method: This is restricted to solving simple circuits. Here there is no need for any changes to be made to the original circuit for analysis. Generally Kirchoff's law, Nodal analysis, Loop analysis, Superposition theorems etc., can do the job.

Network Reduction Method: This is used to solve simple as well as complex networks and here the circuit is converted into a much simpler form for rapid calculations of parameters.

2.3.1 Graph of an Electrical Network

When we sketch the circuit elements of an electrical network by hand, then the figure is known as the graph of the network. The figure – 2 below shows a simple graph of the above network in figure – 1.

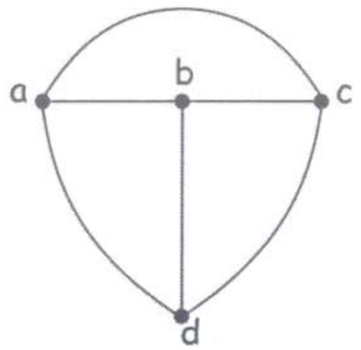


Fig 1.5: Graph of Network

Oriented Graph of the Network

Every line which represents a circuit element is called a branch of the network. The point where two or more branches meet is called a node. The direction of current is represented by an arrowhead. When we have a graph of the network with the direction of current (arbitrary) in each branch, the graph is referred to as an oriented graph of the network.

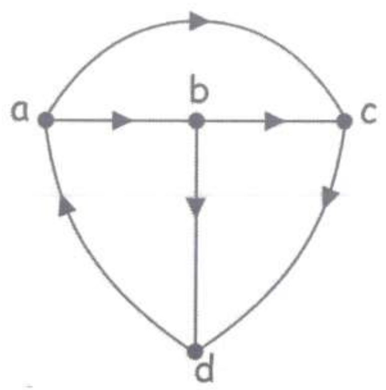


Fig 1.6 – Oriented Graph of Network in fig 1

An active network represented in a graph as a passive network by removing the voltage and current sources is referred to as an oriented topological graph of the network. A short circuit replaced the Voltage source and an Open circuit replaced the current source.

2.3.2 Terms used in Network Analysis

Branch: Each hand drawn line in a graph representing a path for current is called a branch.

Node: The end point of a branch where other branch/es meet is called a node.

Subgraph: A subset of branches of a graph. Is referred to as subgraph

Tree: A subgraph which contains all nodes of the graph, however, not forming a closed circuit is called a tree. If there are n number of nodes in the graph, the tree will have (n – 1) number of branches.

Cotree: A subgraph which contains all those branches which are not included in a tree is referred to as a Cotree. The cotree is the complement of a tree.

Equivalent Circuit: The first step of any network analysis is the simplification of a complex network to simplified form/s. Normally this is done by combining impedances in series and parallel. It may also be necessary to transform some or all of the voltage sources of the network to current source and vice versa.

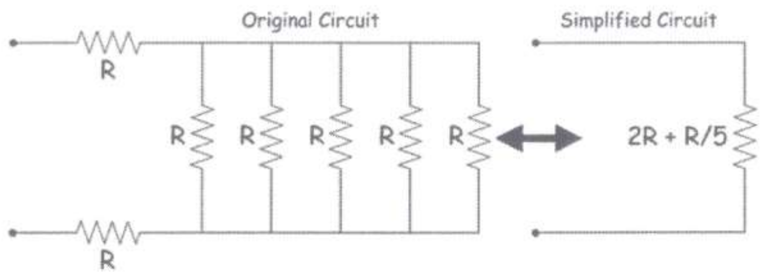


Fig 1.7 – Simplification of a circuit

2.4 Series and Parallel Circuit

As the name implies, when the circuit components (resistor, inductor, capacitor,..) are connected in series., they are referred to as Series circuits. Also, the current flows in each component connected to the circuit.

A series circuit is also called a 'voltage divider' due to its ability to divide or apportion the available voltage into fractions of a constant ratio for all of the components in the circuit.

When the components are connected in Parallel, such circuits are called Parallel circuits. Here the voltage for each component is the same, but the current varies. Current gets divided into each component in the circuit. Kirchoff's current law will be of help in calculating the current in each of the components.

If n number of resistances is connected in series, the value of equivalent resistance would be,

$$R_e = R_1 + R_2 + R_3 + \cdots + R_n$$

Let us say n numbers of resistances are connected in parallel, and then the value of resultant resistance will be,

$$R_e = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_n}\right)^{-1}$$

For Inductance and Impedance the same logic as that of resistances are applicable for both series and parallel connections.
However, for Capacitors it's the exact opposite.

When connected in series, n number of capacitances becomes equal to:

$$C_e = \left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_n}\right)^{-1}$$

And If n number of capacitances are connected in parallel, then it will be,

$$C_e = C_1 + C_2 + C_3 + \cdots + C_n$$

Example Problem (i): Equivalent Resistance, Current and Power in a series circuit

Battery with a terminal voltage of 12 V is connected to a circuit consisting of four 22 Ω and one 12 Ω resistors all in series. Assume the battery has negligible internal resistance.

- Calculate the equivalent resistance of the circuit.
- Calculate the current through each resistor.
- Calculate the potential drop across each resistor.
- Determine the total power dissipated by the resistors and the power supplied by the battery.

Solution

a. The resultant equivalent resistance is the algebraic sum of all the resistances:

$R_S = R1+R2+R3+R4+R5 = 22+22+22+22+12 = 100\ \Omega$.

b. By Kirchoff's current law, the current through the circuit is the same for each resistor in a series circuit. It can be calculated by dividing the applied voltage by the equivalent resistance:

$I = V/R_S = 12/100 = 0.12$ A

c. The power dissipated by a resistor is equal to $P=I^2R$ and the power supplied by the battery is equal to $P=I\epsilon$

$P1=P2=P3=P4=(0.12)^2 \times 22 = 0.3168$ W

$P5=(0.12)^2 X\ (12)=0.1728$ W

Dissipated power = 0.3168+0.3168+0.3168+0.3168+0.1728=1.44 W

$P_{source}=I\epsilon = 0.12 \times 12 = 1.44$ W

Example Problem (ii): Equivalent Resistance, Current and Power in a parallel circuit

Three resistors R1=2.00 Ω, R2=4.00 Ω, and R3=5.00 Ω, are connected in parallel. The parallel connection is attached to a V=12V voltage source.

a. What is the equivalent resistance?
b. Find the current supplied by the source to the parallel circuit.

Solution:

Generalizing to any number of N resistors, the equivalent resistance R_P of a parallel connection is related to the individual resistances by

$$R_P=(1/R_1+1/R_2+1/R_3+\ldots+1/R_{N-1}+1R/_N)^{-1}$$
$$= (1/R1 + 1/R2 + 1/R3)^{-1}$$
$$= (1/2+1/4+1/5)^{-1}$$
$$= 0.95\ \Omega$$

b. The total current can be found from Ohm's law, substituting R_P for the total resistance

$$I = 12/0.95 = 12.63 \text{ A}$$

2.5 Electrical Source Transformation

In a complex circuit, quite often it becomes necessary to convert current source to voltage source and vice versa, for simplifying the network.

When a voltage source is connected it delivers certain current to the circuit which depends on the impedance of the circuit and series internal resistance of the source. When connected across the same terminals, if the current source delivers the same current, the current source can be said to be equivalent to the voltage source.

When two terminals of a voltage source are shorted, the current flowing through the device is the current of the equivalent current source. Very similarly, in an open circuited current source, the voltage appearing across the open terminal of the source would be the voltage of the equivalent voltage source.

Current and voltage drop across the DC circuit elements

Voltage drop means the reduction in voltage or voltage loss across a particular element. Due to the presence of the impedance or passive elements, there will be some loss in voltage as the current moves through the circuit. That is, the energy supplied from the voltage source will get reduced as the current flows through the circuit.

Voltage drop across inductor:

For the case of inductor, the ohm's law takes the form:

$$v = L\,\frac{di}{dt}$$

Where,

v = Instantaneous voltage across the inductor

L = Inductance in Henrys

$\frac{di}{dt}$ = Instantaneous rate of current change (amps per second)

When we connect in series an inductor and a resistor connected to a power source, the inductor resists the change in current, and as such the build of current occurs slowly. The rate of change of current and the amount of impedance to the build-up of current are directly proportional. This is in line with Faraday's law and Lenz's law, That is, the quicker we want it to charge, the more it resists.

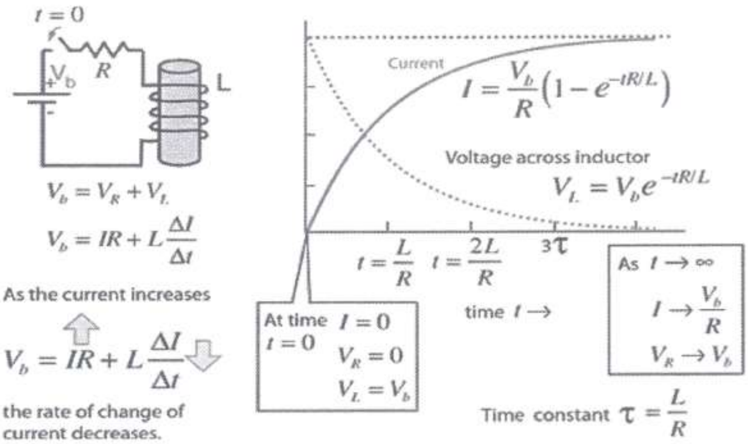


Figure 1.8: Voltage drop across Inductor (Source: Astro.uvic.ca)

Voltage across capacitor

When we have a capacitor in series under a constant voltage source, the following would be applicable:

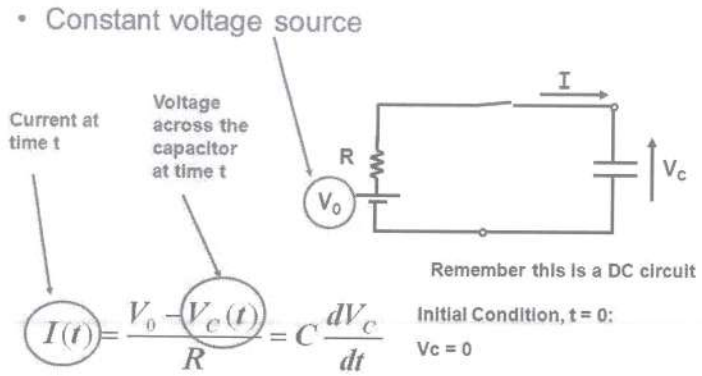


Figure 1.9: (Source: Astro.uvic.ca)

Let us connect a battery in series with a resistor and a capacitor. The capacitor has two plates and there is a high initial current as the battery transports charge with in the capacitor between the two plates. As the capacitor is charged to the level of the battery, the current approaches zero.

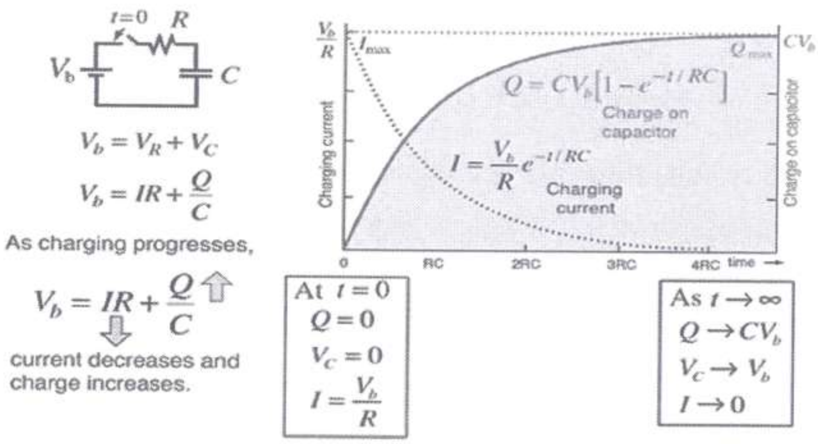


Figure 2.0: (Source: Astro.uvic.ca)

2.6 AC Circuits, with Resistor, Capacitor and Inductor

By its very definition Alternating Current or AC changes its polarity a number of times every second (frequency) due to the very way it is produced using magnetic fields. The frequency is about 50 cycles per sec in India. AC is used for high- power applications, Due to its advantages AC is more predominant in usage compared to DC. The oscillating shape of an AC supply takes the mathematical form of a “sine wave” or “sinusoidal wave”.

Therefore, a sinusoidal voltage can be defined as

$$V(t) = V_{max} \sin \omega t.$$

AC Circuit with a Resistor

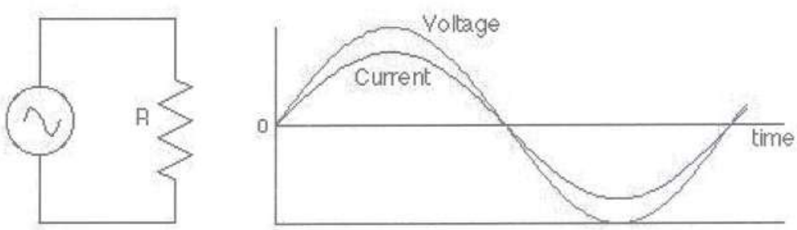


Figure 2.1 AC Circuit with a resistor :showing voltage drop .

In this circuit, the voltage drop across the resistor is exactly in phase with the current as shown in the figure. That means when voltage is zero, the current value at that instant is zero. When the voltage is positive the current is also positive, they rise and fall simultaneously and are said to be “in-phase”. This results in that the AC power in a resistor keeps dissipating as heat while taking it from the source, irrespective of whether the current is positive or negative.

AC Circuit with Inductors

Inductors oppose the change in the current through them. This means when the current is increased, the induced voltage tries to oppose this change of the current by dropping the voltage. The voltage dropped across an inductor is proportional to the rate of change in the current.

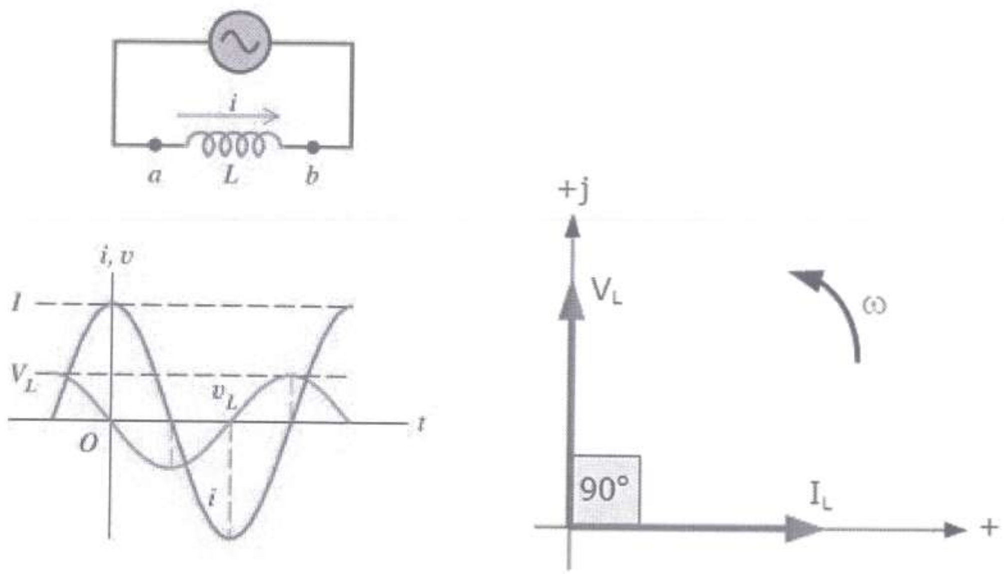


Figure 2.2: AC with Inductor

Therefore, when the current is at its peak (no rate of change in shape), the voltage at that instant is zero, and reverse happens when the current peaks at zero (maximum change of its slope), as shown in the figure. So the power dissipation in the inductor AC circuit is zero.

Thus, the instantaneous power of the inductor is 90 degrees apart so the power is negative, at times, as shown in the figure. The meaning of Negative power means that the power is being released back to the circuit as it absorbs it in the rest of the cycle. Reactance is this opposition to change of current and it depends on the frequency of the operating circuit.

AC Circuit with Capacitors

While an Inductor opposes a change in current capacitor opposes a change in voltage. This happens through either supplying or drawing current from the circuit in proportion to the rate of change of voltage.

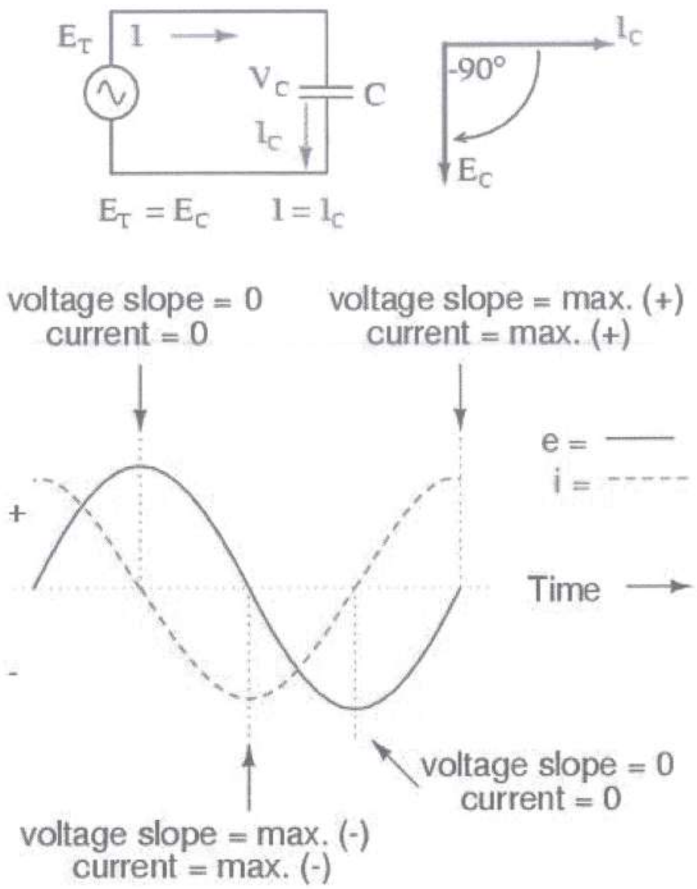


Figure 2.3: Showing how capacitor opposes a change in voltage

2.7 Summary

- Kirchoff's Current Law indicates that "The current flowing into a node in a circuit is equal to the current flowing out of it".
- Kirchoff's Voltage Law states that "In any complete loop within a circuit, the sum of all voltages from components that supply electrical energy, i.e. cells or generators, will equal the sum of all voltage drops across the other components in the same loop"
- Ohm's law may not work for circuits where resistance changes with temperature ie. through heat dissipated through current
- Electrical circuits in practice are never simple single loop paths but can have multiple branches, multiple components and energy sources. Understanding the parameters of the circuits will need Network analysis
- There are many types of network and most networks fall into the type called Lumped Network where each circuit can be physically separated
- Serial and Parallel circuits are a necessity and many times a combination of them

2.8 Keywords

- *Direct Current:* No change in polarity
- *Alternating current:* Exhibits change of polarity for voltage multiple times in a second
- *Resistance:* Resistance is a measure of the opposition to current flow in an electrical circuit. Resistance is measured in Ohm
- Series Circuit: when the circuit components (resistor, inductor, capacitor,..) are connected in series.
- Parallel Circuit: When the circuit components are connected in parallel
- Graph of a network: A simple hand drawn representation of a network

2.9 Review Questions

a. State Kirchoff's First law

b. State Kirchoff's Second Law

c. Describe the process of network analysis

d. What are Series circuits and parallel circuits ?

e. List the different types of Network

f. Explain the behavior of a resistor in series vs in parallel circuits

g. Explain the behavior of a capacitor in series vs in parallel circuits

h. Explain the behavior of an inductor in series vs in parallel circuits

i. Three resistors R1=1.00 Ω, R2=6.00 Ω, and R3=5.00 Ω, are connected in parallel. The parallel connection is attached to a 9V voltage source.

   a. What is the equivalent resistance?

   b. Find the current supplied by the source to the parallel circuit.

j. State the difference between capacitor, Inductor and Resistor.

2.0 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication.
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education.
4. Millman, Electronics Devices and Circuits, TMH.
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi.
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India.

Learning Objectives

After studying this unit, the student will be able to:

1 Explain the types of semiconductor.

2 Understand the applications of semiconductors.

3 Understand about the diodes.

Introduction

On the basis of the relative values of electrical conductivity (σ) or resistivity (ρ =1/σ), the solids are classified as:

Metals: They possess very low resistivity or high conductivity.
$\rho \sim 10^{-2} - 10^{-8}\ \Omega$ m
$\sigma \sim 10^{2} - 10^{8}$ S m –1
Semiconductors: They have resistivity or conductivity intermediate to metals and insulators.
$\rho \sim 10^{-5} - 10^{6}\ \Omega$ m
$\sigma \sim 10^{5} - 10^{-6}$ S m–1
Insulators: They have high resistivity or low conductivity.
$\rho \sim 10^{11} - 10^{19}\ \Omega$ m
$\sigma \sim 10^{-11} - 10^{-19}$ S m –1

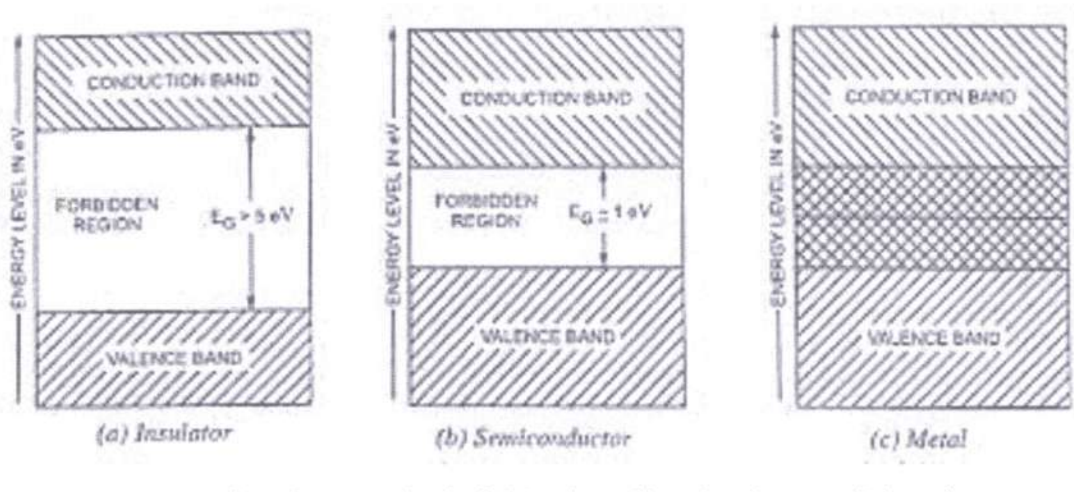


Figure 3.1 energy bands of (a) insulators,(b) semiconductors and (c) metals

The values of ρ (resistivity) and σ (conductivity) given above are indicative of magnitude and could well go outside the ranges as well. Relative values of the resistivity are not the only criteria for distinguishing metals, insulators and semiconductors from each other.

Case I: This refers to a situation, as shown in Figure 3.1(a). One can have a metal either when the conduction band is partially filled or the balanced band is partially empty or when the conduction and valence bands overlap. When there is overlap, electrons from the valence band can easily move into the conduction band. This situation makes a large number of electrons available for electrical conduction. When the valence band is partially empty, electrons from its lower level can move to higher level making conduction possible. Therefore, the resistance of such materials is low or the conductivity is high.

Case II: In this case, as shown in Figure 3.1(b), a large band gap E_g exists ($E_g > 3$ eV). There are no electrons in the conduction band, and therefore no electrical conduction is possible. Note that the energy gap is so large that electrons cannot be excited from the valence band to the conduction band by thermal excitation. This is the case of insulators.

Case III: This situation is shown in Figure 3.1(c). Here a finite but small band gap ($E_g <$ 3 eV) exists. Because of the small band gap, at room temperature some electrons from valence band can acquire enough energy to cross the energy gap and enter the conduction band. These electrons (though small in numbers) can move in the conduction band. Hence, the resistance of semiconductors is not as high as that of the insulators.

3.1 Semiconductor

A semiconductor material is one whose electrical properties lie in between those of insulators and good conductors. Examples are: germanium and silicon.

In terms of energy bands, semiconductors can be defined as those materials which have almost filled valence band and almost an empty conduction band with a very narrow energy gap (of the order of 1 eV) separating the two.

If there is some gap between the conduction band and the valence band, electrons in the valence band all remain bound and no free electrons are available in the conduction band. This makes the material an insulator.

But some of the electrons from the valence band may gain external energy to cross the gap between the conduction band and the valence band. Then these electrons will move into the conduction band. At the same time they will create vacant energy levels in the valence band where other valence electrons can move. Thus the process creates the possibility of conduction due to electrons in the conduction band as well as due to vacancies in the valence band.

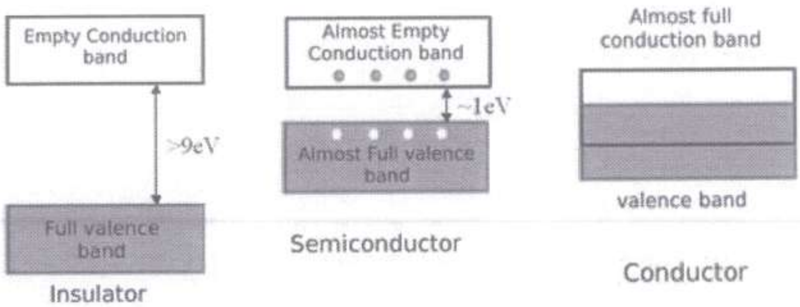


Figure 3.2 Insulator, Semiconductor, conductor

A semiconductor is primarily an insulator at 0K. However, since the energy gap is lower compared to insulators (~1eV), the valence band is slightly thermally populated at room temperature, whereas the conduction band is slightly depopulated. Since electrical conduction is directly connected to the number of electrons in the "almost empty" conduction band and to the number of holes in the "almost fully occupied" valence band, it can be expected that the electrical conductivity of such an intrinsic semiconductor will be very small.

Holes and Electrons in Semiconductors:

Holes and electrons are the types of charge carriers accountable for the flow of current in semiconductors. **Holes** (valence electrons) are the positively charged electric charge carrier whereas **electrons** are the negatively charged particles. Both electrons and holes are equal in magnitude but opposite in polarity.

Mobility of Electrons and Holes

In a semiconductor, the mobility of electrons is higher than that of the holes. It is mainly because of their different band structures and scattering mechanisms.

Electrons travel in the conduction band whereas holes travel in the valence band. When an electric field is applied, holes cannot move as freely as electrons due to their restricted moment. The elevation of electrons from their inner shells to higher shells results in the creation of holes in semiconductors. Since the holes experience stronger atomic force by the nucleus than electrons, holes have lower mobility.

The mobility of a particle in a semiconductor is more if;

- Effective mass of particles is lesser
- Time between scattering events is more

For intrinsic silicon at 300 K, the mobility of electrons is 1500 $cm^2 (V\cdot s)^{-1}$ and the mobility of holes is 475 $cm^2 (V\cdot s)^{-1}$.

The bond model of electrons in silicon of valence 4 is shown below in figure 3.3. Here, when one of the free electrons (blue dots) leaves the lattice position, it creates a hole (grey dots). This hole thus created takes the opposite charge of the electron and can be imagined as positive charge carriers moving in the lattice.

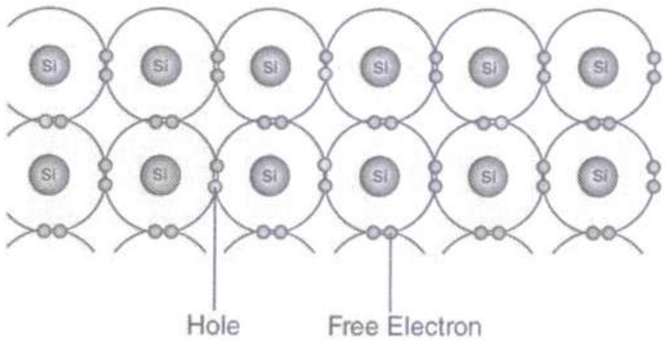


Figure 3.3 Electrons and Holes in Semiconductors

Band Theory of Semiconductors

We know that the electrons in an atom are present in different energy levels. When we try to assemble a lattice of a solid with N atoms, then each level of an atom must split up into N levels in the solid. This splitting up of sharp and tightly packed energy levels forms Energy Bands. The gap between adjacent bands representing a range of energies that possess no electron is called a Band Gap.

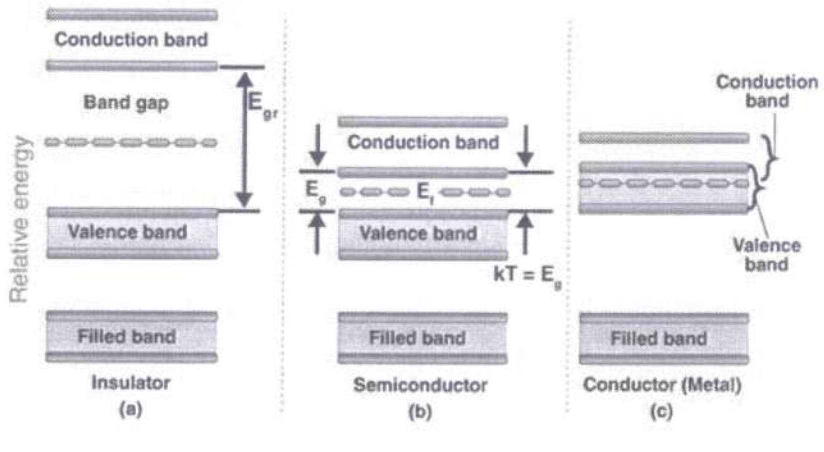


Figure 3.4 Energy Band Diagram for Insulator, Semiconductor and conductor

Conduction Band and Valence Band in Semiconductors:

Valence Band:

The energy band involving the energy levels of valence electrons is known as the valence band. It is the highest occupied energy band. When compared with insulators, the band gap in semiconductors is smaller. It allows the electrons in the valence band to jump into the conduction band on receiving any external energy.

Conduction Band:

It is the lowest unoccupied band that includes the energy levels of positive (holes) or negative (free electrons) charge carriers. It has conducting electrons resulting in the flow of current. The conduction band possesses a high energy level and are generally empty. The conduction band in semiconductors accepts the electrons from the valence band.

Fermi Level in Semiconductors:

Fermi level (denoted by E_F) is present between the valence and conduction bands. It is the highest occupied molecular orbital at absolute zero. The charge carriers in this state have their own quantum states and generally do not interact with each other. When the temperature rises above absolute zero, these charge carriers will begin to occupy states above Fermi level.

In a p-type semiconductor, there is an increase in the density of unfilled states. Thus, accommodating more electrons at the lower energy levels. However, in an n-type semiconductor, the density of states increases, therefore, accommodating more electrons at higher energy levels.

Properties of Semiconductors:

Semiconductors can conduct electricity under preferable conditions or circumstances. This unique property makes it an excellent material to conduct electricity in a controlled manner as required.

Unlike conductors, the charge carriers in semiconductors arise only because of external energy (thermal agitation). It causes a certain number of valence electrons to cross the energy gap and jump into the conduction band, leaving an equal amount of unoccupied energy states, i.e. holes. Conduction due to electrons and holes are equally important.

- **Resistivity:** 10^{-5} to 10^{6} Ωm
- **Conductivity:** 10^{5} to 10^{-6} mho/m
- **Temperature coefficient of resistance:** Negative
- **Current Flow:** Due to electrons and holes

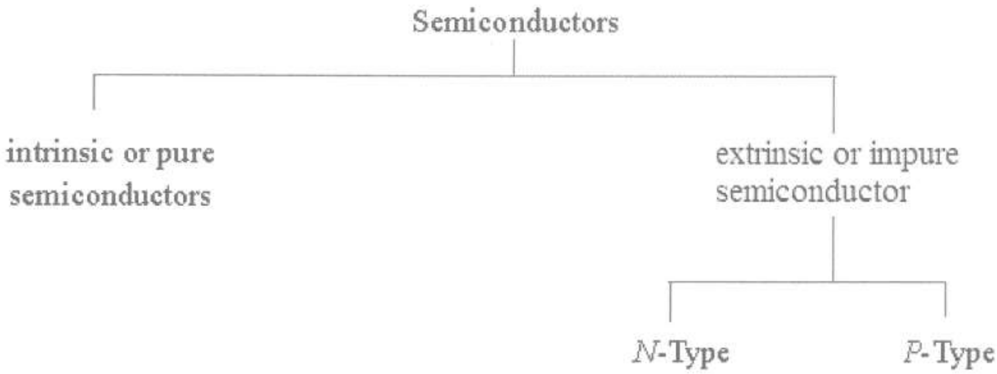
Important Properties of Semiconductors are:

1. Semiconductor acts like an insulator at Zero Kelvin. On increasing the temperature, it works as a conductor.

2. Due to their exceptional electrical properties, semiconductors can be modified by doping to make semiconductor devices suitable for energy conversion, switches, and amplifiers.
3. Lesser power losses.
4. Semiconductors are smaller in size and possess less weight.
5. Their resistivity is higher than conductors but lesser than insulators.
6. The resistance of semiconductor materials decreases with the increase in temperature and vice-versa.

3.2 Types of Semiconductors

Semiconductor may be classified as under:



Intrinsic Semiconductors

An intrinsic semiconductor is one which is made of the semiconductor material in its extremely pure form.

Examples of such semiconductors are: pure germanium and silicon which have forbidden energy gaps of 0.72 eV and 1.1 eV respectively. The energy gap is so small that even at ordinary room temperature; there are many electrons which possess sufficient energy to jump across the small energy gap between the valence and the conduction bands.

Alternatively, an intrinsic semiconductor may be defined as one in which the number of conduction electrons is equal to the number of holes.

Schematic energy band diagram of an intrinsic semiconductor at room temperature is shown in

Figure 3.5 below.

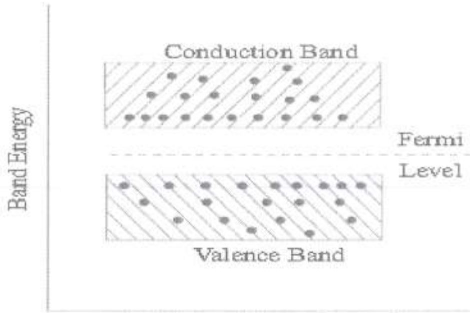


Figure 3.5 energy band diagram of an intrinsic semiconductor

In intrinsic semiconductors, the number of free electrons, n_e is equal to the number of holes, n_h. That is $n_e = n_h = n_i$
where n_i is called intrinsic carrier concentration.

Extrinsic Semiconductors:

Those intrinsic semiconductors to which some suitable impurity or doping agent or doping has been added in extremely small amounts (about 1 part in 10^8) are called extrinsic or impurity semiconductors.

Depending on the type of doping material used, extrinsic semiconductors can be subdivided into two classes:

(i) N-type semiconductors and
(ii) P-type semiconductors.

(i) N-type Extrinsic Semiconductor:

This type of semiconductor is obtained when a pentavalent material like antimonty (Sb) is added to pure germanium crystal. As shown in Figure 3.6 below, each antimony atom forms covalent bonds with the surrounding four germanium atoms with the help of four of its five electrons. The fifth electron is superfluous and is loosely bound to the antimony atom.

Hence, it can be easily excited from the valence band to the conduction band by the application of electric field or increase in thermal energy. It is seen from the above description that in N-type semiconductors, electrons are the majority carriers while holes constitute the minority carriers.

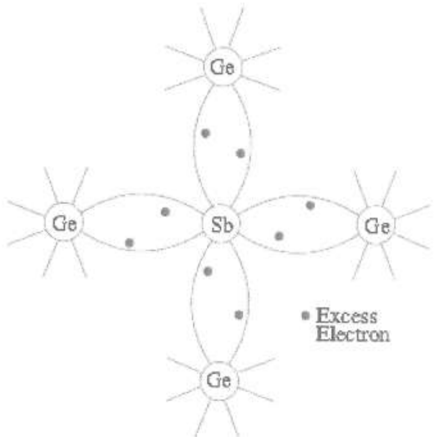


Figure 3.6 Pentavalent material

(ii) P-type Extrinsic Semiconductor:

This type of semiconductor is obtained when traces of a trivalent like boron (B) are added to a pure germanium crystal. In this case, the three valence electrons of a boron atom form covalent bonds with four surrounding germanium atoms but one bond is left incomplete and gives rise to a hole as shown in Figure 3.7 below.

Thus, boron which is called an acceptor impurity causes as many positive holes in a germanium crystal as there are boron atoms thereby producing a P-type (P for positive) extrinsic semiconductor.

In this type of semiconductor, conduction is by the movement of holes in the valence band.

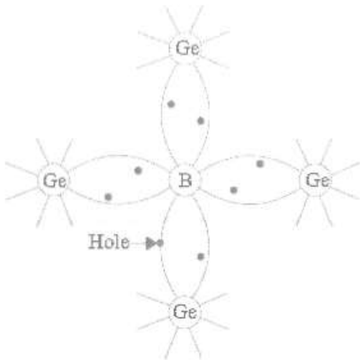


Figure 3.7 Trivalent material

Majority and Minority Carriers:

In a piece of pure germanium or silicon, no free charge carriers are available at 0ºK. However, as its temperature is raised to room temperature, some of the covalent bonds are broken by heat energy and as a result; electron-hole pairs are produced. These are called thermally-generated charge carriers.

They are also known as intrinsically-available charge carriers. Ordinarily, their number is quite small. An intrinsic of pure germanium can be converted into a P-type semiconductor by the addition of an acceptor impurity which adds a large number of holes to it. Hence, a P-type material contains following charge carriers:

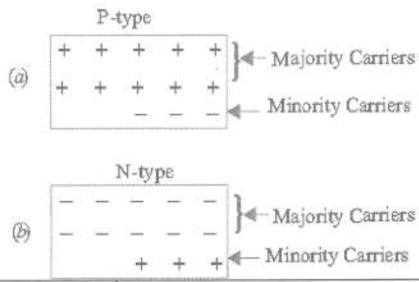
(a) Large number of positive holes—most of them being the added impurity holes with only a very small number of thermally generated ones.

(b) A very small number of thermally-generated electrons (the companions of the thermally generated holes mentioned above).

Obviously, in a P-type material, the number of holes (both added and thermally- generated) is much more than that of electrons. Hence, in such a material, holes constitute majority carriers and electrons form minority carriers as shown in Figure 3.8 (a) below. Similarly, in an N-type material, the number of electrons (both added and thermally-generated) is much larger than the number of thermally-generated holes. Hence, in such a material, electrons are majority carriers whereas holes are minority carriers as shown in Figure 3.8 (b) below.

Figure 3.8 P-type and N-type

Difference Between Intrinsic and Extrinsic Semiconductors:



Intrinsic Semiconductor	Extrinsic Semiconductor
Pure semiconductor	Impure semiconductor
Density of electrons is equal to the density of holes	Density of electrons is not equal to the density of holes
Electrical conductivity is low	Electrical conductivity is high
Dependence on temperature only	Dependence on temperature as well as on the amount of impurity
No impurities	Trivalent impurity, pentavalent impurity

Applications of Semiconductors

Semiconductors are used in almost all electronic devices. Without them, our life would be much different.

Their reliability, compactness, low cost and controlled conduction of electricity make them ideal to be used for various purposes in a wide range of components and devices. transistors, diodes, photosensors, microcontrollers, integrated chips and much more are made up of semiconductors.

Uses of Semiconductors in Everyday life

- Temperature sensors are made with semiconductor devices.
- They are used in 3D printing machines
- Used in microchips and self-driving cars
- Used in calculators, solar plates, computers and other electronic devices.

- Transistor and MOSFET used as a switch in Electrical Circuits are manufactured using the semiconductors.

Industrial Uses of Semiconductors

The physical and chemical properties of semiconductors make them capable of designing technological wonders like microchips, transistors, LEDs, solar cells, etc.

The microprocessor used for controlling the operation of space vehicles, trains, robots, etc is made up of transistors and other controlling devices which are manufactured by semiconductor materials.

Importance of Semiconductors:

We have some advantages of semiconductors which makes them highly useful everywhere.

- They require less input power
- They are highly portable due to the smaller size
- Semiconductor devices are shockproof
- They have a longer lifespan
- They are noise-free while operating

The Mass Action Law

This relationship is valid for both intrinsic and extrinsic semiconductors. In an extrinsic semiconductor the increase in one type of carrier (n or p) reduces the concentration of the other through recombination so that the product of the two (n and p) is a constant at any given temperature.
The carriers whose concentration in extrinsic semiconductors is the larger are designated the majority carriers, and those whose concentration is the smaller the minority carriers. At equilibrium, with no external influences such as light sources or applied voltages, the concentration of electrons, n_e, and the concentration of holes, n_p, are related by
$n_e \times n_p = n_i^2$
n_i denotes the carrier concentration in intrinsic semiconductor.

A material is defined as intrinsic when it consists purely of one element and no outside force (like light energy) affects the number of free carriers other than heat energy. In intrinsic Si, the heat energy available at room temperature generates approximately $1.5x10^{10}$ carriers per cm^3 of each type (holes and electrons) . The number of free carriers doubles for approximately every 11°C increase in temperature. This number represents a very important constant (at room temperature), and we define $n_i = 1.5x10^{10}$ cm^{-3} where n_i denotes the carrier concentration in intrinsic silicon at room temperature (constant for a given temperature).

Based on charge neutrality, for a sample doped with N_D donor atoms per cm^{-3} and N_A acceptor atoms per cm^{-3},
We can write $n_e + N_A = n_p + N_D$

which shows that the sum of the electron concentration plus the ionized acceptor atoms is equal to the sum of the hole concentration plus the ionized donor atoms. The equation assumes that all donors and acceptors are fully ionized, which is generally true at or above room temperature. Given the impurity concentration, the above equations can be solved simultaneously to determine electron and hole concentrations.

In electronic devices, we typically add only one type of impurity within a given area to form either n-type or p-type regions. In n-type regions there are typically only donor impurities and the donor concentration is much greater than the intrinsic carrier concentration, $N_A=0$ and $N_D>>n_i$.

Under these conditions we can write $n_e \approx N_D$ where n_e is the free electron concentration in the n-type material and N_D is the donor concentration (number of added impurity atoms/cm^3). Since there are many extra electrons in n-type material due to donor impurities, the number of holes will be much less than in intrinsic silicon and is given by,

$n_p= n_i^2/N_D$

where n_p is the hole concentration in an n-type material and n_i is the intrinsic carrier concentration in silicon.

Similarly, in p-type regions we can generally assume that $N_D=0$ and $N_A>>n_i$. In p-type regions, the concentration of positive carriers (holes), n_p, will be approximately equal to the acceptor concentration, N_A. $n_p = NA$ and the number of negative carriers in the p-type material, n_p, is given by $n_p = n_i^2 / N_A$

Example 1: Pure Silicon semiconductor at 500K has equal electrons and holes (1.5×10^{16} m ^{3}). Doping by Indium increases n_h to 4.5×10^{22} m^{-3}. Calculate the type and electron concentration of doped semiconductor.

Solution:

Since, $n^2_i = n_e\, n_h$

$(1.5 \times 10^{16})^2 = n_e\, (4.5 \times 10^{22})$

Therefore, $n_e = 5 \times 10^9$

Given $n_h = 4.5\times10^{23}$

$\Rightarrow n_h >> n_e$

Therefore, the semiconductor is p-type and $n_e = 5 \times 10^9$ m^{-3}.

Example 2: Why the valence band in semiconductors is partially empty and the conduction band is partially filled at room temperature?

Solution:

In semiconductors, the conduction band is empty and the valence band is completely filled at Zero Kelvin. No electron from the valence band can cross over to the conduction band at this temperature. But at room temperature, some electrons in the valence band jump over to the conduction band due to a small forbidden gap i.e. 1 eV.

Example 3: In an intrinsic semiconductor, the number of conduction electrons is $7 \times 10^{19}\ m^3$. Find the total number of current carriers in the same semiconductor of size 1 cm × 1 cm × 1 mm.

Solution:

In an intrinsic semiconductor; $n_e = n_h$
Given, ne = 7×10^{19} per m^3
Therefore, $n_h = n_e = 7 \times 10^{19}\ m^3$
So, the total current carrier density = $n_e + n_h = 7\times10^{19} + 7\times10^{19} = 14\times10^{19}$ per m^3
Now, the total number of current carrier = Number density × volume
$= (14 \times 10^{19}$ per $m^3) \times (10^{-2}m \times 10^{-2}m \times 10^{-3}m) = 14\times10^{12}$.

Example 4: The energy gap of silicon is 1.14 eV. What is the maximum wavelength at which silicon will begin absorbing energy?

Solution:

Since hc/λ = Energy (E)
Therefore, $\lambda = hc/E$
$= [(6.628 \times 10^{-34}) \times (3\times10^{8})]/1.14\times1.6\times10^{-19}J$
$= 10.901 \times 10^{-7}\ m = 10901\ Å$.

Example 5: Find the resistivity of an intrinsic semiconductor with intrinsic concentration of 2.5×10^{19} *per* m^3. The mobilities of electrons and holes are 0.40 $m^2/V\text{-}s$ and 0.20 $m^2/V\text{-}s$.

Solution:

Given data are:

Intrinsic concentration $(n_i) = 2.5 \times 10^{19}/m^3$

Mobility of electrons $(\mu_n) = 0.40\ m^2/V\text{-}s$

The mobility of holes (μ_p) = 0.20 m^2/V-s

The conductivity of an intrinsic semiconductor (σ_i) = $n_i e[\mu_n + \mu_p]$

$$\text{The resistivity } (\rho_i) = \frac{1}{\sigma_i} = \frac{1}{n_i e\left[\mu_n + \mu_p\right]}$$

$$= \frac{1}{2.5\times10^{19}\times1.6\times10^{-19}\left[0.40+0.20\right]}$$

$$= \frac{1}{2.5\times1.6\times0.6} = 0.4166\ \Omega\text{-m.}$$

Example 5: A silicon bar is doped with donor impurities N_D= 2.25 $\times10^{15}$ $atoms/cm^3$. Given the intrinsic carrier concentration of silicon at T = 300 K is n_i= 1.5 × 10^{10} cm^{-3}. Assuming complete impurity ionization, what will be the equilibrium electron and hole concentrations ?

Solution:

Donor impurities, N_D= $2.25 \times10^{15} atom/cm^3$ when there is complete impurity ionization the number of electrons provided will be the same as Donor impurities. i.e.

$n = N_D = 2.25 \times10^{15}/cm^3$

$P_0 = n_i^2/n_o$

$n_0 = (1.5 \times 10^{10})^2/2.25 \times15^{15} = 2.25 \times10^{20}/2.25 \times10^{15}$

$= 10^5/cm^3$

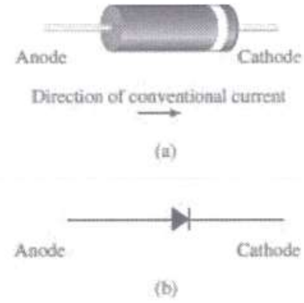
3.3 P-N Junction Diode

PN junction diodes are nonlinear circuit elements. PN junction is an important semiconductor device in itself and used in a wide variety of applications such as rectifiers, Photo detectors, Clipper and Clamper circuits, laser diode (LD) ,light emitting diodes (LED) etc. PN junctions are an integral part of other important semiconductor devices such as BJTs, JFETS and MOSFETs.

Construction:

It is two terminal devices consisting of a P-N junction formed either in Ge or Si crystal. Its circuit symbol is shown in figure 3.9 (a). The P and N type regions are referred to as anode and cathode respectively. In figure 3.9 (b) arrowhead indicates the conventional direction of current flow when forward biased. It is the same direction in which hole flow takes place.

Figure 3.9 PN Junction Diode



Working:

A P-N junction diode is a one way device offering low resistance when forward biased and behaving almost as an insulator when reverse biased. Hence such diodes are mostly used as rectifiers for converting alternating current into direct current.

V/I Characteristic

Figure 3.10 shows the static voltage current characteristics for a low power P-N junction diode.

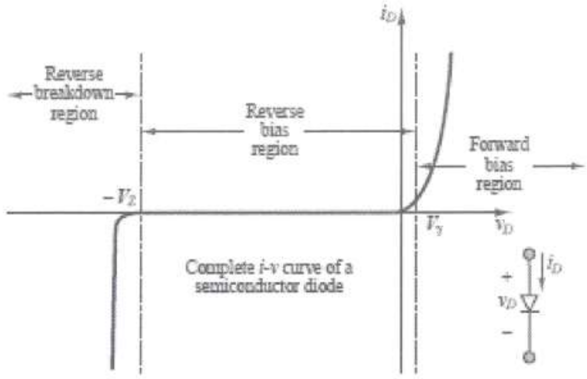


Figure 3.10 V-I Characteristic

Forward characteristic:

When the diode is forward biased and applied voltage is increased from zero hardly any current flows through the device in the beginning. It is so because the external voltage is being opposed by the internal barrier voltage V_B whose value is 0.7 V for Si and 0.3 V for Ge. As soon as V_B is neutralized, current through the diode increases rapidly with increasing applied battery voltage. It is found that as little a voltage as 1.0 V produces a forward current of about 50 mA.

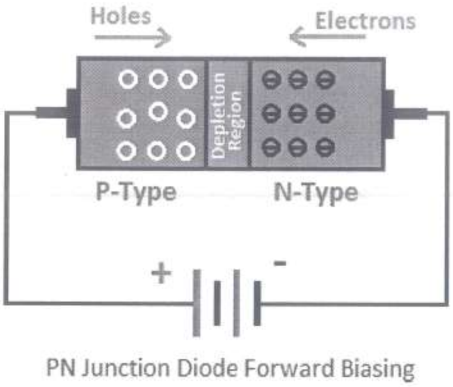


Figure 3.11 Forward biasing of PN junction Diode

Reverse characteristic:

When the diode is reverse biased, majority carriers are blocked and only a small current (due to minority carriers) flows through the diode. As the reverse voltage is increased from zero, the reverse current very quickly reaches its maximum or saturation value I_0 which is also known as leakage current. It is of the order of nano ampere (nA) for Si and micro ampere (μA) for Ge. As seen from figure 12 when reverse voltage V_{BR}, the leakage current suddenly and sharply increases, the curve indicating zero resistance at this point.

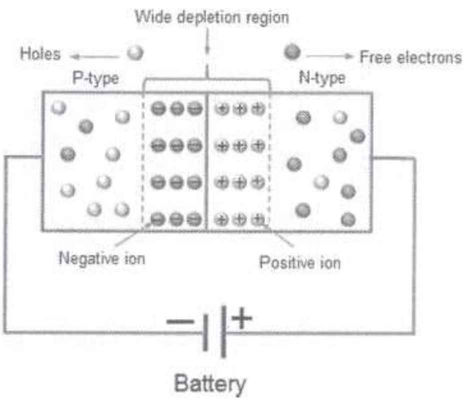


Figure 3.12 Reverse biasing of PN junction Diode

Important terms used for a p-n junction diode:

Breakdown Voltage: It is the minimum voltage at which p-n junction breaks down with sudden rise in reverse current. Knee Voltage: It is the forward voltage at which the current through the junction starts to increase rapidly. Maximum forward Current: It is the highest instantaneous forward current that a p-n junction can conduct without damage to the junction .Current that a p-n junction can conduct without damage to the junction. Peak Inverse voltage (PIV): It is the maximum reverse voltage that can be applied to the p-n junction without damage to the junction. Maximum power rating: It is the maximum power that can be dissipated at the junction without damaging it.

Diode Parameters

Bulk resistance (r_B)

It is the sum of the resistance values of the P and N type semiconductor materials of which the diode is made of

$$r_{B} = r_P + r_N$$

Usually, it is very small, it is given by

$$r_B = \frac{V - V_K}{I_F}$$

It is the resistance offered by the diode well above the knee voltage when current resistance is large.

Obviously, this resistance is offered in the forward direction.

Junction resistance (r_j)

Its value for forward biased junction depends on the magnitude of forward dc current.

$$r_j = \frac{25\,mV}{I_F mA} \dots\dots\dots\dots \quad for\ Ge$$

$$r_j = \frac{50\,mV}{I_F\,mA} \dots\dots\dots\dots \quad for\ Si$$

Dynamic or ac resistance

r_{ac} or $r_d = r_B + r_j$

For large values of forward current, r_j is negligible. Hence, $r_{ac} = r_B$ for small values of I_F, r_B is negligible as compared to r_j

$r_{ac} = r_j$

Forward voltage drop

It is given by the relation

$$forward\ voltage\ drop = \frac{power\ dissipated}{forward\ dc\ current}$$

Reverse saturation current (I_0) Reverse breakdown voltage (V_{BR}) Reverse dc resistance (R_R)

$$R_R = \frac{reverse\ voltage}{reverse\ current}$$

Equation of diode current

The analytical equation which describes both the forward and reverse characteristics is called the Boltzmann's diode equation given

$$I = I_0(e^{40V} - 1) \ldots\ldots\ldots. for\ Ge$$

$$I = I_0 e^{40V} \quad if\ V > 1\ volt.$$

And $I = I_0(e^{20V} - 1) \ldots\ldots\ldots for\ Si$

$$I = I_0 e^{20V} \quad if\ V > 1\ volt.$$

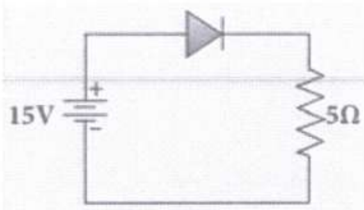
Where I_0 =reverse saturation current.

V = voltage across the diode.

Applications of P-N Junction Diode

- P-N junction diode can be used as a photodiode as the diode is sensitive to the light when the configuration of the diode is reverse-biased.
- It can be used as a solar cell.
- It is used as a rectifier in many electric circuits and as a voltage-controlled oscillator in varactors.
- When the diode is forward-biased, it can be used in LED lighting applications.

Example 6: An ideal diode and a 5 Ω resistor are connected in series with a 15 V power supply as shown in figure below. Calculate the current that flows through the diode.



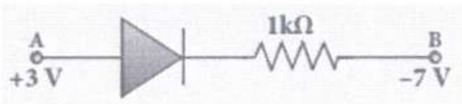
Solution:

The diode is forward biased and it is an ideal one. Hence, it acts like a closed switch with no barrier voltage. Therefore, current that flows through the diode can be calculated using Ohm's law.

$V = IR$

$I = V/R = 15/5 = 3\ A$

Example 7: Consider an ideal junction diode. Find the value of current flowing through AB is



Solution:

The barrier potential of the diode is neglected as it is an ideal diode.

The value of current flowing through AB can be obtained by using Ohm's law

$I = V/R = 3 - (-7) / 1\times 10^3 / 10^3 = 10 = 10^{-2}\ A = 10mA$

Example 8: Determine the wavelength of light emitted from LED which is made up of GaAsP semiconductor whose forbidden energy gap is 1.875 eV. Mention the colour of the light emitted (Take $h = 6.6 \times 10\text{-}34$ Js).

Solution:

$E_g = h_c / \lambda$

Therefore

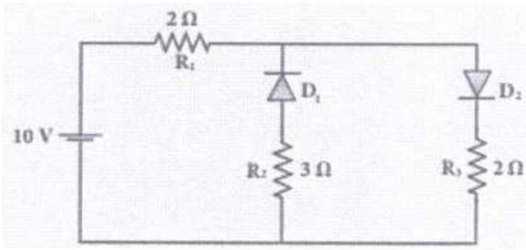
$\lambda = h_c / E_g = 6.6\times10^{-34} \times 3\times10^8 / 1.875\times1.6\times10^{-19}$

For Vivekananda Global University, Jaipur
Registrar

= 660 *nm*

The wavelength 660 nm corresponds to red colour light.

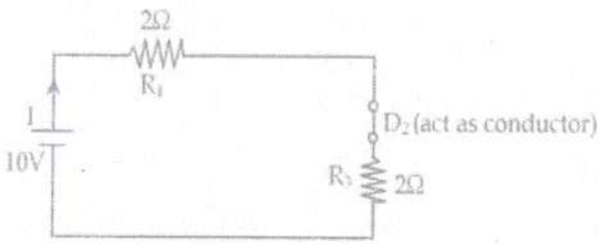
Example 9: The given circuit has two ideal diodes connected as shown in figure below. Calculate the current flowing through the resistance R1



Solution:

Barrier potential for an ideal diode is zero. The diode D_1 is reverse biased, so it will block the current and diode D_2 is forward biased, so it will pass the current.

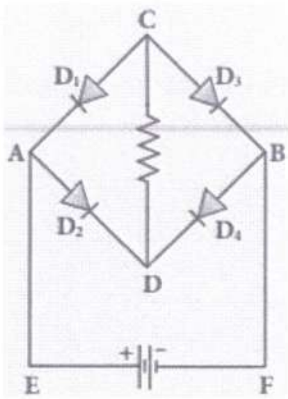
The given circuit becomes



Effective resistance R_{eff} = R1 + R3 = 4Ω

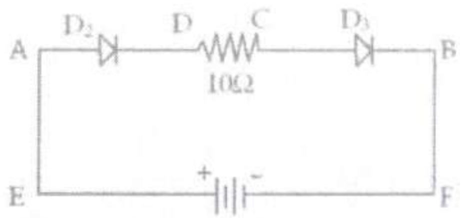
Current through R1 = V/R_{eff} = 10/4 = 2.5A

Example 9: Four silicon diodes and a 10 Ω resistor are connected as shown in figure below. Each diode has a resistance of 1Ω. Find the current flows through the 18Ω resistor.



Solution:

In the given circuit D_2 & D_3 are in forward bias so they conduct current while D_1 &D_4 are in reverse bias so they do not conduct current. So the equivalents circuit will be



the effective resistance is

$R_{eff} = 1\Omega + 10\Omega + 1\Omega = 12\Omega$

Here silicon diodes are used

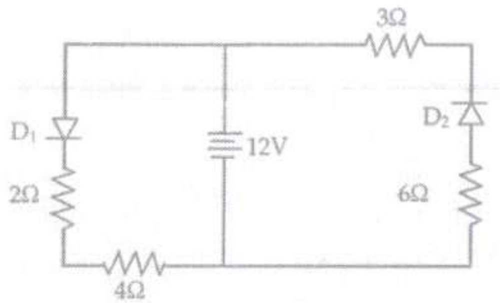
∴ Barrier potential for Si is 0.7 V

Net potential (V_{net}) = 3 – 0.7 - 0.7

V_{net} = 1.6 V

Current (I) = V_{net} / R_{eff} = 1.6 / 12 = 0.133A

Example 9: Determine the current flowing through 3Ω and 4Ω resistors of the circuit given below. Assume that diodes D1 and D2 are ideal diodes.

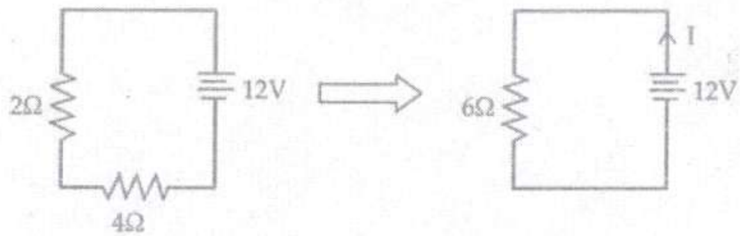


Solution:

The diode D_2 is in reverse biased. So do not conduct current.

∴ current through 3Ω is = 0

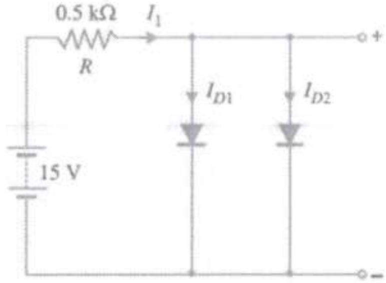
The diode D_1 is in forward biased, and it is an ideal diode. So the given circuit becomes as



The current through 4Ω is

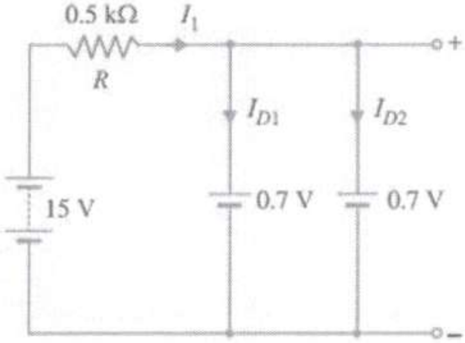
$\therefore I = V/R = 12/6 = 2A$

Example 10: Determine current through each diode in the circuit shown in Fig. 7 (i).Use simplified model. Assume diodes to be similar.



Solution:

The applied voltage forward biases each diode so that they conduct current in the same direction



$$I_1 = \frac{\text{Voltage across } R}{R} = \frac{15 - 0.7}{0.5 \text{ k}\Omega} = 28.6 \text{ mA}$$

Since the diodes are similar, $I_{D1} = I_{D2} = \frac{I_1}{2} = \frac{28.6}{2} = 14.3$ mA

3.4 Diode as a Switch

Diode is a two terminal PN junction that can be used in various applications. One of such applications is an electrical switch. The PN junction, when reverse biased acts as open circuited and when forward biased acts as close circuited. Hence the change of forward and reverse biased states makes the diode work as a switch, the forward being ON and the reverse being OFF state.

Electrical Switches over Mechanical Switches:

Electrical switches are a preferred choice over mechanical switches due to the following reasons-

- Mechanical switches are prone to oxidation of metals whereas electrical switches don't.
- Mechanical switches have movable contacts.

- They are more prone to stress and strain than electrical switches.
- The worn and torn of mechanical switches often affect their working.

Hence an electrical switch is more useful than a Mechanical switch.

Working of Diode as a Switch:

Whenever a specified voltage is exceeded, the diode resistance gets increased, making the diode reverse biased and it acts as an open switch. Whenever the voltage applied is below the reference voltage, the diode resistance gets decreased, making the diode forward biased, and it acts as a closed switch.

The following circuit explains the diode acting as a switch.

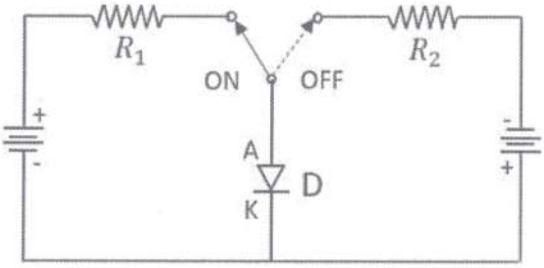


Figure 3.13 Switching Circuit using Diode

A switching diode has a PN junction in which the P-region is lightly doped and the N-region is heavily doped. The above circuit symbolizes that the diode gets ON when positive voltage forward biases the diode and it gets OFF when negative voltage reverse biased the diode.

Diode Switching Times:

While changing the bias conditions, the diode undergoes a **transient response**. The response of a system to any sudden change from an equilibrium position is called a transient response.

The sudden change from forward to reverse and from reverse to forward bias, affects the circuit. The time taken to respond to such sudden changes is the important criterion to define the effectiveness of an electrical switch.

- The time taken before the diode recovers its steady state is called Recovery **Time**.
- The time interval taken by the diode to switch from reverse biased state to forward biased state is called Forward **Recovery Time.** The time interval taken by the diode to switch from forward biased state to reverse biased state is called Reverse **Recovery Time.**

Factors that affect diode switching times:

There are few factors that affect the diode switching times, such as

- **Diode Capacitance** – The PN junction capacitance changes depending upon the bias conditions.
- **Diode Resistance** – The resistance offered by the diode to change its state.
- **Doping Concentration** – The level of doping of the diode affects the diode switching times.
- **Depletion Width** – The narrower the width of the depletion layer, the faster the switching will be. A Zener diode has a narrower depletion region than an avalanche diode, which makes the former a better switch.

Applications:

There are many applications in which diode switching circuits are used, such as –

- High speed rectifying circuits
- Consumer applications
- Automotive applications
- High speed switching circuits
- RF receivers
- General purpose applications
- Telecom applications etc.

3.5 Summary

- A semiconductor material is one whose electrical properties lie in between those of insulators and good conductors. Examples are: germanium and silicon.
- PN junction diodes are nonlinear circuit elements. PN junction is an important semiconductor device in itself and used in a wide variety of applications such as rectifiers, Photo detectors, Clipper and Clamper circuits, laser diode (LD) ,light emitting diodes (LED) etc.
- The PN junction, when reverse biased acts as open circuited and when forward biased acts as close circuited. Hence the change of forward and reverse biased states makes the diode work as a switch, the forward being ON and the reverse being OFF state.

3.6 Keywords

- **LED:** Light Emitting Diode
- **LD:** Laser Diode
- **PIV:** Peak Inverse voltage

3.7 Review Questions

Question 1: What is semiconductor?

Solution:

A semiconductor material is one whose electrical properties lie in between those of insulators and good conductors. Examples are: germanium and silicon.

Question 2: What are the properties of semiconductors?

Solution:

Semiconductors can conduct electricity under preferable conditions or circumstances. This unique property makes it an excellent material to conduct electricity in a controlled manner as required.

Unlike conductors, the charge carriers in semiconductors arise only because of external energy (thermal agitation). It causes a certain number of valence electrons to cross the energy gap and jump into the conduction band, leaving an equal amount of unoccupied energy states, i.e. holes. Conduction due to electrons and holes are equally important.

- **Resistivity:** 10^{-5} to 10^{6} Ωm
- **Conductivity:** 10^{5} to 10^{-6} mho/m
- **Temperature coefficient of resistance:** Negative
- **Current Flow:** Due to electrons and holes

3.8 Exercises

1. Discuss Holes and electrons in Semiconductors.
2. Explain Band Theory.
3. Explain important properties of Semiconductor.
4. Discuss different types of Semiconductor.
5. Differentiate between extrinsic and intrinsic semiconductor.
6. Explain the working of diodes as a switch.
7. Describe P-N junction diode.
8. Discuss applications and importance of P-N junction diodes.
9. Describe working of P-N junction diode with neat diagram.
10. Discuss construction of P-N junction diode. What is meant by majority and minority carriers?

3.9 References

1. The Physics of Information Technology, Neil Gershenfeld, Cambridge University Press (2000)
2. Electronics and Communications for Scientist and Engineers, Martin Plonus, Harcourt Academic Press (2001)
3. http://en.wikipedia.org
4. Introduction to Electrodynamics, 3rd edition, David J. Griffiths, Prentice Hall (1999)

Table of Content

Learning Objectives

After studying this unit the student will be able to:

1 Describe Rectifiers and its applications

2 Explain the concepts Capacitors and inductors and its types

3 Discuss a L-C filters

4 Explain Photovoltaic cell

Introduction

The PN junction can be use to make various electronic devices suc as rectifier, photovoltaic cell etc. capacitors and inductors are also used in various electronic circuits as filters.

4.1 Rectifier

A rectifier is an electronic device that converts an alternating current into a direct current by using one or more P-N junction diodes. A diode behaves as a one-way valve that allows current to flow in a single direction. This process is known as rectification.

Different Types of Rectifier

Rectifiers are mainly classified into two types as:

1. Uncontrolled Rectifier
2. Controlled Rectifier

4.1.1 Uncontrolled Rectifiers

The type of rectifier whose voltage cannot be controlled is known as an uncontrolled rectifier. Uncontrolled rectifiers are further divided as follows:

- Half Wave Rectifier
- Full Wave Rectifier

The type of rectifier that converts only the half cycle of the alternating current into the direct current is known as a half-wave rectifier. Likewise, a full-wave rectifier converts both positive and negative half cycles of the AC. An example of this is a bridge rectifier. A bridge rectifier uses 4 diodes that are connected in the form of a Wheatstone bridge.

4.1.2 Controlled Rectifiers

A type of rectifier whose voltage can be varied is known as the controlled rectifier. We use SCRs, MOSFETs and IGBTs to make an uncontrolled rectifier a controlled one. These rectifiers are preferred over their uncontrolled counterparts. There are two types of controlled rectifiers, and they are Half Wave Controlled Rectifier and Full Wave Controlled Rectifier. Half-wave controlled rectifier has the same design as the half-wave uncontrolled rectifier except we replace the diode with an SCR.

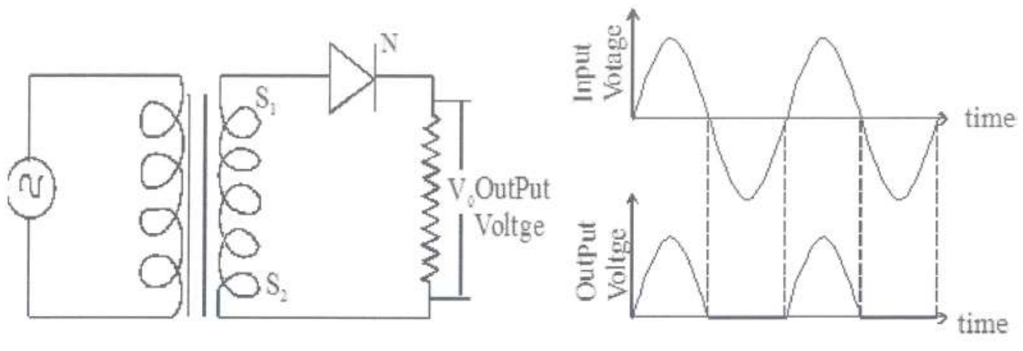
Half wave rectifier

The circuit diagram for half wave rectifier is shown in figure .The source of ac voltage is connected to primary coil of transformer. The secondary coil is connected to the diode D & load resistance R_L in series. For positive half cycle of ac voltage end S1 is at positive voltage

& S2 is at negative voltage. During this state, diode is in forward bias & a current I flow in the circuit.

Thus we get output across RL For negative half cycle S1 negative & S2 is positive, in this state diode is reversed biased & no current flow in the circuit & thus no output across load. The process is repeated. In the output, we have current corresponding to one half of the wave, the other is missing .That is why the process is called half wave rectification.

Input & output voltage is shown in figure.



Full wave rectifier -

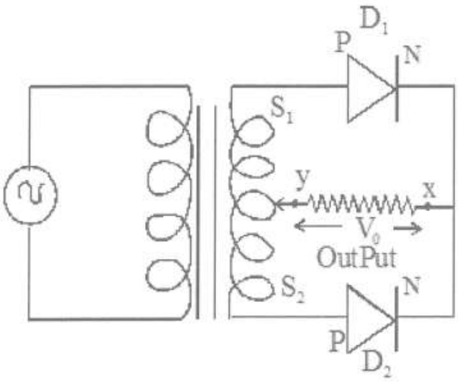
A full wave rectifier uses both parts of input ac voltage. In this rectifiers two diode D1 & D2 are used as shown in figure. In this circuit one diode rectifies one half cycle of input ac voltage & the other diode rectifies other half cycle.

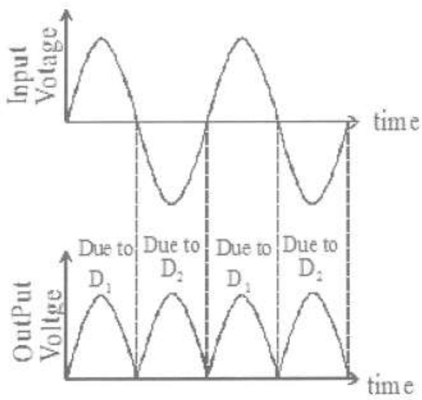
The two end S1 & S2 of secondary coil are connected to the P-terminal of the diode D1 &D2 respectively. The N terminals are connect to each other & between their common point & centre tape of secondary coil of the transformer a load resistance RL is connect .

For first half cycle of input ac voltage the ends S1 of secondary coil is at positive potential & S2 is at negative potential. The potential at centre tape is assumed to be zero. In this state, diode D1 is in forward biased while diode D2 is in reversed biased As a result, a current flow through the diode D1 for the first half of cycle .

During other half cycle, situation is reversed i.e. the end S2 become positive & S1 become negative so that now D2 is in forward biased & D1 is in reversed biased, hence current flows through the diode D2 during this half cycle.

For first half cycle diode D1 conduct & for other half cycle diode D2 conduct but current flowing through load resistance during both half cycle is same therefore output voltage is unidirectional.



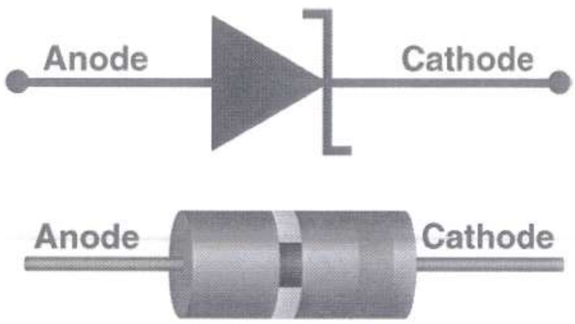


Some common applications of rectifiers are:

- Rectifiers are used in electric welding to provide polarized voltage
- Half-wave rectifiers are used as a mosquito repellent
- Half-wave rectifiers are used as a signal peak detector in AM radio
- Rectifiers are used in modulation, demodulation and voltage multipliers

4.2 Zener Diode:

A zener diode is a special type of diode designed to reliably allow current to flow "backwards" (inverted polarity) when a certain set reverse voltage, known as the zener voltage, is reached.



(Zener diode symbol and package outlines)

4.2.1 Zener Diode working

A Zener diode operates just like a normal diode when it is forward-biased. However, a small leakage current flows through the diode when connected in reverse-biased mode. As the reverse voltage increases to the predetermined breakdown voltage (Vz), current starts flowing through the diode. The current increases to a maximum, which is determined by the series resistor, after which it stabilizes and remains constant over a wide range of applied voltage.

There are two types of breakdowns for a Zener Diode:

- Avalanche Breakdown
- Zener Breakdown

Avalanche Breakdown in Zener Diode

Avalanche breakdown occurs in the normal diode and Zener Diode at high reverse voltage. When a high value of reverse voltage is applied to the PN junction, the free electrons gain sufficient energy and accelerate at high velocities. These free electrons moving at high velocity collide with other atoms and knock off more electrons.

Due to this continuous collision, a large number of free electrons are generated as a result of electric current in the diode rapidly increases. This sudden increase in electric current may permanently destroy the normal diode. However, a Zener diode is designed to operate under avalanche breakdown and can sustain the sudden spike of current.

Avalanche breakdown occurs in Zener diodes with Zener voltage (Vz) greater than 6V.

Zener Breakdown in Zener Diode

When the applied reverse bias voltage reaches closer to the Zener voltage, the electric field in the depletion region gets strong enough to pull electrons from their valence band.

The valence electrons that gain sufficient energy from the strong electric field of the depletion region break free from the parent atom. At the Zener breakdown region, a small increase in the voltage results in the rapid increase of the electric current.

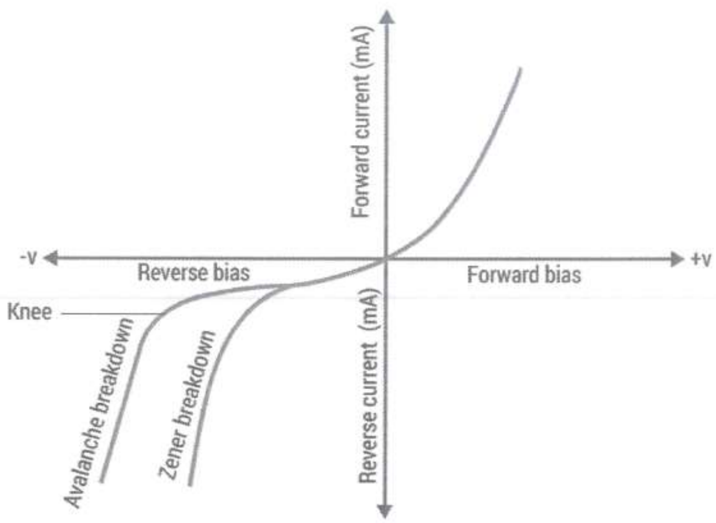
Avalanche Breakdown vs Zener Breakdown

- The Zener effect is dominant in voltages up to 5.6 volts, and the avalanche effect takes over above that.
- They are both similar effects, the difference being that the Zener effect is a quantum phenomenon, and the avalanche effect is the movement of electrons in the valence band like in an electric current.
- Avalanche effect also allows a larger current through the diode than what a Zener breakdown would allow.

When the input d.c. voltage across zener diode decreases the current through the circuit decreases sharply, causing a sufficient decrease in voltage drop across the resistor R. As a result of it, the voltage across the zener diode remains constant & hence output voltage remains constant.

4.2.2 V-I Characteristics of Zener Diode

The diagram given below shows the V-I characteristics of the Zener diode.



The V-I characteristics of a Zener diode can be divided into two parts as follows:

(i) Forward Characteristics

(ii) Reverse Characteristics

Forward Characteristics of Zener Diode

The first quadrant in the graph represents the forward characteristics of a Zener diode. From the graph, we understand that it is almost identical to the forward characteristics of any other P-N junction diode.

Reverse Characteristics of Zener Diode

When a reverse voltage is applied to a Zener voltage, a small reverse saturation current Io flows across the diode. This current is due to thermally generated minority carriers. As the reverse voltage increases, at a certain value of reverse voltage, the reverse current increases drastically and sharply. This is an indication that the breakdown has occurred. We call this voltage breakdown voltage or Zener voltage, and Vz denotes it.

4.2.3 Zener Diode Specifications

Some commonly used specifications for Zener diodes are as follows:

- Zener/Breakdown Voltage – The Zener or the reverse breakdown voltage ranges from 2.4 V to 200 V, sometimes it can go up to 1 kV while the maximum for the surface-mounted device is 47 V.

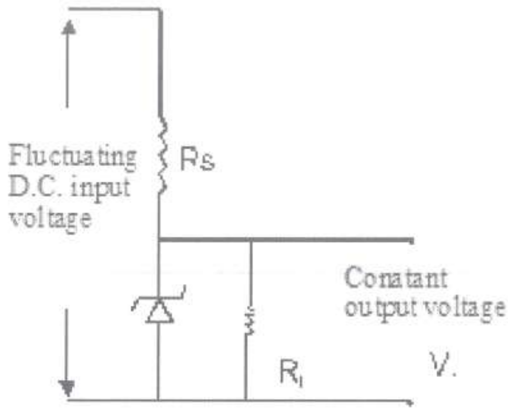
- Current Iz (max) – It is the maximum current at the rated Zener Voltage (Vz – 200μA to 200 A)
- Current Iz (min) – It is the minimum value of current required for the diode to break down.
- Power Rating – It denotes the maximum power the Zener diode can dissipate. It is given by the product of the voltage of the diode and the current flowing through it.
- Temperature Stability – Diodes around 5 V have the best stability
- Voltage Tolerance – It is typically ±5%
- Zener Resistance (Rz) – It is the resistance to the Zener diode exhibits.

4.2.4 Application of Zener Diode

Following are the applications of Zener diode:

Zener diode as Voltage Regulator

Its working is based on the fact in reverse breakdown region, a very small change in voltage across the zener diode produce a very large change in current through the circuit but the voltage across the zener remains constant. The circuit detail is shown in figure(c). Here the zener diode is joined in reverse bias to the fluctuating d.c. input voltage through a resistance R .The constant output voltage is taken across zener diode.



(c)

When the input d.c. voltage across zener diode increases beyond a break down voltage ,the current through the circuit rises sharply, causing a sufficient increase in voltage drop across the resistor R. As a result of it, the voltage across the zener diode remains constant & hence output voltage remains constant.

Zener diode in over-voltage protection:

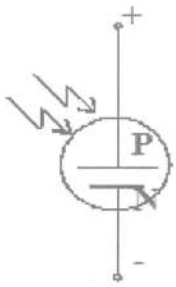
When the input voltage is higher than the Zener breakage voltage, the voltage across the resistor drops resulting in a short circuit, this can be avoided by using the Zener diode.

Zener diode in clipping circuits:

Zener diode is used for modifying AC waveform clipping circuits by limiting the parts of either one or both the half cycles of an AC waveform.

4.3 Photovoltaic cell

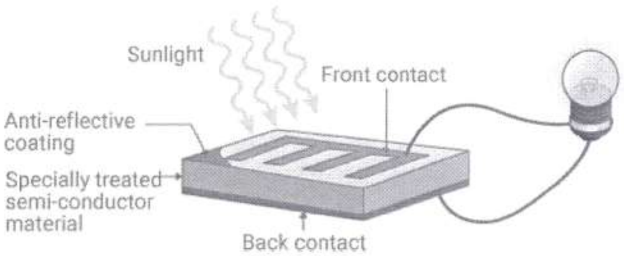
It is a cell which converts light energy into electric energy. In fact it is photodiode which is unbiased. When photons of light of energy (h $> E_g$) fall at the junction, electron - hole pairs are generated in the depletion layer. The electrons & holes produced move in opposite directions due to junction field .The photo generated electrons move towards n side of p-n junction.



(Photovoltaic cell symbol)

The photo generated hole move towards p-side of p-n junction .They will be collected at the two sides of the junction, giving rise to a photo voltage top & bottom at electrodes .

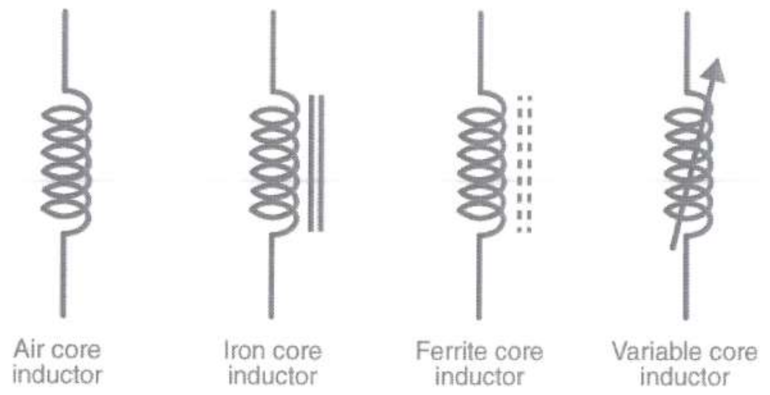
The top metal contact act as positive electrode & bottom metal contact acts as negative electrode. When an external load is connected across metal electrodes a photo current flows. The symbol of solar diode is shown in figure.



(Photovoltaic cell working)

4.4 Inductor

An inductor is a passive component that is used in most power electronic circuits to store energy in the form of magnetic energy when electricity is applied to it. One of the key properties of an inductor is that it impedes or opposes any change in the amount of current flowing through it. Whenever the current across the inductor changes it either acquires charge or loses the charge in order to equalize the current passing through it. The inductor is also called a choke, reactor or just coil.



4.4.1 Inductance

Inductance is the tendency of an electrical conductor to oppose a change in the electric current flowing through it. L is used to represent the inductance, and Henry is the SI unit of inductance.

4.4.2 Different Types of Inductors

Depending on the type of material used inductors can be classified as follows:

1. Iron Core Inductor
2. Air Core Inductor
3. Iron Powder Inductor
4. Ferrite Core Inductor which is divided into,

   - Soft Ferrite
   - Hard Ferrite

Iron Core Inductor

As the name suggests the core of this type of inductor is made of iron. These inductors are low space inductors that have high power and high inductance value. However, they are limited in high-frequency capacity. These inductors are used in audio equipment.

Air Core Inductor

These inductors are used when the amount of inductance required is low. Since there is no core, it does not have a core loss. But the number of turns the inductor must have is more for this type when compared to the inductors with the core. This results in a high-Quality factor. Usually, ceramic inductors are often referred to as air-core inductors.

Iron Powder Inductor

In this type of inductor, the core is Iron Oxide. They are formed by very fine and insulating particles of pure iron powder. High magnetic flux can be stored in it due to the air gap. The permeability of the core of this type of inductor is very less. They are usually below 100. They are mainly used in switching power supplies.

Ferrite Core Inductor

In this type of Inductor, ferrite materials are used as core. The general composition of ferrites is XFe_2O_4. Where X represents transition material. Ferrites can be classified into two types. Soft ferrites and hard ferrites.

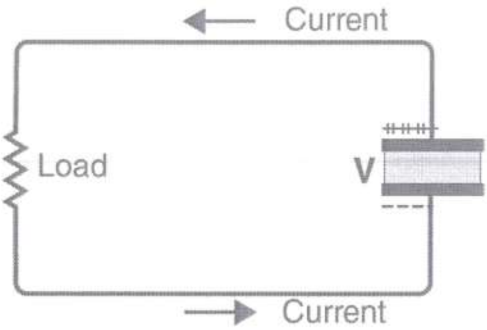
- Soft Ferrite: Materials that have the ability to reverse their polarity without any external energy.
- Hard Ferrite: These are permanent magnets. That is their polarity will not change even when the magnetic field is removed.

Choke

A choke is a type of inductor that is used mainly for blocking high-frequency alternating current (AC) in an electrical circuit. On the other hand, it will allow DC or low-frequency signals to pass. As the function of this inductor is to restrict the changes in current it is called a choke. This inductor consists of a coil of insulated wire wound on a magnetic core. The main difference between chokes and other inductors is that in their cases they do not require high Q factor construction techniques meant to reduce the resistance in inductors found in tuned circuits.

4.5 Capacitor

A capacitor is a little like a battery but works completely differently. A battery is an electronic device that converts chemical energy into electrical energy, whereas a capacitor is an electronic component that stores electrostatic energy in an electric field. In this article, let's learn about capacitors in detail.



(Capacitor acting as a source of energy)

(Capacitor symbol)

4.5.1 Capacitance

Capacity of storing electric charge is called electric capacitance. When a charge is given to a conductor its potential increases in the ratio of charge. Increase in potential is directly proportional to the charge given to the conductor. If charge given to conductor is 'Q' and its potential increases by 'V' then

$$V \alpha Q$$

Where 'C' is a constant known as capacitance of conductor

$$C = Q/V$$

Therefore capacitance of a conductor is equal to the ratio of charge given to the conductor to increase in potential

SI unit of Capacitance = Farad (F)

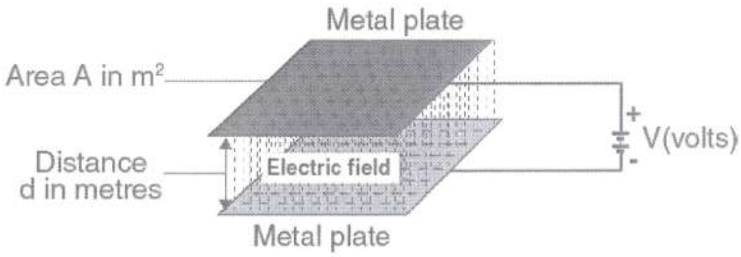
4.5.2 Types of Capacitor

- Parallel Plate Capacitor
- Spherical capacitor
- Cylindrical capacitor

4.5.2.1 Parallel Plate Capacitor

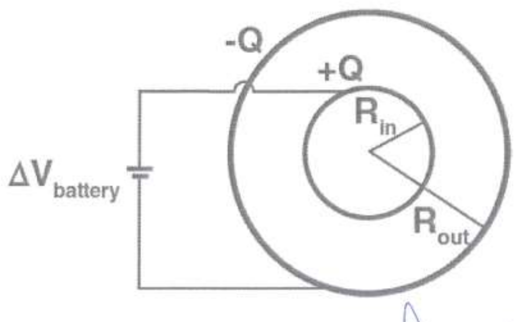
The parallel plate capacitor consists of two metal plates of Area, A and is separated by a distance d. The plate on the top is given a charge +Q and that at the bottom is given the charge –Q. A potential difference of V is developed between the plates.

The separation is very small compared to the dimensions of the plate so that the effect of bending outward of electric field lines at the edges and the non-uniformity of surface charge density at the edges can be ignored.



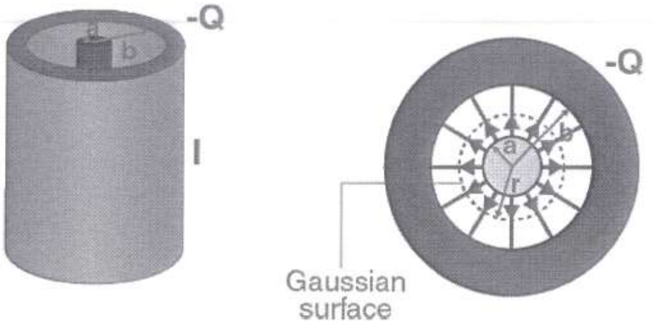
4.5.2.1 Spherical Capacitor

Let's consider a spherical capacitor that consists of two concentric spherical shells. Suppose the radius of the inner sphere, R_{in} = a and radius of the outer sphere, R_{out} = b. The inner shell is given a positive charge +Q and the outer shell is given –Q.



4.5.2.3 Cylindrical Capacitor

Consider a solid cylinder of radius, a surrounded by a cylindrical shell, b. The length of the cylinder is l and is much larger than a-b to avoid edge effects. The capacitor is charged so that the charge on the inner cylinder is +Q and the outer cylinder is –Q.



4.6. Filters

A frequency filter or also known as a frequency selective circuit is a special type of a circuit, which is used for filtering out some of the input signals on the basis of their frequencies.

A filter circuit passes some frequency signal's without any attenuation (Reduction in amplitude) or with some amplification, & attenuate other frequency depending on the types of the filter.

4.7 LC Filters:-

LC filters refer to circuits consisting of a combination of inductors (L) and capacitors (C) to cut or pass specific frequency bands of an electric signal. Capacitors block DC currents but pass AC more easily at higher frequencies. Conversely, inductors pass DC currents as they are, but pass AC less easily at higher frequencies. In other words, capacitors and inductors are passive components with completely opposite properties.

By combining these components with opposite properties, noise can be cut and specific signals can be identified.

LC filters are broadly classified into three types-

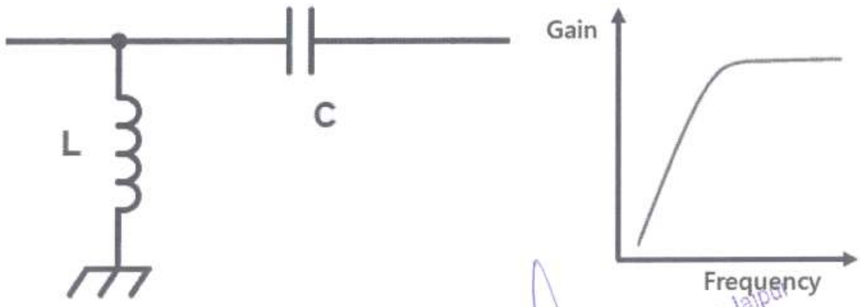
Low-pass Filters(LPF)

Low-pass filters are filter circuits that pass DC and low-frequency signals and cut high-frequency signals. They are the most widely used filter circuits and are mainly used to cut high-frequency noise. In audio, they are also used to cut treble/mid-range sound components of bass speak

(Low-pass Filters circuit and Gain Vs Frequency graph)

6.1.2 High-pass Filters (HPF)

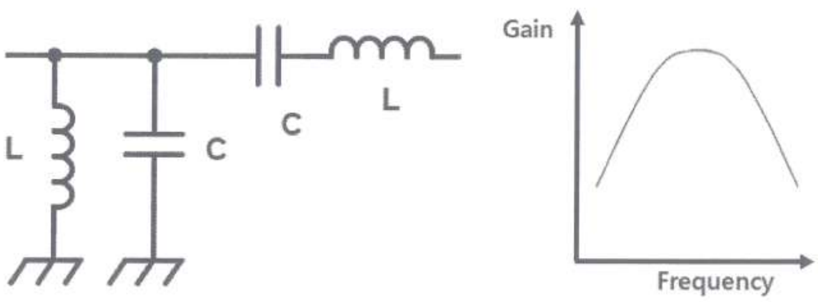
High-pass filters are filter circuits that cut DC and low-frequency signals and pass high-frequency signals. They are used to cut low-frequency noise in the audible range, cut mid-range/bass sound components of treble speakers, etc.



(High-pass Filters circuit and Gain Vs Frequency graph)

Band-pass Filters (BPF)

Band-pass filters are filter circuits that pass only signals at a specific frequency and cut signals at other frequencies. They are used for radio tuning (frequency adjustment) or for cutting the bass/treble sound components of mid-range speakers, etc.



(Band-pass Filters circuit and Gain Vs Frequency graph)

4.7. Summary

A rectifier is an electronic device that converts an alternating current into a direct current by using one or more P-N junction diodes.

The type of rectifier that converts only the half cycle of the alternating current into the direct current is known as a half-wave rectifier.

The capacitor is a device in which electrical energy can be stored. It is an arrangement of two-conductor generally carrying charges of equal magnitudes and opposite sign and separated by an insulating medium.

A frequency filter or also known as a frequency selective circuit is a special type of a circuit, which is used for filtering out some of the input signals on the basis of their frequencies

4.9 Keywords

- Choke - A choke is a type of inductor that is used mainly for blocking high-frequency alternating current (AC) in an electrical circuit.
- Power Rating – It denotes the maximum power the Zener diode can dissipate. It is given by the product of the voltage of the diode and the current flowing through it.
- Temperature Stability – Diodes around 5 V have the best stability
- Voltage Tolerance – It is typically ±5%
- Zener Resistance (Rz) – It is the resistance to the Zener diode exhibits.

4.10 Review Questions

1. What is self-induction?
2. What is mutual induction?
3. When is emf induced in a circuit?
4. Define the capacitance of a capacitor?
5. How can the capacitance of the parallel plate capacitor be increased?
6. Does the capacitance of the capacitor change when the charge on it is doubled?
7. What happens to the charge on the parallel plate capacitor when the potential difference between the plates is doubled?
8. What is a rectifier?
9. What are the types of rectifiers?
10. What is Filter and types of filters?

4.11. References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education
4. Millman, Electronics Devices and Circuits, TMH
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India

Table of Content

Learning Objectives

After studying this unit the student will be able to:

1 Describe transistor and its modes of operation.

2 Explain the concepts and Characteristics, Current components of transistor.

3 Discuss Current gains: alpha, beta and gamma.

4 Explain CE, CB and CC configuration

Introduction

The PN junction can be use to make various electronic devices such as transistors, FET, MOSFETs. Transistors are also used in various electronic circuits as filters.

In N-P-N transistor both emitter & collector are N types of semiconductor & base B is of P type. In PNP transistor both emitter & collector are P types of semiconductor & base B is of N type.

In the symbol of junction transistor shown in figure, arrow indicates the following -

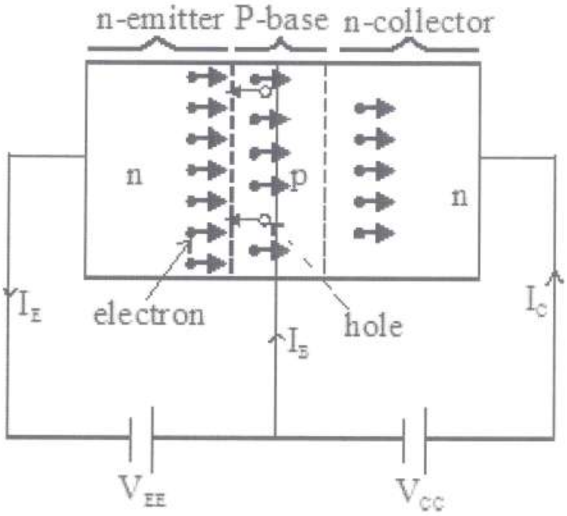
(a) Position of emitter (b) direction of current (c) Type of transistor

In the NPN transistor electrons are majority charge carriers which flow from emitter to base i.e. the conventional current flow from base to emitter as shown in figure.

In the same way holes are majority charge carrier in PNP transistor which flow from emitter to base therefore conventional current from emitter to base.

5.1 Operation of transistor

For the active operation of transistor its emitter- base junction is in forward bias & collector - base junction is in reverse bias. The biasing of NPN junction is shown in figure.



(Biasing of NPN transistor)

Due to forward biasing of emitter - base junction of NPN transistor, the effect of depletion layer is reduced. Electrons which are majority charge carrier in emitter (N type semiconductor) are repelled towards base by negative potential of V_{EE} on emitter, resulting current I_e.

The base being thin & lightly doped (P type) has low density of hole. When the electron enter the base region , the only a few hole get neutralized by electron-hole combination , resulting base current I_b.The remaining electrons pass over to the collector , on account of high positive potential of collector to battery V_{CC} , resulting in current I_C.

As one electron reaches the collector, it flows to the positive terminal of battery V_{CC} through connecting wire. At the same time, one electron flow from negative terminal of V_{CC} to positive terminal of V_{EE} and one electron flow from negative terminal of V_{EE} to emitter. When the electron coming from emitter combines with the hole in base, the deficiency of hole in base is compensated by breakage of covalent bond there. The electron so released flows to the positive terminal of battery V_{EE}, through connecting wire.

Thus

$$I_e = I_b + I_c$$

In transistor base current is very small as compared to collector current. Therefore

$$I_c \cong I_e \qquad (\because I_b \cong 0)$$
$$\text{i.e } I_c \le I_e$$

5.2 Configuration of Bipolar junction transistor

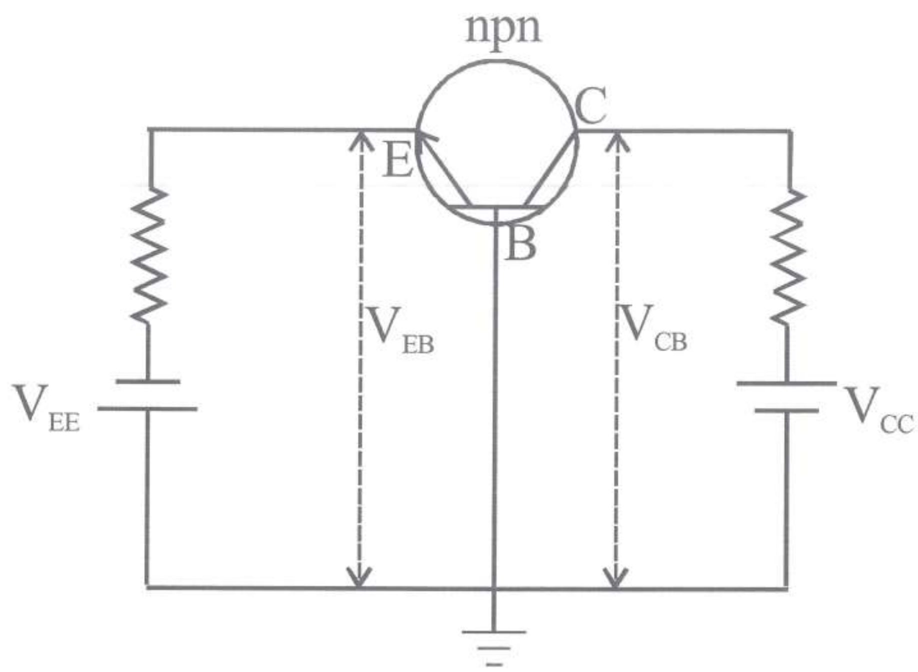
Bipolar Junction transistor is three terminal device but usually it is as four terminal device as one of its three terminal (E, B, C) is used as common terminal for input & output circuit. So there are three basic configurations in which a transistor can be connected in circuit.

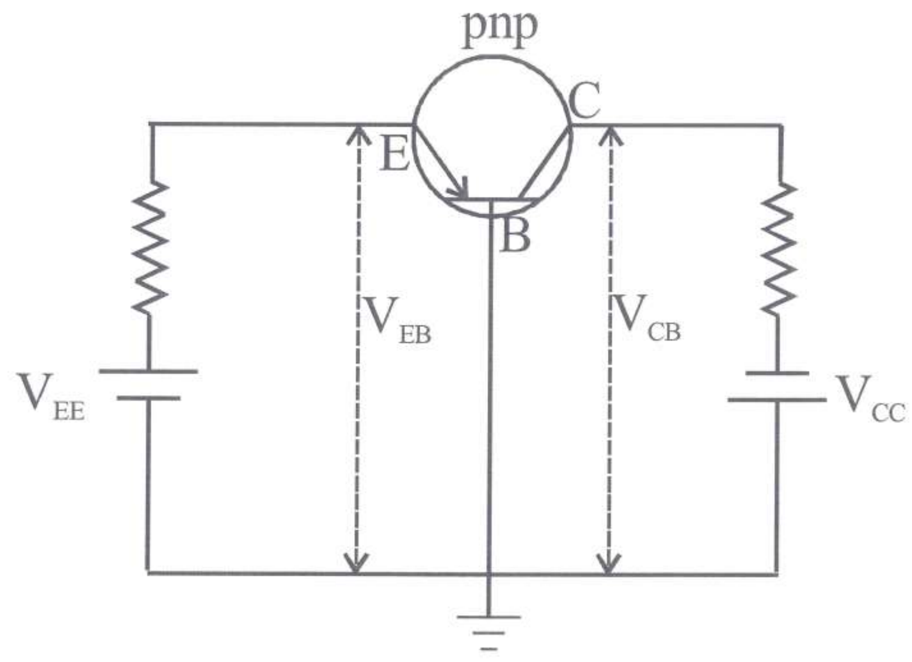
(a) Common base configuration

(b) Common emitter configuration

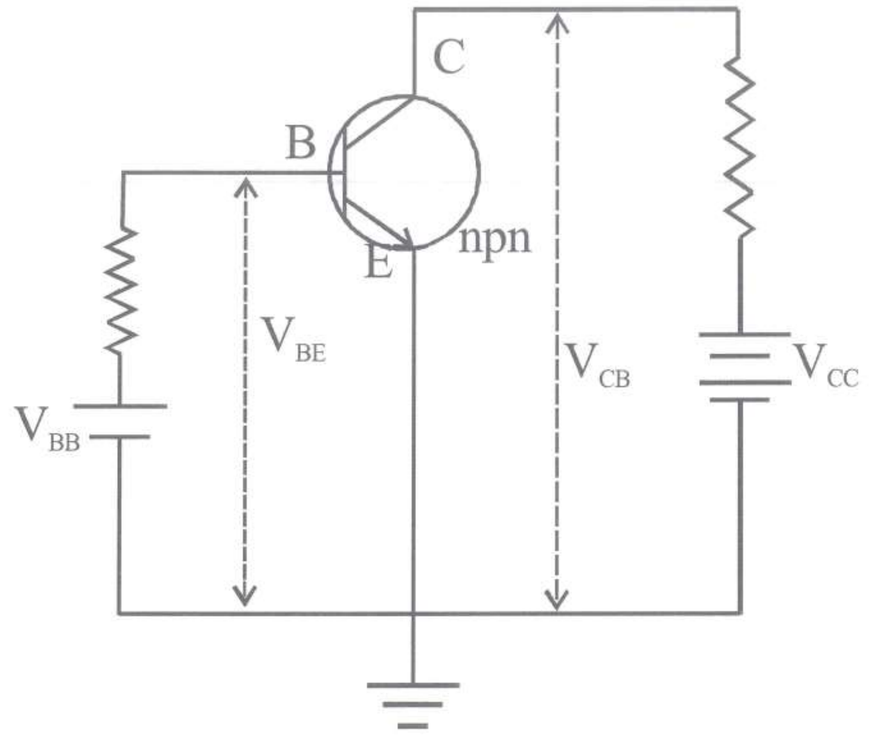
(c) Common collector configuration

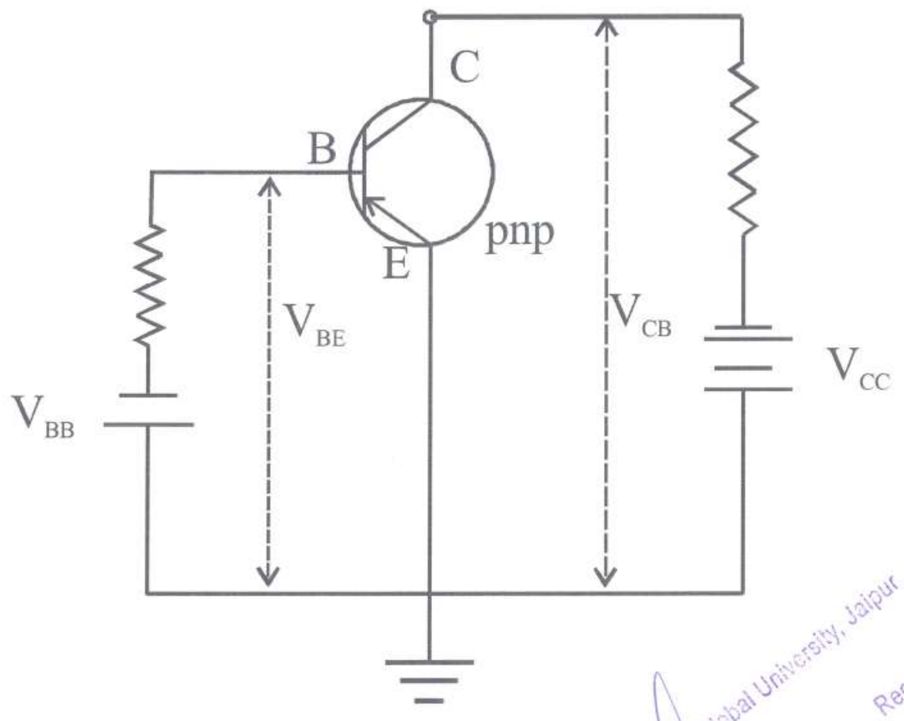
5.2.1 Common base configuration



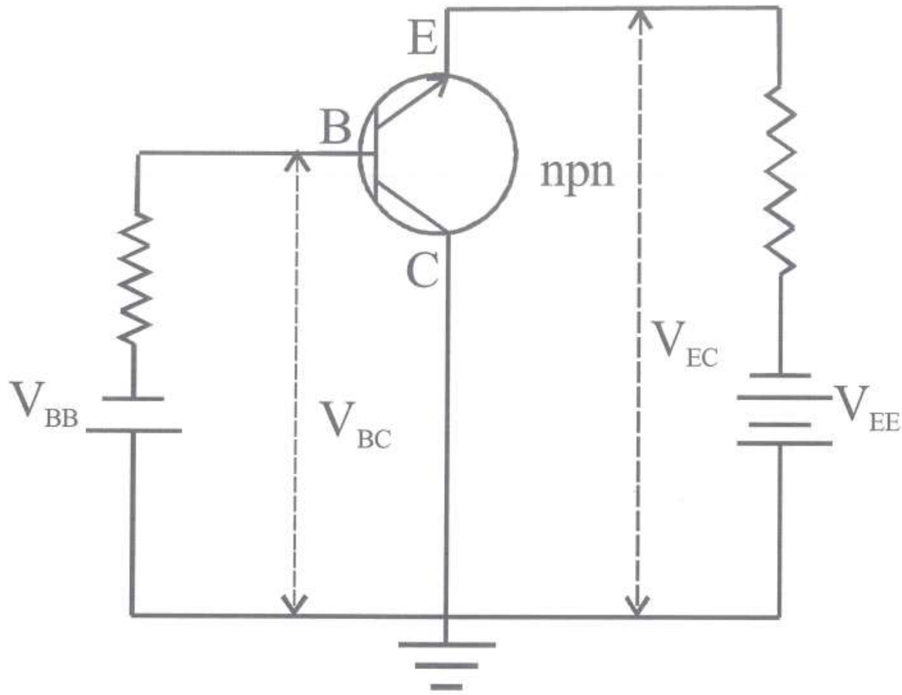


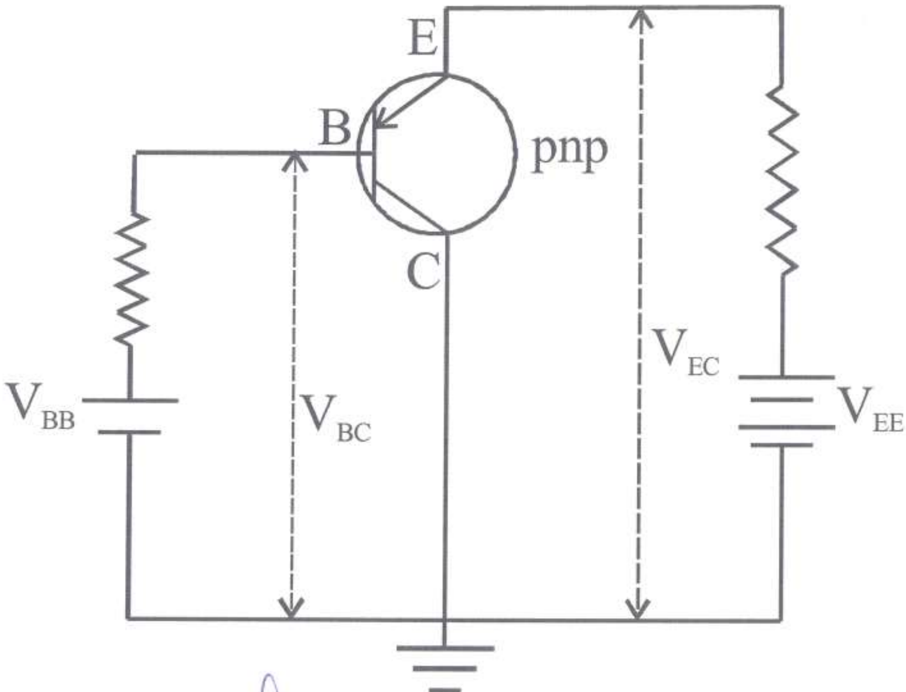
5.2.2 Common emitter configuration





5.2.3 Common collector configuration





5.3 Characteristic curve for transistor

There are two type of characteristic curve

(a) Input characteristic curve (b) Output characteristic curve

5.3.1 Input characteristic curve

The curve between Input current & input voltage keeping output voltage as constant

1.3.2 Output characteristic curve

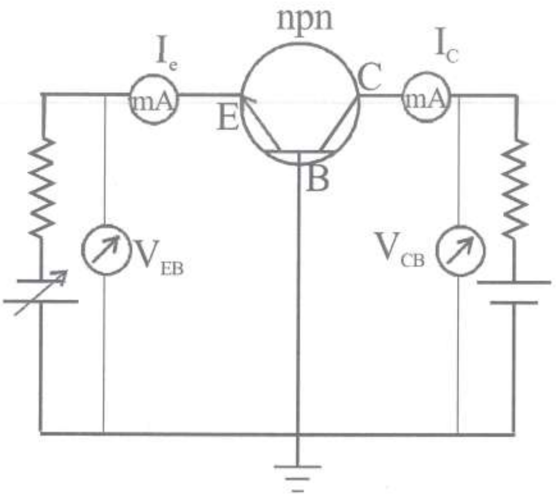
The curve between output current & output voltage keeping input current as constant.

5.4 Characteristic curve for common base configuration

In this configuration the input current is emitter current I_b & input voltage is emitter -base voltage V_{BE}. The output current is I_C & output voltage is collector- base voltage V_{CB}.

To study the characteristic of common base transistor using NPN transistor, we complete the circuit as shown in figure. Here emitter-base circuit is biased in forward direction with battery V_{EE} & collector base circuit is biased in reverse direction with battery V_{CC}. Emitter voltage & emitter current can be studied by VEB and ammeter I_e respectively.

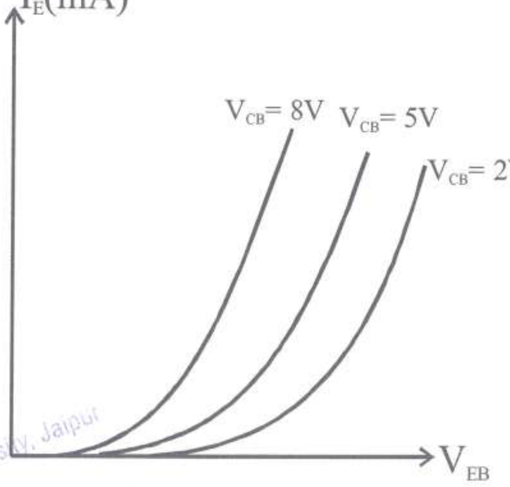
Whereas the collector voltage & collector current by voltmeter V_{CB} & ammeter I_C respectively.



Input characteristic curve -

To get input characteristic curve, apply a suitable constant voltage on collector & by applying the various value of emitter voltage note the corresponding values of emitter current. Repeat the experiment for various constant collector voltages.

Plot a graph between emitter voltage & emitter current, we get curves of the type shown in figure, called input characteristics of common base transistor.



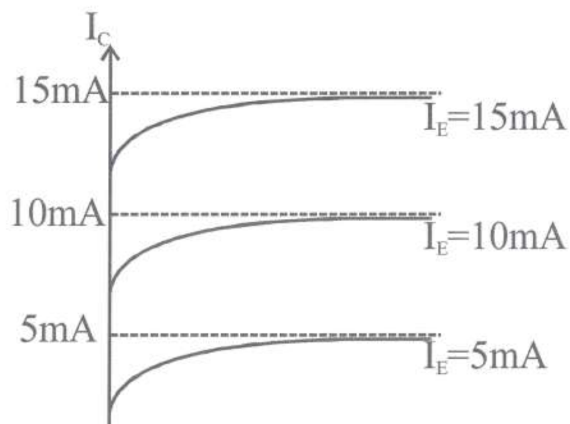
From the graph, it can be concluded that

(i) for a given collector voltage , the emitter current increases rapidly with increasing values of emitter base voltage .It means input resistance is very small.

(ii) For higher collector voltage, the emitter current rises more rapidly with emitter voltage.

Output characteristic curve -

To get the output characteristic curve , adjust suitable value of emitter current & applying the various values of collector voltage , note the corresponding values of collector current. Repeat the experiment for the various constant emitter currents. Plot a graph between collector voltage V_{CB} & collector current I_C .We get the curves of the type shown in figure , called output characteristics curve of common base transistor.



From the graph, it can be concluded that

(i) For a given of emitter current, the collector current is not zero when collector voltage is zero.

(ii) For a given emitter current, there is a rapid increase in the collector current for an increase in low negative collector voltage. This shows the region of low collector resistance. The transistor is never operated in this region.

(iii) For a given emitter current, the collector current becomes saturated for a certain collector voltage shown by horizontal line. Beyond this there is no change in collector current for further increase in negative collector voltage. This indicates a region of high collector resistance. This means output resistance is very high.

Current amplification factor for common base transistor: –

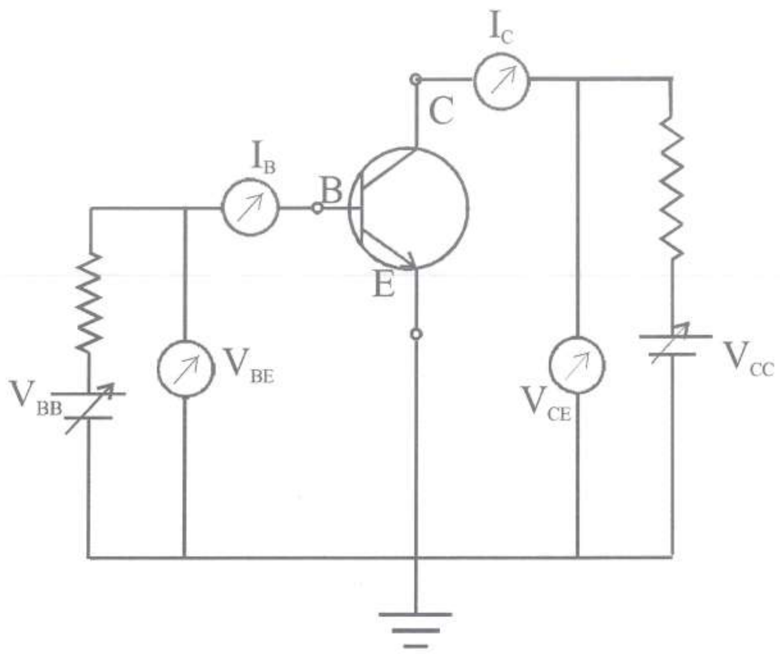
Current amplification factor in the ratio of output current & input current when output voltage is constant. In common base Configuration input current is I_E & output current is Ie current amplification factor for common base configuration is represented by a

$$\alpha = \left(\frac{I_C}{Te} \right)_{V_{CB} \rightarrow \text{Const ant}}$$

Since collector current is less than emitter current therefore, current amplification factor for common base on fig. is always less than one (a < 1) the value of a can never be greater than one

5.5 Characteristic Curve for common – emitter configuration

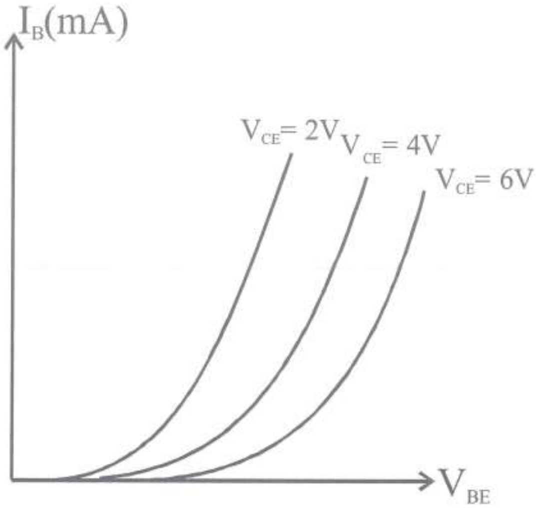
The common emitter on fig. N — P – N transistor is shown in fig. In this fig, the emitter terminal is common in input & output circuit this on fig. is also called grounded emitter on fig.



Input characteristic curve

In common emitter configuration the input current is base current I_B & input voltage is base emitter voltage V_{BE}. Output current is collector current I_C & output voltage is V_{CE}.

The graph between input current I_B & input voltage V_{BE} for constant value of output voltage V_{CE} is called input characteristic curve. The input characteristic curve for N -P-N transistor is shown in figure (a).



(a)

It is clear that from these characteristics that these are identical to the characteristic curve of forward biased PN junction diode. For a constant value of V_{BE} if V_{CE} is increased, the value of current I_B decreases.

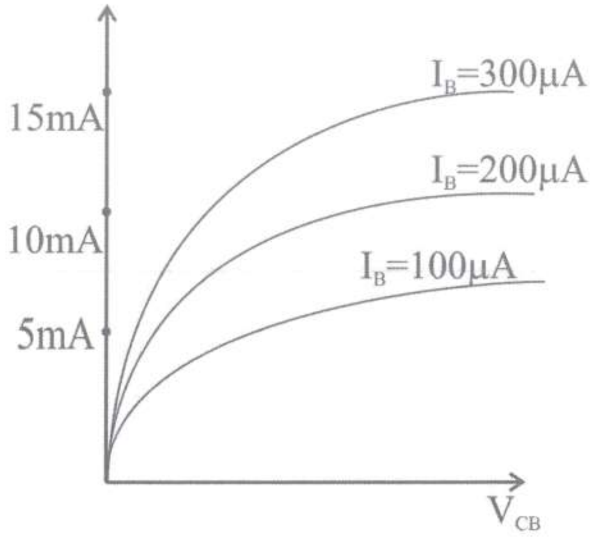
Input Dynamic resistance R_{in} =

$$\left(\frac{\Delta V_{BE}}{\Delta I_B}\right)_{V_{CE} \to \text{Constant}}$$

Output characteristic curve

In common emitter on fig the collector current I_E is Output current & collector emitter voltage V_{CE} is output Voltage

The graph between output current I_C & output voltage V_{CE} for a constant value of input current I_B is called output characteristic curve. The output characteristic curve for N -P-N transistor is shown in figure.



Output Dynamic resistance: – R_{out} =

$$\left(\frac{\Delta V_{CE}}{\Delta I_C}\right)_{I_B \to \text{Constant}}$$

Current Amplification factor for common emitter configuration

It is the ratio of output current I_C& input current I_B It is represented by b=

$$\left(\frac{I_C}{I_B}\right) V_{CE} \to \text{cons tan t}$$

$$I_C >> I_B$$

$$b >> 1$$

Relation between a and b

$$\text{Since } I_E = I_C + I_B$$

Dividing by I_C

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

$$\frac{1}{\alpha} = 1 + \frac{1}{\beta} = \frac{\beta + 1}{\beta}$$

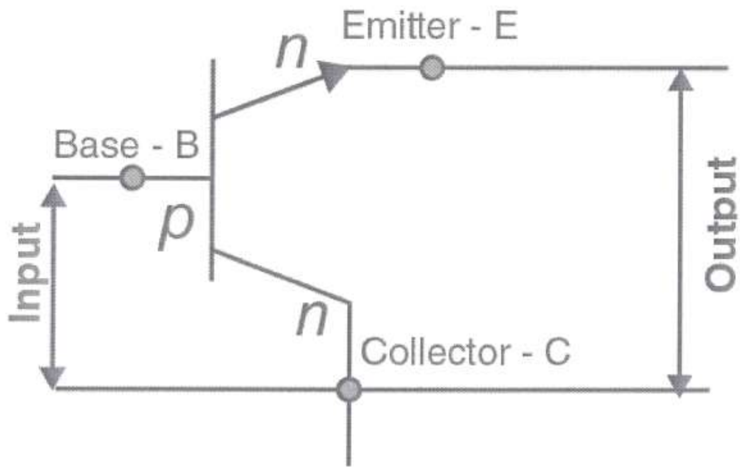
$$\alpha = \frac{\beta}{1+\beta}$$

$$a + a\,b = b$$

$$a = b\,(1 - a)$$

$$\boxed{\beta = \frac{\alpha}{1-\alpha}}$$

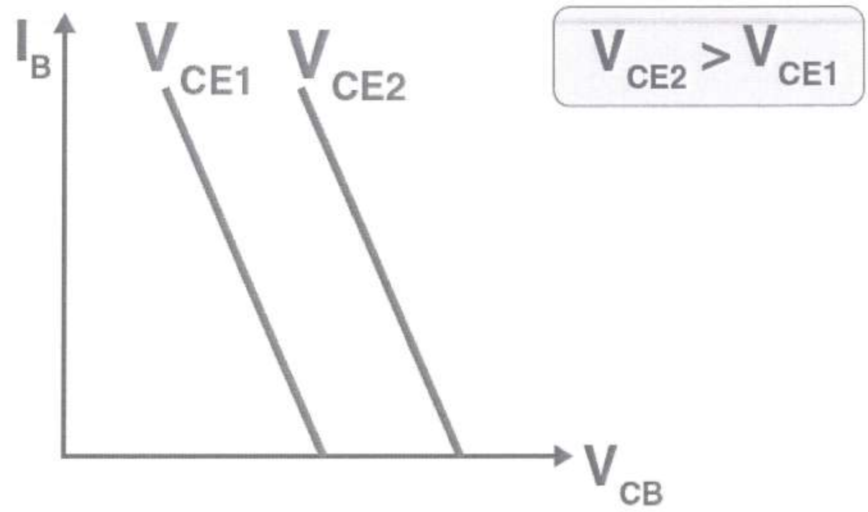
5.6 Characteristic Curve for Common Collector configuration



(Common Collector (CC) Configuration)

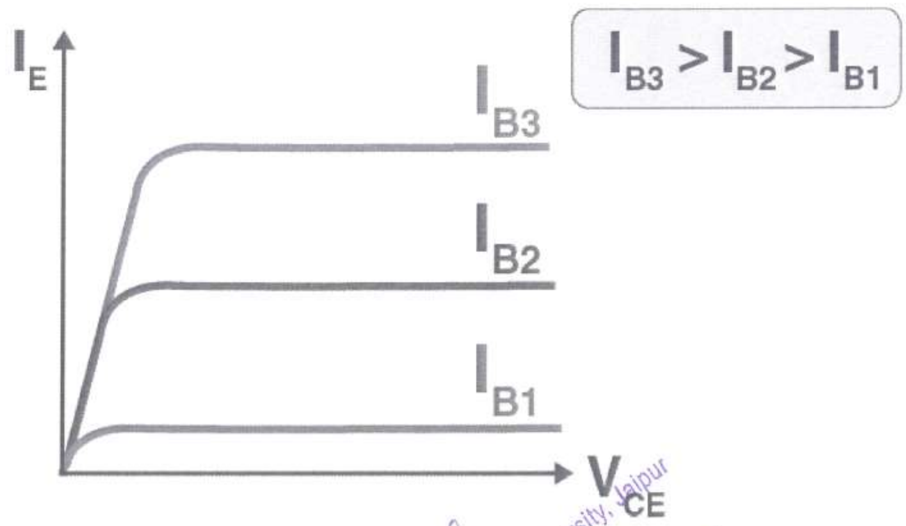
Input Characteristics

The variation of emitter current (I_B) with Collector-Base voltage (V_{CB}), keeping Collector Base voltage (V_{CB}) constant.



Output Characteristics

The variation of emitter current (I_E) with Collector-Emitter voltage (V_{CE}), keeping the base current (I_B) constant.



Current Amplifier Factor for common Collector configuration (γ)

The current amplification factor is defined as the ratio of the output current to the input current. In common emitter configuration, the output current is emitter current I_E, whereas the input current is base current I_B.

Thus, the ratio of change in emitter current to the change in base current is known as the current amplification factor. It is expressed by the γ.

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

Relation Between γ and α

The γ is the current amplification factor of common collector configuration and the α is current amplification factor of common base connection.

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

$$\alpha = \frac{\Delta I_E}{\Delta I_B}$$

and,

$$I_E = I_C + I_B$$

$$\Delta I_E = \Delta I_C + \Delta I_B$$

$$\Delta I_B = \Delta I_E - \Delta I_C$$

Substituting the value of ΔI_B in above first equation, we get,

$$\gamma = \frac{\Delta I_E}{\Delta I_E - \Delta I_C}$$

$$\gamma = \frac{\Delta I_E/\Delta I_E}{\Delta I_E/\Delta I_E - \Delta I_C/\Delta I_E}$$

$$\gamma = \frac{1}{1 - \Delta I_C/\Delta I_E}$$

$$\gamma = \frac{1}{1-\alpha}$$

The above relation shows that the value of Y is nearly equal to β. This circuit is mainly used for amplification because of this arrangement input resistance is high, and output resistance is very low. The voltage gain of the resistance is very low. This circuit arrangement is mainly used for impedance matching.

Collector Current

We know that,

$$I_C = \alpha I_\beta + I_{CBO}$$

$$I_E = I_C + I_B = (\alpha I_\beta + I_{CBO}) + I_B$$

$$I_E(1-\alpha) = I_\beta + I_{CBO}$$

$$I_E = I_B\left(\frac{1}{1-\alpha}\right) + I_{CBO}\left(\frac{1}{1-\alpha}\right)$$

$$= (\beta + 1)I_B + (\beta + 1)I_{CBO}$$

5.7 Summary

- Transistor is a Solid-state semiconductor device for amplifying, controlling, and generating electrical signals.
- It displaced the vacuum tube in many applications.
- Transistors consist of layers of different semiconductors produced by addition of impurities (such as arsenic or boron) to silicon.
- These impurities affect the way electric current moves through the silicon.
- Transistors were pivotal in the advancement of electronics because of their small size, low power requirements, low heat generation, modest cost, reliability, and speed of operation.
- Single transistors were superseded in the 1960s and 70s by integrated circuits; present-day computer chips contain billions of transistors.
- Today transistors perform many different functions in nearly every type of electronic equipment.

5.8 Keywords

- **Emitter –** This segment is on the left side of the transistor. It is moderately sized and heavily doped.
- **Base –** This segment is at the center of the transistor. It is thin and lightly doped.
- **Collector –** This segment is on the right side of the transistor. It is larger than the emitter and is moderately doped.

5.9 Review Questions

1. Which region of the transistor is lightly doped?
2. How is the emitter region of the transistor different from the collector region?

3. What is the normal biasing of the diodes of the transistor?
4. How many depletion regions does a transistor have?
5. What is a transistor?
6. What are the three terminals of the transistor?
7. Applications of BJT?
8. Which region in BJT transistors works as an amplifier?
9. What are the current gains in transistor?
10. What is CC configuration?

5.10 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education
4. Millman, Electronics Devices and Circuits, TMH
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India

Contents

Learning Objectives

After studying this unit the student will be able to:

1 Describe transistor, JFET, MOSFETs and its modes of operation.
2 Explain the concepts and Characteristics, Current components of FETs.
3 Discuss applications of FETs.
4 Explain Transistor as an amplifier.

Introduction

The PN junction can be use to make various electronic devices such as transistors, FET, MOSFETs. Transistors are also used in various electronic circuits as filters.

6.1 Transistor as an amplifier

The process by which electrical signal is increased is called amplification and the device by which amplification is done is called amplifier.

Electric current and voltage can be amplified.

There are two types of gain

(a) Current gain or Current amplification factor

$$\text{Current gain} = \frac{\text{output current}}{\text{input current}}$$

(b) Voltage gain or voltage amplification factor

$$\text{Voltage gain} = \frac{\text{output voltage}}{\text{input voltage}}$$

6.1.1 Common – emitter transistor as amplifiers

Circuit of n-p-n transistor as a common emitter amplifier is shown in figure. Emitter base is forward bias with battery V_{BB}. and collector base circuit is reversed biased with battery V_{CC}. Due to this the resistance of input circuit is low & that of output circuit is high. R_C is load resistance connected in collector circuit.

To operate the transistor as an amplifier it is necessary to fix its operating point somewhere in the middle of its active region. If we fix the value of V_{BB} corresponding to a point in the middle of linear part of the transfer curve then the dc base current I_B would be constant & corresponding collector current I_C will also be constant.

The dc voltage $V_{CE}=V_{CC} - I_C R_C$

would also remain constant.

The operating values of V_{CE} & I_B determine the operating point of the amplifier.

If a small sinusoidal voltage is superimposed on the dc base by connecting the source of that signal in series with the supply V_{BB}, then the base current will have sinusoidal variations superimposed on the value of I_B.

As a consequence the collector current also will have sinusoidal variations superimposed on the value of I_C producing in turn corresponding change in the value of V_{CE}. Let us superimposed an ac signal V_i (to be amplify) on bias V_{BB}. The output is taken between the collector and the ground.

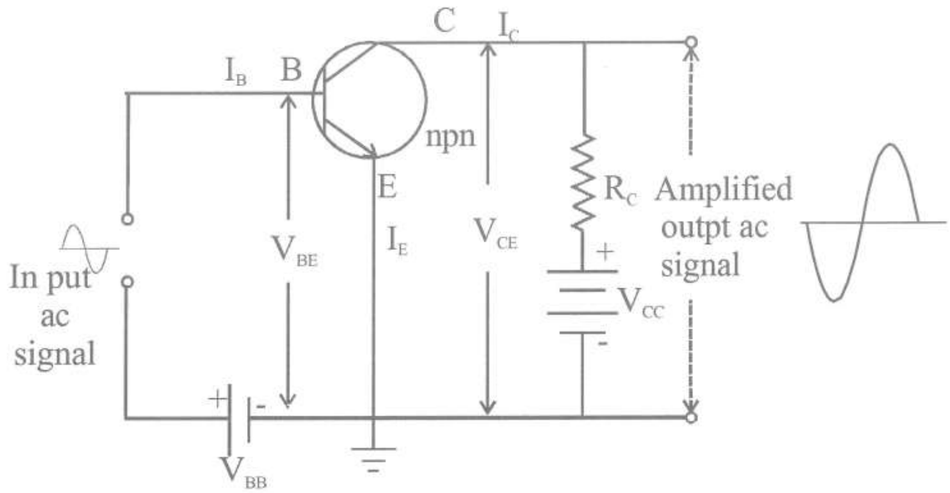


Figure 1: Transistor as an amplifier Circuit

The working of an amplifier can be easily understood. If we first assume that $V_i = 0$.

Input loop gives $V_{BB} = V_{BE} + I_B R_B$

Applying Kirchhoff's law to the output loop we get

$$V_{CC} = V_{CE} + I_C R_C$$

When V_i is not zero

$$V_{BB} + V_i = V_{BE} + I_B R_B + \Delta V_{BE} + \Delta I_B R_B$$

$$\text{As } \Delta V_{BE} = r_i \, \Delta I_B$$

$$\text{So } V_i = \Delta I_B (R_B + r_i) = \Delta I_B \, r$$

Change in I_B cause a change in I_C, we define a parameter β_{ac}

$$\beta_{ac} = \frac{\Delta I_C}{\Delta I_B} = \frac{i_c}{i_b}$$

A field-effect transistor's terminals have applied a voltage through which conductivity is regulated. The voltage that was applied to the gate creates an electric field in the device which causes repulsion and attraction to charges that are carried amid the two terminals.

The conductivity is also affected due to the density of those charge carriers.

In this topic, we will explore the principles behind FETs, their different forms (JFETS, MOSFET), advantages and disadvantages, and their various applications.

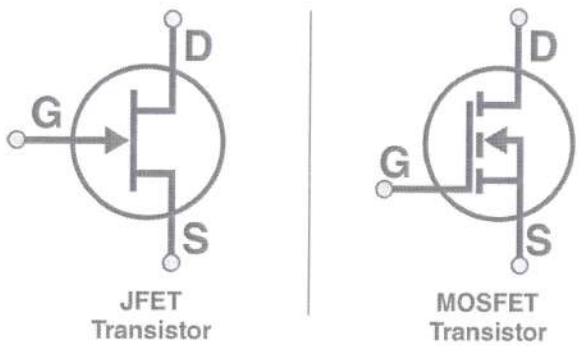


Figure 2: Junction Field Effect Transistor symbol

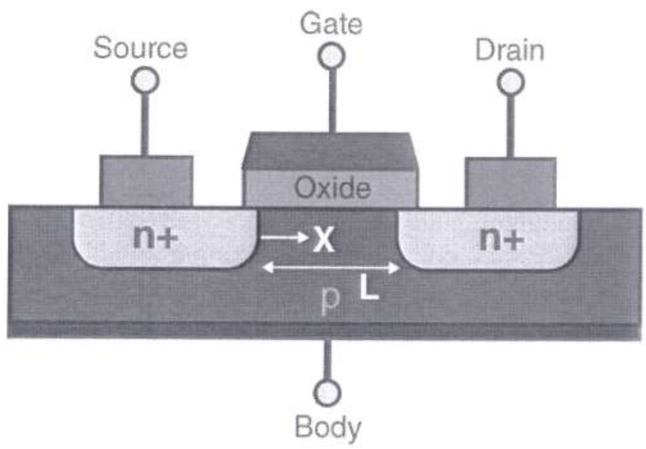


Figure 3: Field Effect Transistor structure

6.2.2 Principles of Operation

The basic principle behind a FET is the creation of a channel between two semiconductor regions by applying a voltage to a gate electrode. This channel allows current to flow between the sources and drain electrodes, and the width of the channel can be controlled by adjusting the voltage applied to the gate.

One important feature of FETs is their high input impedance, which means that they require very little current to control the flow of current through the device. This makes them particularly useful in applications where low power consumption is important, such as in battery-powered devices.

6.2.3 Different forms of FETs

There are several different forms of FETs, including Metal-Oxide-Semiconductor FETs (MOSFETs), Junction FETs (JFETs), and Insulated Gate Bipolar Transistors (IGBTs).

- **Junction Field Effect Transistor (JFET)**

JFET or Junction Field Effect Transistor is one of the simplest types of field-effect transistor. Contrary to the Bipolar Junction Transistor, JFETs are voltage-controlled devices. In JFET, the current flow is due to the majority of charge carriers. However, in BJTs, the current flow is due to both minority and majority charge carriers. Since only the majority of charge carriers are responsible for the current flow, JFETs are unidirectional.

- *JFET Construction*

In an N-channel JFET, the material is of P-type, and the substrate is N-type, while in a P channel JFET the material is of N-type, and the substrate used is p-type. JFET is made of a long channel of semiconductor material. Ohmic contacts are provided at each end of the semiconductor channels to form source and drain connections. A P-type JFET contains many positive charges, and if the JFET contains a large number of electrons, it is called an N-type JFET.

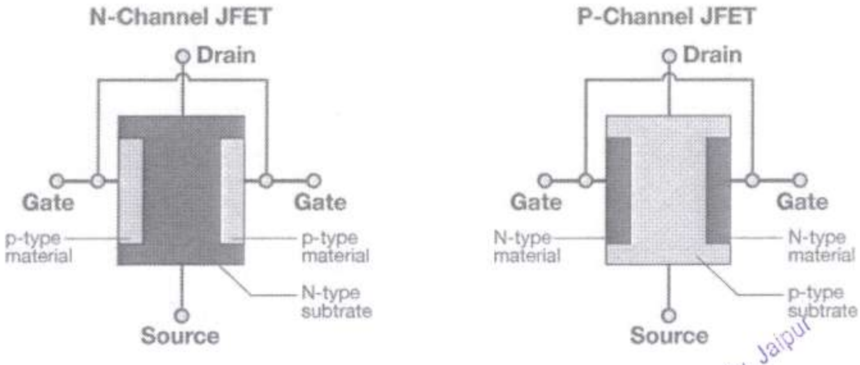


Figure 4: Junction Field Effect Transistor structure

- *JFET Operation*

Let us understand the working of JFET by comparing it to a garden hose pipe. Water flows smoothly through a garden hose pipe if there is no obstruction, but if we squeeze the pipe slightly, the water flow slows down. This is precisely how a JFET works. Here the hose is analogous to JFET, and the water flow is equivalent to a current. By constructing the current carrying-channel according to our needs, we could control the current flow.

When no voltage is applied across the source and gate, the channel is a smooth path for the electrons to flow through. When the polarity that makes the P-N junction reverse biased is applied, the channel narrows by increasing the depletion layer and could put the JFET in the cut-off or pinch-off region.

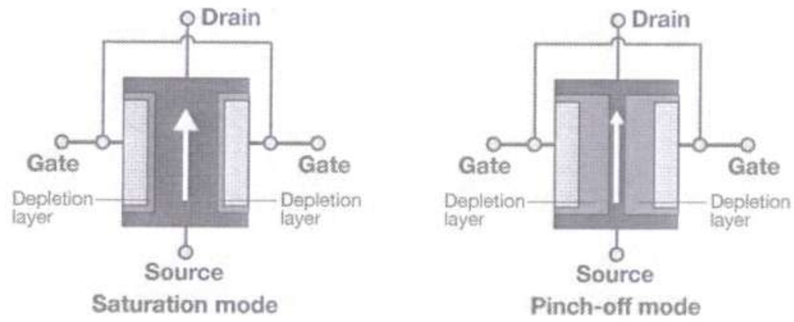


Figure 5: Junction Field Effect Transistor mode of operation

The image shows the depletion region becoming wider and narrower during the saturation and the pinch-off mode.

- *Types of JFET –*

There are two types of JFETs commonly used in the field semiconductor devices: **N-Channel JFET** and **P-Channel JFET**.

❖ **N-Channel JFET**

It has a thin layer of N type material formed on P type substrate. Following figure shows the crystal structure and schematic symbol of an N-channel JFET. Then the gate is formed on top of the N channel with P type material. At the end of the channel and the gate, lead wires are attached and the substrate has no connection.

When a DC voltage source is connected to the source and the drain leads of a JFET, maximum current will flow through the channel. The same amount of current will flow from the source and the drain terminals. The amount of channel current flow will be determined by the value of V_{DD} and the internal resistance of the channel.

A typical value of source-drain resistance of a JFET is quite a few hundred ohms. It is clear that even when the gate is open full current conduction will take place in the channel. Essentially, the amount of bias voltage applied at ID, controls the flow of current carriers passing through the channel of a JFET. With a small change in gate voltage, JFET can be controlled anywhere between full conduction and cutoff state.

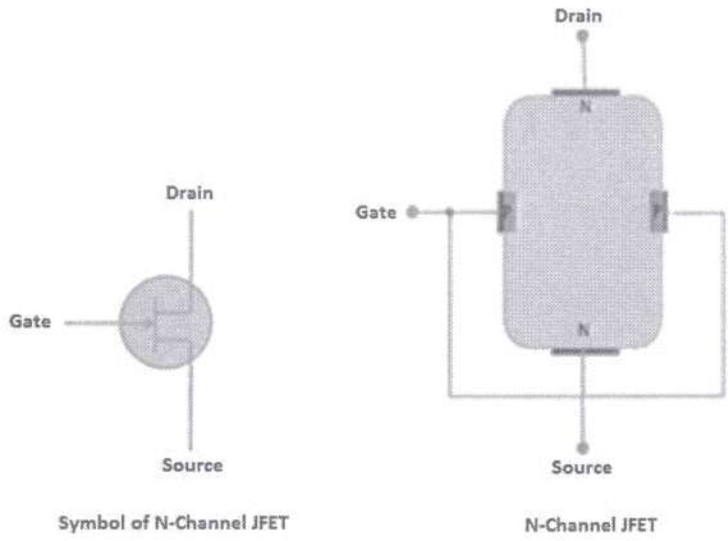
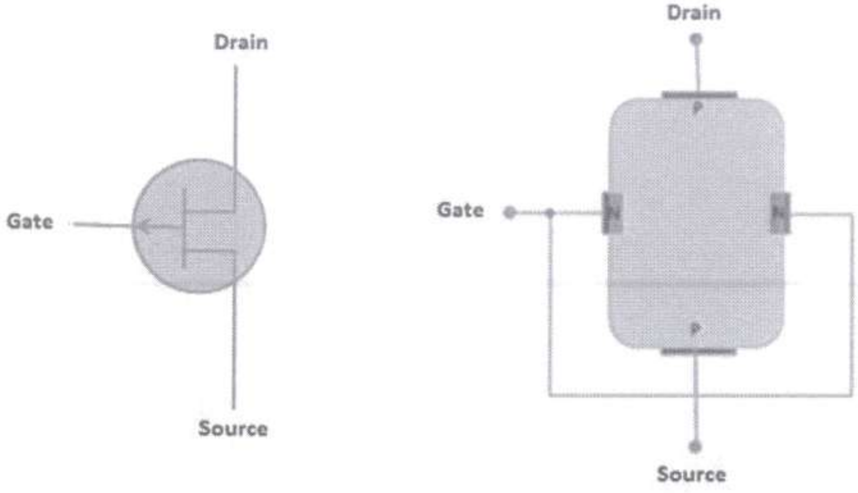


Figure 6: N- Channel JFET

❖ **P-Channel JFET**

It has a thin layer of P type material formed on N type substrate. The following figure shows the crystal structure and schematic symbol of an N-channel JFET. The gate is formed on top of the P channel with N type material. At the end of the channel and the gate, lead wires are attached. Rest of the construction details is similar to that of N- channel JFET



Symbol of P-Channel JFET **P-Channel JFET**

Figure 7: P- Channel JFET

Normally for general operation, the gate terminal is made positive with respect to the source terminal. The size of the P-N junction depletion layer depends upon fluctuations in the values of reverse biased gate voltage. With a small change in gate voltage, JFET can be controlled anywhere between full conduction and cutoff state.

- ***Output Characteristics of JFET***

The output characteristics of JFET are drawn between drain current (I_D) and drain source voltage (V_{DS}) at constant gate source voltage (V_{GS}) as shown in the following figure.

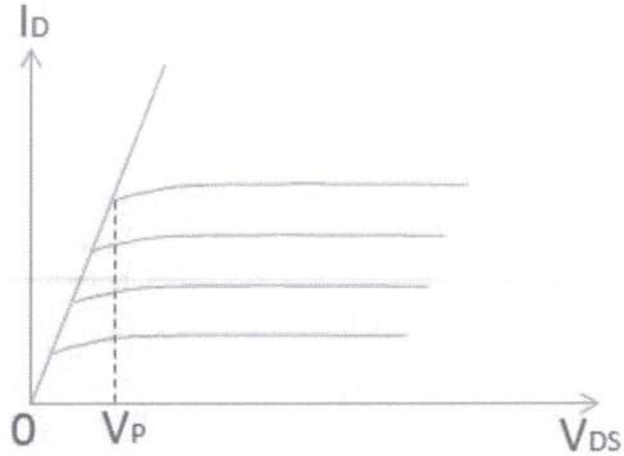


Figure 8: Output Characteristics of JFET

Initially, the drain current (I_D) rises rapidly with drain source voltage (V_{DS}) however suddenly becomes constant at a voltage known as pinch-off voltage (V_P). Above pinch-off voltage, the channel width becomes so narrow that it allows very small drain current to pass through it. Therefore, drain current (I_D) remains constant above pinch-off voltage.

- *Parameters of JFET*

The main parameters of JFET are –

- AC drain resistance
- Trans conductance
- Amplification factor

❖ **AC drain resistance (Rd) –** It is the ratio of change in the drain source voltage (ΔV_{DS}) to the change in drain current (ΔI_D) at constant gate-source voltage. It can be expressed as,

$$R_d = (\Delta V_{DS})/(\Delta I_D) \text{ at Constant } V_{GS}$$

- **Trans conductance (gfs) –** It is the ratio of change in drain current (ΔI_D) to the change in gate source voltage (ΔV_{GS}) at constant drain-source voltage. It can be expressed as,

$$g_{fs} = (\Delta I_D)/(\Delta V_{GS}) \text{ at constant } V_{DS}$$

- **Amplification Factor (u) –** It is the ratio of change in drain-source voltage (ΔV_{DS}) to the change in gate source voltage (ΔV_{GS}) constant drain current (ΔI_D). It can be expressed as,

$$u = (\Delta V_{DS})/(\Delta V_{GS}) \text{ at constant } I_D$$

Metal Oxide Silicon Field Effect Transistors (MOSFET)

MOSFETs or Metal Oxide Silicon Field Effect Transistors were invented to overcome the disadvantages posed by FETs, such as the slow operation, high drain resistance, and moderate input impedance. In this article, let us learn about the basics of MOSFET.

- *MOSFET Basics*

Metal Oxide Silicon Field Effect Transistors commonly known as MOSFETs are electronic devices used to switch or amplify voltages in circuits. It is a voltage controlled device and is constructed by three terminals. The terminals of MOSFET are named as follows:

- Source
- Gate
- Drain
- Body

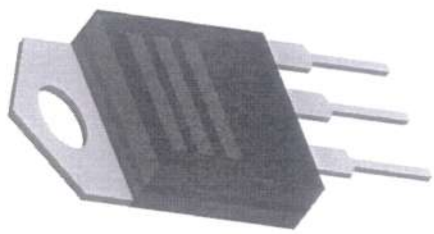


Figure 9: Practical MOSFET

- *MOSFET Construction*

The circuit of MOSFET is typically represented as follows:

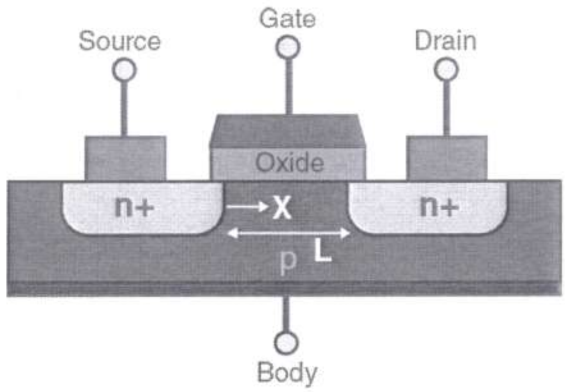


Figure 10: MOSFET Construction

- The p-type semiconductor forms the base of the MOSFET.
- The two types of the base are highly doped with an n-type impurity which is marked as n+ in the diagram.
- From the heavily doped regions of the base, the terminals source and drain originate.
- The layer of the substrate is coated with a layer of silicon dioxide for insulation.
- A thin insulated metallic plate is kept on top of the silicon dioxide and it acts as a capacitor.
- The gate terminal is brought out from the thin metallic plate.
- A DC circuit is then formed by connecting a voltage source between these two n-type regions.

- ***Working Principle of MOSFET***

When voltage is applied to the gate, an electrical field is generated that changes the width of the channel region, where the electrons flow. The wider the channel region, the better conductivity of a device will be.

- ***MOSFET Types***

The classification of MOSFET based on the construction and the material used is given below in the flowchart.

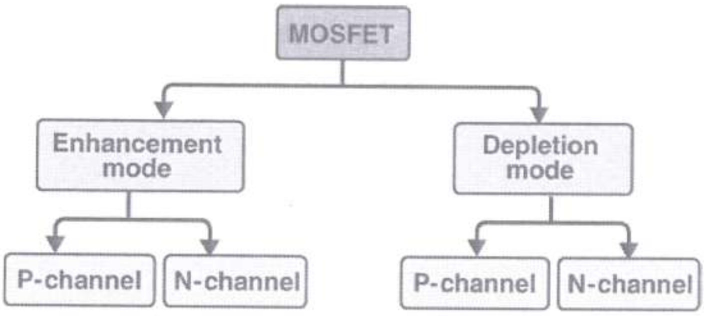


Figure 11: MOSFET Types

MOSFETs are of two classes: *Enhancement mode* and *depletion mode*. Each class is available as n-channel or p-channel; hence overall they tally up to four types of MOSFETs.

❖ **Depletion Mode**

When there is no voltage across the gate terminal, the channel shows maximum conductance. When the voltage across the gate terminal is either positive or negative, then the channel conductivity decreases.

❖ **Enhancement Mode**

When there is no voltage across the gate terminal, then the device does not conduct. When there is the maximum voltage across the gate terminal, then the device shows enhanced conductivity.

The N-channel MOSFETs are abbreviated as NMOS and are symbolically represented as shown in the figure below:

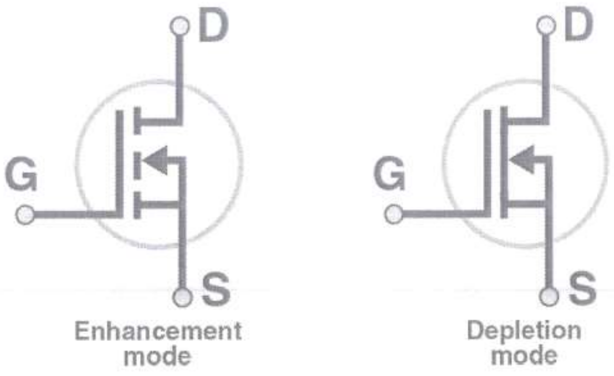


Figure 12: N- channel MOSFET Symbols

Similarly, the P-channel MOSFETs are abbreviated as PMOS and are symbolically represented as follows

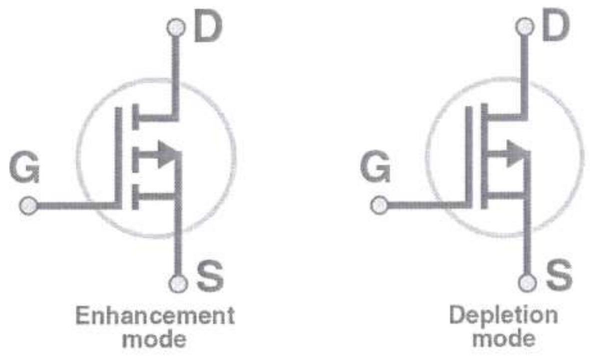


Figure 13: P- channel MOSFET Symbols

- *Operating Regions of MOSFET*

A MOSFET is seen to exhibit three operating regions. Here, we will discuss those regions.

- **Cut-Off Region**

The cut-off region is a region in which there will be no conduction and as a result, the MOSFET will be OFF. In this condition, MOSFET behaves like an open switch.

- **Ohmic Region**

The ohmic region is a region where the current (I_{DS})increases with an increase in the value of V_{DS}. When MOSFETs are made to operate in this region, they are used as amplifiers.

❖ Saturation Region

In the saturation region, the MOSFETs have their I_{DS} constant in spite of an increase in V_{DS} and occurs once V_{DS} exceeds the value of pinch-off voltage V_P. Under this condition, the device will act like a closed switch through which a saturated value of I_{DS} flows. As a result, this operating region is chosen whenever MOSFETs are required to perform switching operations.

- ***MOSFET as a Switch***

MOSFETs are commonly used as switches. The circuit below shows the configuration of MOSFET when it is used as a switch.

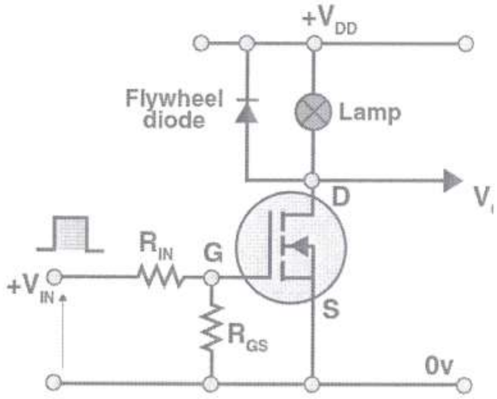


Figure 14: Circuit for MOSFET as a Switch

In the circuit arrangement, an Enhancement-mode N-channel MOSFET is used to switch a simple lamp "ON" and "OFF." The input fate voltage V_{gs} is adjusted to an appropriate positive voltage to switch "ON" the device and the voltage level is set to a negative value or zero to turn it "OFF."

The switching characteristics for both N-channel and P-channel type MOSFET are summarized in the table below:

MOSFET Type	VGS ≪ 0	VGS = 0	VGS ≫ 0
N-channel Enhancement	OFF	OFF	ON
N-channel Depletion	OFF	ON	ON
P-channel Enhancement	ON	OFF	OFF
P-channel Depletion	ON	ON	OFF

- *MOSFET applications*

- Radiofrequency applications use MOSFET amplifiers extensively.
- MOSFET behaves as a passive circuit element.
- Power MOSFETs can be used to regulate DC motors.
- MOSFETs are used in the design of the chopper circuit.

6.2.4 FETs Advantages and Disadvantages

One advantage of FETs is their high switching speed, which makes them useful in applications such as digital logic circuits and switching power supplies. Another advantage is their high input impedance, which allows them to be used in applications where low power consumption is important.

One disadvantage of FETs is their sensitivity to static electricity and other forms of electrical interference. This can cause damage to the device and can be a problem in high-precision applications. Another disadvantage is their relatively low output impedance, which can make them less suitable for use in high-power applications.

6.2.5 FETs Applications

FETs are used in a wide variety of applications, including amplifiers, oscillators, switching circuits, and voltage regulators. MOSFETs are particularly important in digital logic circuits, where their high switching speeds make them well-suited for use in microprocessors and other digital devices. JFETs are commonly used in low-noise amplifiers and voltage-controlled oscillators, while IGBTs are used in high-power applications such as motor control and power electronics.

6.3 Summary

- In conclusion, Field Effect Transistor is an important type of transistor that is widely used in modern electronics.
- Their high input impedance, high switching speed, and low power consumption make them particularly useful in applications where these features are important.
- However, their sensitivity to electrical interference and relatively low output impedance can be disadvantages in some applications.

- Overall, FETs are an important tool in the design and construction of modern electronic devices.
- Transistor is a Solid-state semiconductor device for amplifying, controlling, and generating electrical signals.
- It displaced the vacuum tube in many applications.
- Transistors consist of layers of different semiconductors produced by addition of impurities (such as arsenic or boron) to silicon.

6.4 Keywords

- **Gate** – By using diffusion or alloying technique, both sides of N type bar are heavily doped to create PN junction. These doped regions are called gate (G).
- **Source** – It is the entry point for majority carriers through which they enter into the semiconductor bar.
- **Drain** – It is the exit point for majority carriers through which they leave the semiconductor bar.
- **Channel** – It is the area of N type material through which majority carriers pass from the source to drain.

6.5 Review Questions

1. What is JFET?
2. When was the junction field-effect transistor invented?
3. How many diodes do junction field-effect transistors contain?
4. List a difference between JFET and BJT.
5. What are the important uses of transistors?
6. What is a MOSFET?
7. Can MOSFET conduct in both directions?
8. What is the difference between a MOSFET and a BJT?
9. Which type of transistor is called a unipolar transistor?
10. What are the operating regions of MOSFET?

6.6 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication.
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education.
4. Millman, Electronics Devices and Circuits, TMH.
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi.
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India.

Contents

Learning Objectives

After studying this unit, the student will be able to:

1 Describe Representation of Data.
2 Explain the concepts Digital versus Analog data.
3 Discuss their Digital number system.
4 Explain types of Digital number system.

Introduction

The Data Representation refers to the form in which data is stored, processed, and transmitted. The Data can be represented in both analog and digital format. Representation of digital data has great advantage. There are various types of number system.

Digital systems are electronic systems that use binary signals (0 and 1) to represent and process information. These systems are the backbone of modern communication, computing, and control systems. They are used in a wide range of applications such as personal computers, smartphones, home appliances, medical equipment, and transportation systems.

The fundamental building block of digital systems is the logic gate, which is a device that performs a logical operation on one or more binary inputs to produce a binary output. The most common types of logic gates are AND, OR, NOT, NAND, and NOR gates, which can be combined to create more complex circuits such as adders, multiplexers, and flip-flops.

Digital systems also rely on a clock signal, which synchronizes the operations of the various components in the system. The clock signal ensures that each operation is performed at the right time and in the right sequence.

The performance of digital systems is measured in terms of speed, power consumption, and reliability. With advances in technology, the speed and power efficiency of digital systems have increased significantly, while their size and cost have decreased.

However, digital systems also face challenges such as security threats, data privacy concerns, and the need for continuous updates and maintenance. To address these challenges, digital system designers and engineers need to stay up-to-date with the latest technologies and industry standards.

7.1 Data

Data refers to the symbols that represent people, events, things, and ideas. Data can be a name, a number, the colors in a photograph, or the notes in a musical composition.

7.2 Representation of Data

Data Representation refers to the form in which data is stored, processed, and transmitted. Devices such as smart phones, iPods, and computers store data in digital formats that can be handled by electronic circuitry. Data can be represented in two ways: 1) Analog 2) Digital

7.2.1 Analog representation of data

Analog data is data that is represented in a physical way. Analog data is stored in physical media.

Analog data may also be known as organic data or real-world data. One way to characterize analog data is that it simply exists without being measured.

For an example the analog data is like the actual water surface in motion, which human senses would perceive as the changes to physical motions as well as the color, texture and even smell of the water itself.

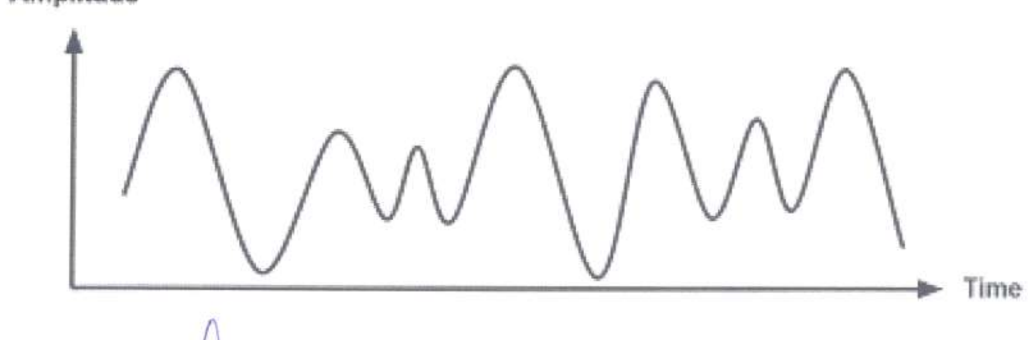


Figure 1: Analog representation of Data

7.2.2 Digital representation of data

A representation of something in the form of digital data i.e., computerized data that is represented using discrete (discontinuous) values to embody information, as contrasted with continuous or analog signals that behave in a continuous manner or represent information using a continuous function.

Digital data is also known as a set of individual symbols.

A digital system can understand positional number system only where there are a few symbols called digits and these symbols represent different values depending on the position they occupy in the number.

A value of each digit in a number can be determined using

- The digit
- The position of the digit in the number
- The base of the number system (where base is defined as the total number of digits available in the number system).

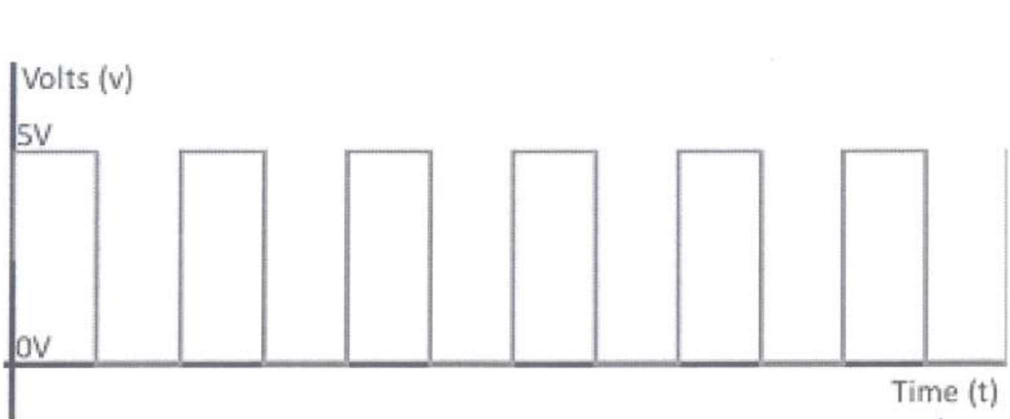


Figure 2: Digital representation of Data

❖ *Data representation on optical media*

In optical devices, the presence of light is interpreted as '1' while its absence is interpreted as '0'. Optical devices use this technology to read or store data.

Take example of a CD-ROM, if the shiny surface is placed under a powerful microscope, the surface is observed to have very tiny holes called **pits**.

The areas that do not have pits are called **land**. The laser beam reflected from the land is interpreted, as 1.

The laser entering the pot is not reflected. This is interpreted as 0. The reflected pattern of light from the rotating disk falls on a receiving photoelectric detector that transforms the patterns into digital form.

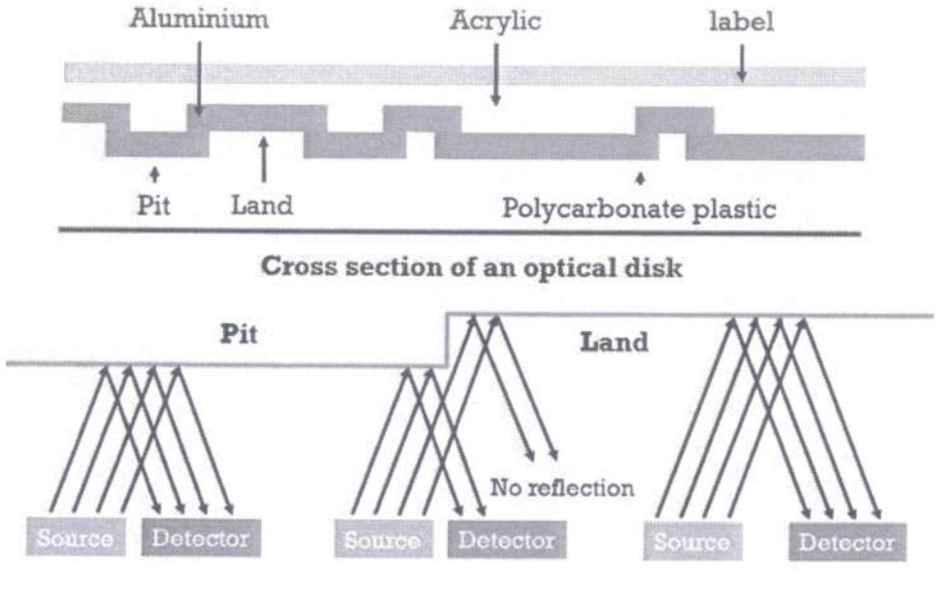


Figure 4: Data representation on optical media

- **Reason for use of binary system in computers**

It has proved difficult to develop devices that can understand natural language directly due to the complexity of natural languages. However, it is easier to construct electric circuits based on the binary or ON and OFF logic.

All forms of data can be represented in binary system format. Other reasons for the use of binary are that digital devices are more reliable, small and use less energy as compared to analog devices.

❖ *Bits, bytes, nibble and word*

- The terms bits, bytes, nibble and word are used widely in reference to computer memory and data size.

- **Bits:** can be defined as either a binary, which can be 0, or 1.It is the basic unit of data or information in digital computers.

- **Byte:** a group of bits (8 bits) used to represent a character. A byte is considered as the basic unit of measuring memory size in computer.

- **A nibble**: is half a byte, which is usually a grouping of 4 bytes.

- **Word:** two or more bits make a word. The term **word length** is used as the measure of the number of bits in each word. For example, a word can have a length of 16 bits, 32 bits, 64 bits etc.

7.2.3 Number systems and their representation

A number system is a set of symbols used to represent values derived from a common base or radix.

The magnitude of a number can be considered using these parameters which are Absolute value, Place value or positional value, Total value, Base value.

- **The absolute** value is the magnitude of a digit in a number. For example the digit 5 in 7458 has an absolute value of 5 according to its value in the number line.

- **The place value** of a digit in a number refers to the position of the digit in that number i.e. whether; tens, hundreds, thousands etc.

- **The total value** of a number is the sum of the place value of each digit making the number.

- **The base value** of a number also k known as the **radix**, depends on the type of the number systems that is being used .The value of any number depends on the radix. For example the number 100_{10} is not equivalent to 100_2.

As far as computers are concerned, number systems can be classified into four major categories:

- decimal number system
- binary number system
- octal number system

- hexadecimal number system

➢ **Decimal number system**

The term decimal is derived from a Latin prefix Deci, which means ten. Decimal number system has ten digits ranging from 0-9. Because this system has ten digits; it is also called a base ten number system.

A decimal number should always be written with a subscript 10 e.g. X_{10}. But since this is the most widely used number system in the world, the subscript is usually understood and ignored in written work.

However, when many number systems are considered together, the subscript must always be put so as to differentiate the number systems.

The decimal number system is a positional notation system that uses ten digits (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) to represent numbers. It is also known as the base-10 number system, as each digit in a decimal number represents a different power of 10.

In the decimal system, the value of each digit is determined by its position in the number. For example, the number 123 has a value of 1 x 10^2 + 2 x 10^1 + 3 x 10^0, which is equal to 100 + 20 + 3 = 123. The first digit from the right represents the ones place, the second digit represents the tens place, the third digit represents the hundreds place, and so on.

The decimal system is widely used in everyday life, from measuring time to money to scientific calculations. It is also the most commonly used number system in mathematics, as it is easy to understand and manipulate.

One advantage of the decimal system is that it allows for easy conversion between fractions and decimals, as well as between different units of measurement. For example, it is easy to convert 2.5 meters to centimeters by multiplying by 100, or to convert 3/4 to a decimal by dividing 3 by 4.

Overall, the decimal system is a fundamental aspect of modern mathematics and plays a crucial role in many aspects of our daily lives.

Example of Decimal Number System:

The decimal number 1457 consists of the digit 7 in the units position, 5 in the tens place, 4 in the hundreds position, and 1 in the thousands place whose value can be written as:

$= (1\times10^3) + (4\times10^2) + (5\times10^1) + (7\times10^0)$

$= (1\times1000) + (4\times100) + (5\times10) + (7\times1)$

$= 1000 + 400 + 50 + 7$

$= 1457$

- **Binary number system**

It uses two digits namely, 1 and 0 to represent numbers. Unlike in decimal numbers where the place value goes up in factors of ten, in binary system, the place values increase by the factor of 2. The binary number system is a positional notation system that uses only two digits, 0 and 1, to represent numbers. It is also known as the base-2 number system, as each digit in a binary number represents a different power of 2.

In the binary system, the value of each digit is determined by its position in the number, just like in the decimal system. For example, the binary number 1011 has a value of 1 x 2^3 + 0 x 2^2 + 1 x 2^1 + 1 x 2^0, which is equal to 8 + 0 + 2 + 1 = 11. The first digit from the right represents the ones place, the second digit represents the twos place, the third digit represents the fours place, and so on. The binary system is widely used in computing and digital electronics, as digital circuits are designed to work with binary signals. In fact, all data in a computer is represented in binary form, as it can be easily manipulated and processed by electronic devices.

One advantage of the binary system is that it is very easy to understand and work with, as there are only two digits to remember. It also allows for easy conversion between binary and decimal numbers, as well as between different units of digital storage. Overall, the binary system is a fundamental aspect of modern computing and digital electronics, and plays a crucial role in many aspects of our daily lives, from smartphones to the internet.

Binary numbers are written as X_2. Consider a binary number such as 1011_2.The right most digit has a place value of 1×2^0 while the left most has a place value of 1×2^3.

For example, 110101 is a binary number.

We can convert any system into binary and vice versa.

Example: Write $(14)_{10}$ as a binary number.

Solution:

2	14	
2	7	0
2	3	1
	1	1

$\therefore (14)_{10} = 1110_2$

- **Octal number system**

Consists of eight digits ranging from 0 - 7. The place value of octal numbers goes up in factors of eight from right to left. Octal numbers are written as X_8.

The octal number system is a positional notation system that uses eight digits (0, 1, 2, 3, 4, 5, 6, and 7) to represent numbers. It is also known as the base-8 number system, as each digit in an octal number represents a different power of 8.

In the octal system, the value of each digit is determined by its position in the number, just like in the binary and decimal systems. For example, the octal number 345 has a value of 3 x 8^2 + 4 x 8^1 + 5 x 8^0, which is equal to 192 + 32 + 5 = 229. The first digit from the right represents the ones place, the second digit represents the eights place, the third digit represents the sixty-fours place, and so on.

The octal system is commonly used in computer programming and digital electronics, particularly in older systems that used eight-bit bytes. It is also used in some non-digital systems, such as the markings on resistors in electronic circuits.

One advantage of the octal system is that it allows for easy conversion between binary and octal numbers, as each octal digit can be represented by a group of three binary digits (bits). For example, the binary number 11011011 can be divided into groups of three bits (110, 110, and 11), which can be represented as the octal number 333.

Overall, while the octal system is not as widely used as the binary and decimal systems, it is still an important part of computer programming and digital electronics. Octal numbers are commonly used in computer applications. Converting an octal number to decimal is the same as decimal conversion and is explained below using an example.

Example: Convert 215_8 into decimal.

Solution:

$215_8 = 2 \times 8^2 + 1 \times 8^1 + 5 \times 8^0$

$= 2 \times 64 + 1 \times 8 + 5 \times 1$

$= 128 + 8 + 5$

$= 141_{10}$

- **Hexadecimal number system**

This is a base 16 number system that consists of sixteen digits ranging from 0-9 and letters A-F where A is equivalent to 10, B to 11 up to F which is equivalent to 15 in base ten system. Hexadecimal numbers are written as X_{16}.

The hexadecimal number system is a positional notation system that uses sixteen digits (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, and F) to represent numbers. It is also known as the base-16 number system, as each digit in a hexadecimal number represents a different power of 16.

In the hexadecimal system, the value of each digit is determined by its position in the number, just like in the binary, decimal, and octal systems. The first ten digits (0-9) represent the same values as in the decimal system, while the remaining six digits (A-F) represent values from 10 to 15.

For example, the hexadecimal number 2AF has a value of $2 x 16^2 + 10 x 16^1 + 15 x 16^0$, which is equal to 688 + 160 + 15 = 863. The first digit from the right represents the ones place, the second digit represents the sixteens place, the third digit represents the 256s place, and so on.

The hexadecimal system is widely used in computer programming and digital electronics, particularly in representing memory addresses and color codes. It is also used in some non-digital systems, such as music notation.

One advantage of the hexadecimal system is that it allows for easy conversion between binary and hexadecimal numbers, as each hexadecimal digit can be represented by a group of four binary digits (bits). For example, the binary number 11011011 can be divided into groups of four bits (1101 and 1011), which can be represented as the hexadecimal number DB.

Overall, the hexadecimal system is an important part of computer programming and digital electronics, and is often used in combination with other number systems like binary and decimal.

The place value of hexadecimal numbers goes up in factors of sixteen. A hexadecimal number can be denoted using 16 as a subscript (X_{16}) or capital letter H to the right of the number. For example, 94B can be written as $94B_{16}$ or 94BH.

The below-given table shows the representation of numbers in the hexadecimal number system.

Hexadecimal	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
Decimal	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**

Example 4:

Convert hexadecimal 2C to decimal number.

Solution:

We need to convert $2C_{16}$ into binary numbers first.

$2C \rightarrow 0010\ 1100$

Now convert 00101100_2 into a decimal number.

$101100 = 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0$

$= 32 + 8 + 4$

$= 44$

7.3 Summary

- Data refers to the symbols that represent people, events, things, and ideas. Data can be a name, a number, the colors in a photograph, or the notes in a musical composition.
- Data Representation refers to the form in which data is stored, processed, and transmitted.
- Analog data is data that is represented in a physical way. Analog data is stored in physical media.
- A representation of something in the form of digital data i.e., computerized data.
- The number system or the numeral system is the system of naming or representing numbers.
- There are different types of number systems in Maths like decimal number system, binary number system, octal number system, and hexadecimal number system.
- There are various types of number systems in mathematics. The four most common number system types are: Decimal number system (Base- 10), Binary number system (Base- 2), octal number system (Base-8) and Hexadecimal number system (Base- 16).

7.4 Keywords

- **Bits:** can be defined as either a binary, which can be 0, or 1.It is the basic unit of data or information in digital computers.
- **Byte:** a group of bits (8 bits) used to represent a character. A byte is considered as the basic unit of measuring memory size in computer.
- **A nibble**: is half a byte, which is usually a grouping of 4 bytes.
- **Word:** two or more bits make a word. The term word length is used as the measure of the number of bits in each word. For example, a word can have a length of 16 bits, 32 bits,64bits etc.

7.5 Review Questions

1. What is Number System and it's Types?
2. Why is the Number System Important?
3. What is Base 10 Number System Called?
4. What is the equivalent binary number for the decimal number 43?
5. How to convert 30_8 into a decimal number?
6. What is Base 8 Number System Called?
7. What is Base 16 Number System Called?
8. What is Data representation on magnetic media>
9. What is Data representation on optical media?
10. What is optical representation of Data?

7.6 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication.
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education.
4. Millman, Electronics Devices and Circuits, TMH.
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi.
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India.

Contents

Learning Objectives

After studying this unit the student will be able to:

1 Describe Conversion from one form to another.

2 Explain the concepts of fractional numbers and signed numbers.

3 Discuss various number systems.

4 Explain number's Complements.

Introduction

There are various types of number system. There are some set of rules for conversion in numbers system. In this unit we will know about what are fractional numbers and signed numbers and their complements.

Converting numbers from one form to another is a fundamental operation in digital systems and computer science. There are several number systems commonly used, including decimal, binary, octal, and hexadecimal. Converting between these systems requires an understanding of the fundamental principles of each system and the conversion process.

Decimal is the most commonly used number system in everyday life, and it uses the base 10 system, which means it has 10 digits, from 0 to 9. Binary, on the other hand, uses the base 2 system, which means it has only two digits, 0 and 1. Octal uses the base 8 system, which means it has eight digits, 0 to 7. Finally, hexadecimal uses the base 16 system, which means it has 16 digits, 0 to 9 and A to F.

Converting from decimal to binary involves dividing the decimal number by 2 repeatedly until the quotient becomes zero. The binary number is then formed by writing the remainders in reverse order. For example, to convert decimal 23 to binary, we divide 23 by 2, which gives a quotient of 11 and a remainder of 1. We then divide 11 by 2, which gives a quotient of 5 and a remainder of 1. We repeat the process, dividing 5 by 2 to get a quotient of 2 and a remainder of 1, and then divide 2 by 2 to get a quotient of 1 and a remainder of 0. Finally, we divide 1 by 2 to get a quotient of 0 and a remainder of 1. The binary number is formed by writing the remainders in reverse order, which gives 10111.

Converting from binary to decimal involves multiplying each digit of the binary number by its corresponding power of 2, and adding up the results. For example, to convert binary 10111 to decimal, we start by writing down the powers of 2: 2^4, 2^3, 2^2, 2^1, 2^0. We then multiply each digit of the binary number by its corresponding power of 2, and add up the results: $1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 16 + 0 + 4 + 2 + 1 = 23$.

Converting from decimal to octal and hexadecimal is similar to converting to binary, but uses base 8 and base 16, respectively. The conversion process involves dividing the decimal number by the base repeatedly until the quotient becomes zero, and then forming the octal or hexadecimal number by writing the remainders in reverse order.

In summary, converting numbers from one form to another requires an understanding of the fundamental principles of each number system and the conversion process. The conversion process involves dividing the number by the base repeatedly until the quotient becomes zero, and then forming the new number by writing the remainders in reverse order. Converting between number systems is an important operation in digital systems and computer science, and is essential for designing and analyzing complex digital circuits and systems.

8.1 Conversion of numbers from one number system to another

To convert numbers from one system to another, the following conversions will be considered

- Conversion between binary and decimal numbers.
- Converting octal numbers to decimal and binary numbers.
- Converting hexadecimal numbers to decimal and binary numbers.

8.1.1 Conversion between binary and decimal number

- **Converting binary number to a decimal number**

For this, we proceed as follows:

- First, write the place values starting from the right hand side.
- Write each digit under its place value.
- Multiply each digit by its corresponding place value.
- Add up the products. The answer will be the decimal number in base ten.

- By placing all the remainders in order in such a way, the Least Significant Bit (LSB) at the top and Most Significant Bit (MSB) at the bottom, the required binary number will be obtained.

Now, let us convert the given decimal number 294 into a binary number.

Divide by 2	**Result**	**Remainder**	**Binary Value**
294 ÷ 2	147	0	0 (LSB)
147 ÷ 2	73	1	1
73 ÷ 2	36	1	1
36 ÷ 2	18	0	0
18 ÷ 2	9	0	0
9 ÷ 2	4	1	1
4 ÷ 2	2	0	0
2 ÷ 2	1	0	0
1 ÷ 2	0	1	1 (MSB)

Therefore, the binary equivalent for the given decimal number 294_{10} is 100100110_2

$$294_{10} = 100100110_2$$

Example 1: Convert 195.25 into binary.

Solution:

195 / 2 = 97 with remainder 1

97 / 2 = 48 with remainder 1

48 / 2 = 24 with remainder 0

24 / 2 = 12 with remainder 0

12 / 2 = 6 with remainder 0

6 / 2 = 3 with remainder 0

3 / 2 = 1 with remainder 1

1 / 2 = 0 with remainder 1

Thus, the binary equivalent of 195 is 11000011.

Now, we have to convert the fractional part of the given number into binary.

Multiply 0.25 by 2 and observe the resulting integer and fractional parts. Renew multiplying the resultant fractional part by 2 until we get a resulting fractional part equal to zero.

Then we need to write the integer parts from the results of each multiplication to make the equivalent binary number.

$0.25 \times 2 = 0.5$

$0.5 \times 2 = 1.0$ (fraction becomes zero)

Here, 0.25 is equivalent to the binary number 0.01.

Therefore, $(195.25)_{10} = (11000011.01)_2$

Example 2: Convert decimal number 0.375_{10} into binary form.

Solution:

$0.375 \times 2 = 0.750$

$0.750 \times 2 = 1.500$

$0.500 \times 2 = 1.000$ (fraction becomes zero)

Therefore $0.375_{10} = 0.011_2$

8.1.2 Conversion between hexadecimal numbers to decimal and binary numbers

- **Converting hexadecimal numbers to decimal number**

There is a conversion table –

Hexadecimal digit	Decimal equivalent	Binary equivalent
00	00	0000
01	01	0001
02	02	0010
03	03	0011
04	04	0100
05	05	0101
06	06	0110
07	07	0111
08	08	1000
09	09	1001
A	10	1010
B	11	1011
C	12	1100
D	13	1101
E	14	1110
F	15	1111

To convert hexadecimal number to base 10 equivalent we proceed as follows:

- First, write the place values starting from the right hand side.
- If a digit is a letter such as 'A', write its decimal equivalent.
- Multiply each hexadecimal digit with its corresponding place value and then add the products.

Example 1: Convert 7CF (hex) to decimal.

Solution: Given hexadecimal number is 7CF.

In hexadecimal system,

$$7 = 7$$

$$C = 12$$

$$F = 15$$

To convert this into a decimal number system, multiply each digit with the powers of 16 starting from units place of the number.

$$7CF = (7 \times 16^2) + (12 \times 16^1) + (15 \times 16^0)$$

$$= (7 \times 256) + (12 \times 16) + (15 \times 1)$$

$$= 1792 + 192 + 15$$

$$= 1999$$

From this, the rule can be defined for the conversion from hex numbers to decimal numbers.

- **Converting hexadecimal numbers to binary numbers**

The simplest method of converting a hexadecimal number to binary is to express each hexadecimal digit as a four bit binary digit number and then arranging the group according to their corresponding positions as shown in example.

Example 1: Convert 321_{16} into binary.

Solution:

Hexadecimal digit	3	2	1
Binary equivalent	0011	0010	0001

Combining the three sets of bits, we get 001100100001_2

$$321_{16} = 001100100001_2$$

Example 2: Convert $5E6_{16}$ into binary.

Solution: Hexadecimal digit $5E6_{16}$

Hexadecimal digit	5	E	6
Binary equivalent	0101	1110	0110

$$5E616 = 010111100110_2$$

8.1.3 Conversion between octal numbers to decimal and binary numbers

- **Converting octal numbers to decimal numbers**
  - To convert a base 8 number to its decimal equivalent we use the same method as we did with binary numbers.
  - However, it is important to note that the maximum absolute value of an octal digit is 7.

Example 1: Convert 512_8 to its base 10 equivalent.

Solution:

Octal digit	5	1	2

Place value	8^2	8^1	8^0
	64	8	1

Write each number under its place value as shown below

Multiply each number by its place value.

$$N_{10}=(5 \times 8^2)+(1 \times 8^1)+(2 \times 8^0)$$

$$=(5 \times 64)+8+2$$

$$=320+8+2$$

$$N_{10}=330_{10}$$

Therefore $512_8 = 330_{10}$

- **Converting octal numbers to binary numbers**

To convert an octal number to binary, each digit is represented by three binary digits because the maximum octal digit i.e. 7 can be represented with a maximum of seven digits. See table:

Octal digit	Binary equivalents
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

Example 1: Convert octal number 321_8 to its binary equivalent.

Solution: Working from left to the right, each octal number is represented using three digits and then combined. We will get the final binary equivalent. Therefore:

octal number	3	2	1
binary equivalent	011	010	001

Combining the three binary digit from left to right

$$321_8 = 011010001_2$$

8.2 Fractional Numbers

Fractional numbers, also known as real numbers, are an important part of digital systems as they allow for more precise calculations and measurements. In digital systems, fractional numbers are represented using a fixed or floating-point format, where the number is divided into two parts: the integer part and the fractional part. The integer part represents the whole number part of the value, while the fractional part represents the decimal or fractional part of the value.

In fixed-point representation, the number of bits allocated to the integer and fractional parts are fixed and predetermined. This means that the range and precision of the fractional number representation are limited. Fixed-point representation is commonly used in applications where the number range and precision requirements are known in advance.

In floating-point representation, the number of bits allocated to the integer and fractional parts can vary depending on the value being represented. This allows for a wider range of values and precision levels than fixed-point representation. Floating-point representation is commonly used in applications where the range and precision requirements are not known in advance, such as scientific and engineering applications.

Both fixed-point and floating-point representation have their advantages and disadvantages, and the choice of representation depends on the specific requirements of the application. However, both representations allow for the use of fractional numbers in digital systems, enabling precise and accurate calculations and measurements.

8.3 Signed Numbers Format

In digital systems, signed numbers are used to represent both positive and negative values. Signed numbers are used in a wide range of applications, from simple calculations to complex algorithms and control systems. Signed numbers are essential in digital systems because they allow for the representation of negative values, which are necessary in many applications.

Signed numbers can be represented using different formats, including signed magnitude, one's complement, and two's complement. In signed magnitude representation, the most significant bit represents the sign of the number, with 0 representing a positive value and 1 representing a negative value. The remaining bits represent the magnitude of the number. While this format is easy to understand and implement, it can lead to issues with arithmetic operations, particularly with negative numbers.

One's complement representation is similar to signed magnitude representation, but with a twist. In this format, the negative number is obtained by flipping all the bits of the positive number. While this format solves some of the issues with signed magnitude representation, it can still result in inconsistencies with arithmetic operations.

Two's complement representation is the most commonly used format for representing signed numbers in digital systems. In this format, the negative number is obtained by taking the one's complement of the positive number and adding 1. This format ensures that the negative number is unique and solves the issues with arithmetic operations.

While signed numbers are essential in digital systems, they can also present some challenges. One challenge is overflow, which occurs when the result of an arithmetic operation exceeds the maximum value that can be represented by the number of bits allocated. Another challenge is underflow, which occurs when the result of an arithmetic operation is smaller than the minimum value that can be represented by the number of bits allocated.

In conclusion, signed numbers are an essential part of digital systems, allowing for the representation of both positive and negative values. Different formats, including signed magnitude, one's complement, and two's complement, can be used to represent signed numbers, with two's complement being the most used. While signed numbers can present some challenges, such as overflow and underflow, they are necessary for a wide range of applications, from simple calculations to complex control systems.

8.4 Complements in Digital Systems

In digital systems, complements are used to represent negative numbers, perform subtraction, and simplify arithmetic operations. There are three types of complements used in digital systems: one's complement, two's complement, and ten's complement.

One's complement is a way of representing negative numbers by flipping all the bits of a positive number. This is done by subtracting each bit from 1 to obtain its complement. For example, the one's complement of the binary number 10110 is 01001. While one's complement can be used to represent negative numbers, it has some limitations, including the existence of two representations of zero and inconsistencies with arithmetic operations.

Two's complement is the most widely used complement in digital systems. In two's complement representation, negative numbers are represented by taking the one's complement of the positive number and adding 1 to the result. For example, the two's complement of the binary number 10110 is 01010. Two's complement representation ensures that there is only one representation of zero and solves the inconsistencies with arithmetic operations found in one's complement representation.

Ten's complement is used in decimal systems to perform subtraction by adding the complement of the subtrahend to the minuend. The ten's complement of a decimal number is obtained by subtracting each digit from 9. For example, the ten's complement of 1234 is 8766. To subtract 456 from 1234 using ten's complement, we add the ten's complement of 456, which is 5434, to 1234, resulting in 6668.

In addition to negative number representation and subtraction, complements are also used to simplify arithmetic operations, such as addition and multiplication. In addition, complements can be used to perform logical operations, such as bit-wise negation and complementation.

While complements are useful in digital systems, they can also present some challenges, such as overflow and underflow. Overflow occurs when the result of an arithmetic operation exceeds the maximum value that can be represented by the number of bits allocated. Underflow occurs when the result of an arithmetic operation is smaller than the minimum value that can be represented by the number of bits allocated. These challenges can be addressed using techniques such as modulo arithmetic and saturation arithmetic.

In conclusion, complements are an essential part of digital systems, used to represent negative numbers, perform subtraction, simplify arithmetic operations, and perform logical operations. One's complement, two's complement, and ten's complement are the three main types of complements used in digital systems. While complements are useful, they can also present some challenges, such as overflow and underflow, which must be addressed using appropriate techniques.

8.5 Summary

In digital systems, numbers can be represented in various forms, such as binary, octal, decimal, and hexadecimal. Converting numbers from one form to another is a common task in digital systems and can be done using various methods, such as direct conversion, conversion by substitution, and conversion by successive division.

Direct conversion involves converting a number from one base to another using a conversion table, which lists the equivalent values of each digit in different bases. Conversion by substitution involves converting a number by substituting the equivalent values of each digit in the target base. Conversion by successive division involves dividing the original number by the target base repeatedly until the quotient is zero, and then reading the remainders in reverse order to obtain the target number. In addition to integer numbers, digital systems also deal with fractional numbers, which can be represented in various forms, such as fixed-point and floating-point formats. Fixed-point formats represent fractional numbers with a fixed number of digits after the decimal point, while floating-point formats represent fractional numbers using a mantissa and an exponent.

Signed numbers are also used in digital systems to represent negative numbers, and complements are used to simplify arithmetic operations and represent negative numbers. One's complement, two's complement, and ten's complement are the three main types of complements used in digital systems. Converting numbers in digital systems is an essential task that requires a good understanding of different number systems, their representations, and conversion methods. With proper knowledge and skills, converting numbers from one form to another can be done efficiently and accurately, enabling digital systems to perform various computations and operations.

8.6 Key Words

- Binary
- Octal
- Decimal
- Hexadecimal
- Direct conversion
- Conversion by substitution
- Conversion by successive division
- Fixed-point
- Floating-point
- Complements

8.7 Review Questions

1. What are the different number systems used in digital systems?
2. What are the three main methods used for converting numbers from one base to another?
3. How is direct conversion used to convert numbers in digital systems?
4. How is conversion by substitution used to convert numbers in digital systems?
5. How is conversion by successive division used to convert numbers in digital systems?
6. What are the different formats used to represent fractional numbers in digital systems?
7. How are signed numbers represented in digital systems?
8. What is the purpose of using complements in digital systems?
9. What are the different types of complements used in digital systems?
10. Why is the conversion of numbers important in digital systems?

8.8 References

- Floyd, T. L. (2015). Digital fundamentals. Pearson.
- Brown, S. H., & Vranesic, Z. G. (2017). Fundamentals of digital logic with VHDL design. McGraw-Hill Education.
- Roth, J. P. (2015). Fundamentals of logic design. Cengage Learning.
- Mano, M. M., & Ciletti, M. D. (2017). Digital design. Pearson.
- Givone, D. (2018). Digital principles and design. Oxford University Press.
- Wakerly, J. F. (2017). Digital design: principles and practices. Pearson.
- Palnitkar, S. (2013). Verilog HDL: A guide to digital design and synthesis. Prentice Hall.
- Hwang, K., & Briggs, F. A. (2016). Computer arithmetic: principles, architecture, and design. Elsevier.

Contents

UNIT-9

Operation on Binary Numbers

Learning Outcomes

After studying this unit, the student will be able to:

1. Describe operations on binary numbers
2. Explain the concepts used in various operations
3. Explain the concepts in building higher order models using basic building blocks
4. Explain Behavior of fixed- and floating-point representation of binary numbers

9 Operations on Binary Numbers

Operations on binary numbers form the bedrock of digital computation, enabling the manipulation and processing of information in the form of zeros and ones. Binary numbers, composed solely of these two digits, serve as the backbone of all digital systems, from simple electronic circuits to complex computer algorithms. Arithmetic operations on binary numbers, such as addition, subtraction, multiplication, and division, are fundamental to performing numerical computations in binary form. These operations allow for the manipulation of binary data, enabling tasks ranging from basic calculations to complex mathematical algorithms. By leveraging binary arithmetic, computers can accurately process and manipulate numerical data, forming the basis for countless applications in areas such as scientific research, data analysis, and financial modelling.

Logical operations on binary numbers, including bitwise AND, OR, XOR, and NOT, enable the manipulation of individual bits within binary representations. These operations are vital for tasks like data masking, bitwise comparisons, and encoding and decoding of binary structures. Logical operations underpin Boolean algebra, forming the basis for logical decision-making and circuit design. A firm understanding of operations on binary numbers is crucial for computer scientists, programmers, and engineers working with digital systems. Proficiency in these operations enables the design and implementation of efficient algorithms, the development of robust software, and the construction of sophisticated hardware systems.

In summary, operations on binary numbers are at the heart of digital computation. They allow for the manipulation and processing of binary data, facilitate numerical computations, and form the foundation for logical decision-making. Mastering these operations is key to unlocking the vast potential of digital systems and technology. The classifications of operations on binary numbers are defined as

- Arithmetic Operations
- Logical Operations

9.1 Arithmetic operation on binary numbers

Arithmetic operations on binary numbers involve performing addition, subtraction, multiplication, and division. Binary numbers use a base-2 system with only two digits, 0 and 1. Understanding these operations can be helpful in various fields such as computer science, digital electronics, and information technology. In this explanation, we'll explore each operation and how it is performed on binary numbers.

9.1.1 Addition

To add binary numbers, follow these steps:

Start from the rightmost digit (least significant bit) and add the corresponding bits from both numbers.

- If the sum is 0 or 1, write it down.
- If the sum is 2, write down 0 and carry over 1 to the next column.
- If the sum is 3, write down 1 and carry over 1 to the next column.
- Repeat the above steps for each column, moving from right to left.
- If there's a carry after adding the leftmost digits, include an additional bit to represent it.

Example:

1011 (decimal 11) (let us consider this as binary number 1 (BN 1)

1101 (decimal 13) (let us consider this as binary number 2 (BN 2)

Table 1 represents the detail solution for the given binary addition

Table 1 Solution

Carry	1	1	1	1	
BN1		1	0	1	1
BN2		1	1	0	1
Answer	**1**	**1**	**0**	**0**	**0**

Final answer of the given binary addition is = 11000 (decimal 24)

9.1.2 Subtraction

To subtract binary numbers, following steps are performed:

- Start from the rightmost digit (least significant bit) and subtract the corresponding bits.
- If the minuend bit is smaller than the subtrahend bit, borrow 1 from the next higher bit.
- Subtract the borrowed 1 from the minuend bit and proceed.
- Repeat the above steps for each column, moving from right to left.

Example:

1101 (decimal 13) (let us consider this as binary number 1 (BN 1)

1011 (decimal 11) (let us consider this as binary number 2 (BN 2)

Table 2 shows the detail solution of the given binary subtraction

Table 2 Solution

Borrow		0	0	1	
BN1		1	1	0	1
BN2		1	0	1	1
Answer		**0**	**0**	**1**	**0**

Here the final answer of the given problem is 0010 (Decimal number is 2)

9.1.3 Multiplication

To multiply binary numbers, follow these steps:

- Multiply each bit of the multiplier with every bit of the multiplicand, starting from the rightmost bit of the multiplier.
- Place the resulting products in their appropriate positions, shifting them to the left according to their column.
- Add the shifted products together to get the final result.

Example:

1101 (decimal 13) (let us consider this as binary number 1 (BN 1)

x 11 (decimal 3) (let us consider this as binary number 2 (BN 2)

Table 3 shows the detail solution of the given binary multiplication

Table 3 Solution

BN1			1	1	0	1
BN2			0	0	1	1
		1	1	1	0	1
		1	1	0	1	X
Final Answer	1	0	0	1	1	1

Final answer for the given example is 100111 (Decimal number =39)

9.1.4 Division

To divide binary numbers, follow these steps:

- Compare the divisor with the leftmost bits of the dividend.
- If the divisor is larger or equal, subtract it from the dividend and write down 1 in the quotient.

- If the divisor is smaller, write down 0 in the quotient.
- Shift the dividend to the left by 1 bit, bringing down the next bit from the dividend.
- Repeat the above steps until the entire dividend has been processed.

Example:

11101 (decimal 29)

÷ 11 (decimal 3)

1001 (decimal 9)

These are the basic arithmetic operations on binary numbers. Understanding and applying these operations is essential in binary arithmetic and in various fields where binary representation is used.

9.2 Logical operations of binary numbers

Logical operations on binary numbers involve performing bitwise AND, OR, XOR, and NOT operations. These operations manipulate individual bits in binary numbers. Let's explore each operation in detail:

Bitwise AND

The bitwise AND operation compares each corresponding bit of two binary numbers and produces a new binary number where each bit is the logical AND of the corresponding bits in the input numbers.

If both bits are 1, the result is 1.

Otherwise, the result is 0.

Example:

1101 (decimal 13)

& 1011 (decimal 11)

1001 (decimal 9)

Bitwise OR

The bitwise OR operation compares each corresponding bit of two binary numbers and produces a new binary number where each bit is the logical OR of the corresponding bits in the input numbers.

If at least one of the bits is 1, the result is 1.

Otherwise, the result is 0.

Example:

1101 (decimal 13)

1011 (decimal 11)

1111 (decimal 15)

Bitwise XOR (Exclusive OR)

The bitwise XOR operation compares each corresponding bit of two binary numbers and produces a new binary number where each bit is the logical XOR of the corresponding bits in the input numbers.

If the bits are different (one is 0 and the other is 1), the result is 1.

If the bits are the same (both 0 or both 1), the result is 0.

Example:

1101 (decimal 13)

^ 1011 (decimal 11)

1110 (decimal 14)

Bitwise NOT

The bitwise NOT operation, also known as complement, flips each bit in a binary number.

If the bit is 0, it becomes 1.

If the bit is 1, it becomes 0.

Example:

~ 1101 (decimal 13)

0010 (decimal 2)

These logical operations on binary numbers are essential in various fields, including computer science, digital logic design, and information technology. They allow us to manipulate and analyse binary data efficiently.

9.3 Examples of arithmetic operations on binary numbers

In this part, we will discuss the numerical example of various arithmetic operations on binary numbers in detail. We will start with addition and its numerical examples and try to understand the basic phenomenon of addition operation on binary numbers. Similarly, we will discuss other operations such as subtraction, multiplication and division with similar approach.

Let us take sufficient number of numerical examples to better understand the concepts of arithmetic operations on binary numbers.

9.3.1 Numerical examples of addition on binary numbers

Here are 10 examples of addition of binary numbers:

Binary number 1= 101

Binary number 2= 110

BN1		1	0	1
BN2		1	1	0
Answer	**1**	**0**	**1**	**1**

Ans: 1011

2.

Binary number 1: 1001

Binary number 2: 1110

Ans: 10111

3.

Binary number 1: 11011

Binary number 2: 10101

Ans: 110100

4.

Binary number 1:11100

Binary number 2: 11010

Ans: 101010

5.

Binary number 1:10010

Binary number 2: 11001

Ans: 101011

Example Number 6	
Binary number 1	10101
Binary number 2	10110
Answer	101111

Example Number 7	
Binary number 1	111111
Binary number 2	101010
Answer	1011001

Example Number 8	
Binary number 1	110001
Binary number 2	100011
Answer	1010010

Example Number 9	
Binary number 1	101101
Binary number 2	110011
Answer	1001110

Example Number 10	
Binary number 1	111000
Binary number 2	110011
Answer	1010011

These examples illustrate the addition of binary numbers by following the same principles as explained earlier, where corresponding bits are added, carrying over to the next column if necessary.

9.3.2 Numerical examples of subtraction on binary numbers

Here are 10 examples of subtraction of binary numbers:

Binary number 1: 1011

Binary number 2: 110

BN1	1	0	1	1
BN2		1	1	0
Answer	**0**	**1**	**0**	**1**

Answer: 101

2.

Binary Number 1: 10101

Binary Number 2: 1110

Answer: 10011

3. 11011

10101

yaml

Copy code

1010

4. 11100

11010

Copy code

10

5. 10010

11001

yaml

For Vivekananda Global University, Jaipur

Copy code

1011

6. 10101

10110

markdown

Copy code

1

7. 111111

101010

Copy code

10101

8. 110001

100011

Copy code

10010

9. 101101

110011

yaml

Copy code

1110

10. 111000

110011

yaml

Copy code

1101

These examples demonstrate the subtraction of binary numbers by applying the borrowing principle when the minuend bit is smaller than the subtrahend bit. The borrow allows for the subtraction to be carried out correctly in each column, resulting in the correct binary difference.

9.3.3 Numerical examples of multiplication on binary numbers

Here are 10 examples of multiplication of binary numbers:

101

x 110

Copy code

10100

2. 1001

x 1110

1000110

11011

x 10101

1111111

11100

x 11010

10001100

10010

x 11001

11001010

10101

x 10110

11011110

111111

x 101010

10000010110

110001

x 100011

1100111111

101101

x 110011

11101100011

111000

x 110011

1010011000

These examples illustrate the multiplication of binary numbers by performing bitwise multiplication and shifting operations to obtain the final product. Each bit of the multiplier is multiplied by the multiplicand, and the partial products are shifted and added together to obtain the final result.

9.3.4 Numerical examples of division on binary numbers

Here are 10 examples of division of binary numbers:

1011

÷ 10

Quotient: 10

Remainder: 11

10101

÷ 11

Quotient: 101

Remainder: 0

11011

÷ 101

Quotient: 110

Remainder: 1

11100

÷ 110

Quotient: 101

Remainder: 0

10010

÷ 111

Quotient: 11

Remainder: 101

10101

÷ 10

Quotient: 1010

Remainder: 1

111111

÷ 1010

Quotient: 1101

Remainder: 111

110001

÷ 1001

Quotient: 11111

Remainder: 0

101101

÷ 101

Quotient: 1001

Remainder: 0

111000

÷ 1101

Quotient: 10110

Remainder: 10

These examples demonstrate the division of binary numbers by performing bitwise division and shifting operations. The divisor is compared to the leftmost bits of the dividend, and the quotient is determined based on the comparison. The process is repeated until the entire dividend has been processed, resulting in a quotient and remainder.

9.4 Representation of binary numbers

9.4.1 Fixed point representation of binary numbers

Fixed-point and floating-point representations are two ways to represent real numbers in binary form.

Fixed-Point Representation:

In fixed-point representation, a specific number of bits are allocated to represent the integer and fractional parts of a number. The position of the decimal point is fixed. The range and precision of fixed-point numbers are determined by the allocation of bits.

For example:

In a 8-bit fixed-point representation with 4 bits for the integer part and 4 bits for the fractional part, the number 6.75 would be represented as 0110.1100.

Fixed-point representation is commonly used in applications where a fixed range and precision are required, such as digital signal processing (DSP) and embedded systems.

Examples of Fixed point representation for binary numbers are:

8-bit Fixed-Point Representation (4 integer bits, 4 fractional bits):

0101.1001 represents 5.5625

16-bit Fixed-Point Representation (8 integer bits, 8 fractional bits):

11001011.01110010 represents -37.71484375

12-bit Fixed-Point Representation (6 integer bits, 6 fractional bits):

101100.001010 represents -12.15625

10-bit Fixed-Point Representation (5 integer bits, 5 fractional bits):

00101.11110 represents 0.96875

6-bit Fixed-Point Representation (3 integer bits, 3 fractional bits):

110.011 represents -1.375

In fixed-point representation, the decimal point is at a fixed position, separating the integer and fractional parts. The number of bits allocated for the integer and fractional parts determines the range and precision of the fixed-point number. The examples above demonstrate different fixed-point representations with varying bit allocations and corresponding binary numbers representing real values.

9.4.2 Floating point representation of binary numbers

Floating-point representation uses a combination of a sign bit, an exponent, and a fraction (also called mantissa) to represent real numbers. The position of the decimal point is not fixed and can be adjusted by the exponent.

The IEEE 754 standard is widely used for floating-point representation, which specifies formats for single precision (32 bits) and double precision (64 bits) numbers. The representation allows for a wider range and varying levels of precision.

For example:

In IEEE 754 single precision, the number 6.75 would be represented as:

Sign bit: 0 (positive)

Exponent: 10000010 (130 in decimal, bias of 127 added)

Fraction: 10110000000000000000000 (normalized fraction)

The final representation is: 0 10000010 10110000000000000000000.

Floating-point representation is used in various applications where a wider range and variable precision are required, such as scientific calculations, graphics processing, and numerical simulations.

Single Precision Floating Point (32 bits):

0 10000011 10100000000000000000000 represents -6.5

Single Precision Floating Point (32 bits):

1 10000000 01010101010101010101011 represents -0.33333334

Double Precision Floating Point (64 bits):

0 10000000011 1100010000000000000000000000000000000000000000000000 represents - 456.25

Single Precision Floating Point (32 bits):

0 01111101 01000000000000000000000 represents 0.15625

Double Precision Floating Point (64 bits):

1 10000000010 1011001000000000000000000000000000000000000000000000 represents - 123.75

In the IEEE 754 floating-point representation, the number is expressed in scientific notation as (-1)^s * 1.m * 2^e, where s represents the sign bit, m represents the normalized fraction, and e represents the biased exponent. The examples above showcase different binary representations using single and double precision formats, with varying sign bits, exponents, and fractions, representing real values in binary form.

9.5 Summary

Binary numbers are a fundamental part of digital systems and computing. They use a base-2 numeral system, consisting of only two digits: 0 and 1. Each digit in a binary number is called a bit, and the position of each bit represents a power of 2. This positional notation allows binary numbers to efficiently represent and manipulate data in digital systems. The process of converting decimal numbers to binary, known as binary conversion, involves repeatedly dividing the decimal number by 2 and recording the remainders. The binary representation is obtained by arranging the remainders in reverse order. Binary numbers can also be converted back to decimal by multiplying each bit by its corresponding power of 2 and summing the results.

Binary numbers find widespread use in various digital applications. They form the basis of digital communication and storage systems, such as computers, calculators, and electronic devices. Binary arithmetic operations, including addition, subtraction, multiplication, and division, are performed on binary numbers to process numerical data in digital systems.

Logical operations on binary numbers, such as AND, OR, XOR, and NOT, manipulate individual bits within binary representations. These operations are essential for tasks such as data masking, bitwise comparisons, and encoding and decoding of binary structures.

Understanding the representation of binary numbers is vital for professionals in computer science, electrical engineering, and related fields. Proficiency in binary representation and operations enables efficient algorithm design, hardware system development, and data manipulation in digital systems.

In summary, binary numbers are a foundational concept in digital systems. Their representation and operations provide the basis for efficient data manipulation, digital computation, and logical decision-making in numerous applications across various industries.

9.6 Keywords

- Binary numbers
- Base-2 numeral system
- Bit
- Positional notation
- Binary conversion
- Decimal to binary
- Binary to decimal
- Binary arithmetic
- Addition
- Subtraction
- Multiplication
- Division
- Logical operations
- Bitwise operations
- Digital systems

9.7 Review Questions

- What are binary numbers and how are they different from decimal numbers?
- Explain the process of converting a decimal number to binary.
- How does positional notation work in binary representation?
- What are the key arithmetic operations performed on binary numbers?
- Describe the steps involved in performing binary addition.
- How do you perform binary subtraction?
- What are the common methods for binary multiplication?
- Explain the process of binary division.
- What are logical operations on binary numbers, and why are they important?
- How do bitwise operations like AND, OR, XOR, and NOT work on binary numbers?

9.8 References

Title: "Computer Organization and Design: The Hardware/Software Interface"
Authors: David A. Patterson, John L. Hennessy
Publisher: Morgan Kaufmann Publishers
Year: 2014

Title: "Digital Design and Computer Architecture"
Authors: David Money Harris, Sarah L. Harris
Publisher: Morgan Kaufmann Publishers
Year: 2012

Title: "Computer Systems: A Programmer's Perspective"
Authors: Randal E. Bryant, David R. O'Hallaron
Publisher: Pearson
Year: 2016

Title: "Computer Science: An Overview"
Authors: J. Glenn Brookshear, Dennis Brylow
Publisher: Pearson
Year: 2018

Title: "Logic and Computer Design Fundamentals"
Authors: M. Morris Mano, Charles R. Kime
Publisher: Pearson
Year: 2018

Contents

Learning Objectives

After studying this unit the student will be able to:

1 Describe Logic gates and its types.

2 Explain the concepts and Boolean algebra.

3 Discuss Their Truth Tables.

4 Explain various types of logic gates.

Introduction

The PN junction can be use to make various electronic devices such as Diodes, transistors, FET, MOSFETs. These devices can be further use to make logic gates to perform various functions in electronic circuits. Boolean Algebra is a branch of algebra that deals with the mathematical operations and relationships between logic gates, which are fundamental building blocks of digital circuits. In Boolean Algebra, variables can only take on two values, true or false, represented by the binary digits 1 and 0, respectively. Logical operations such as AND, OR, and NOT can be used to manipulate these values and create complex expressions that describe the behavior of digital circuits. Boolean Algebra is a powerful tool for analyzing and designing digital circuits, and is widely used in computer science and engineering.

Logic gates are physical devices that implement Boolean logic operations. There are several types of logic gates, including AND, OR, NOT, XOR, NAND, and NOR gates, each with a specific function and behavior. The behavior of each logic gate can be described using a truth table, which shows the output of the gate for all possible combinations of input values. For example, an AND gate has two input terminals and one output terminal, and its truth table shows that the output is high (1) only when both input terminals are high (1), and low (0) otherwise. Truth tables are an essential tool for analyzing the behavior of digital circuits and designing complex systems that meet specific requirements. By combining multiple logic gates together, it is possible to create complex digital circuits that can perform a wide range of functions, from simple arithmetic operations to complex control systems.

10.1 Analogue Signal

A continuous time varying current or voltage signal is called analogue signal.

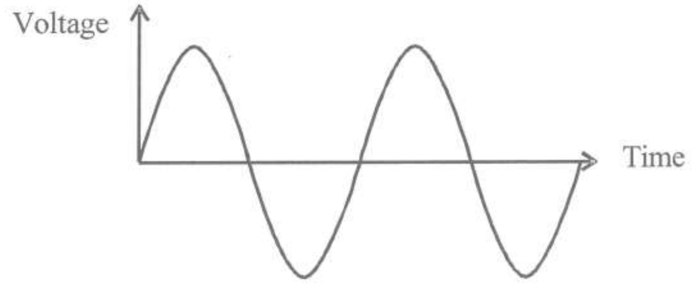


Figure 1: Analogue Signal

10.2 Digital Signal

The signal which have two levels of current or voltage (represented by 0 &a) are called digital signal

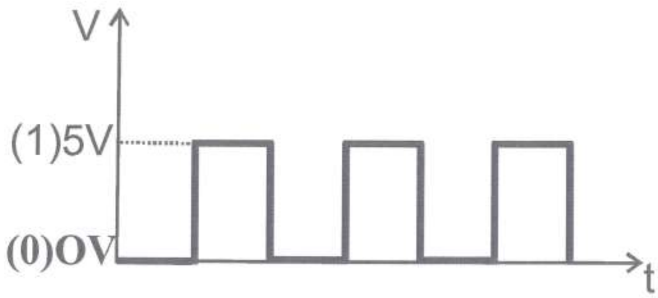


Figure 2: Digital Signal

10.3 Logic Gates

A Logic Gate is a digital circuit which follows some logical relationship between input & output signal. The logic gates are building block of digital electronics & its function is defined either by a truth table or by Boolean algebra.

In digital electronic signals have only two levels. In digital circuit high, true, yes, closed such conditions are represented by level 1 and low, false, no, open such conditions are represented by level 0.

10.4 Truth table

It is a table that shows all possible combinations and the corresponding output combinations for a logic gate. It is also called table of combinations.

10.5 Boolean expression

A	B	Y = A + B
0	0	0
0	1	1
1	0	1
1	1	1

Figure 4: Truth Table of OR gate

An OR gate can be realized by the electronic circuit, making use of two ideal p-n junction diode D1 and D2 as shown in figure.

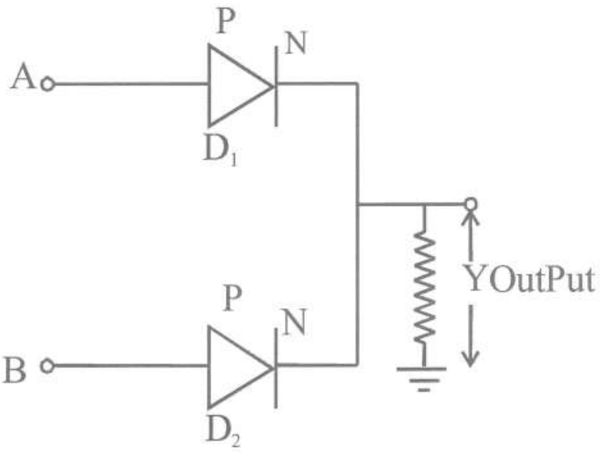


Figure 5: Circuit for OR gate

Here negative terminal of the battery is grounded and corresponds to the 0 level and the positive terminals of the battery (i.e. Voltage 5V in the present case) corresponds to the level 1. The output Y is voltage at C w.r.t. earth. The operation of OR gate can be understood by the following our cases-

(i) When A and B are connected to earth (A=0 and B=0) both the diodes do not conduct and therefore no voltage develops across resistance R. The voltage at C is zero w.r.t. earth. Hence the output Y is 0 (in level).

(ii) When A is connected to earth & B is connected to positive terminal of battery 5V (A=0 and B=1), the junction diode D_1 does not conduct while D_2 conducts being forward biased.

Since diode D2 is ideal, no voltage drop take place across D_2. Now a voltage drop of 5V takes place across R with C at +5V w.r.t earth. Therefore the output Y will be 1 (in level).

(iii) When B is connected to earth & A is connected to positive terminal of battery 5V (A=1 and B=0), the junction diode D_2 does not conduct while D_1 conducts being forward biased. Since diode D_2 is ideal , no voltage drop take place across D_1 .Now a voltage drop of 5V take place across R with C at +5V w.r.t earth. Therefore the output Y will be 1 (in level)

(iv) When A & B are connected to positive terminal of battery 5V (A=1 and B=1), both junction diode D_1 & D_2 conducts being forward biased. Since both diode D_1& D_2 are ideal, no voltage drop take place across D_1& D_2. Now a voltage drop of 5V takes place across R with C at +5V w.r.t earth. Therefore the output Y will be 1 (in level).

If Switch S_1 is input A, Switch S_2 & the output bulb is Y then we have the truth table of the OR operation. In this table, the false value is represented by zero and true value by one.

The analogue electric circuit having function similar to the OR gate is shown in figure. In this arrangement, off (or open) corresponds to 0 and on (or closed) corresponds to 1. The inputs are introduced through the switch A & B. The lighting of the bulb is the output. Here we find that the bulb glow glows (i.e. output is 1) when either switch A & B is closed or both the switches are closed. The bulb remain off (output is 0) only when both the switches A are B are (A=0 B=0)

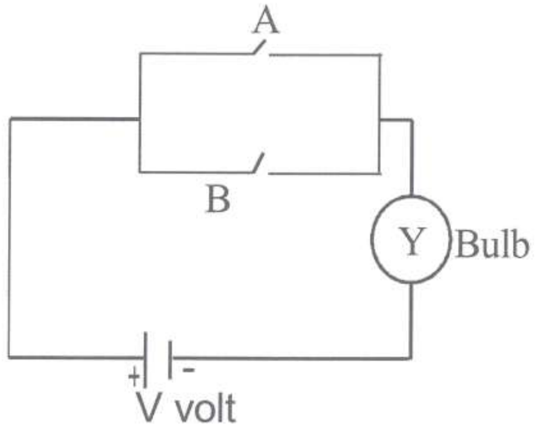


Figure 6: analogue electric circuit for OR gate

10.6.2 AND Gate

The AND gate has two or more inputs and one output. AND gate follows the logic in which output has value 1 only when all inputs have 1 value, output has 0 value when one or more input have 0 value.

The logic symbol of AND gate with A and B input & Y is output is shown in figure.

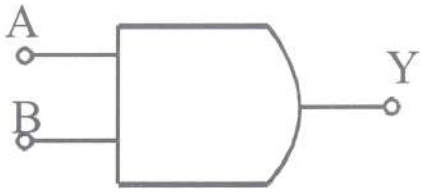


Figure 7: AND gate symbol

A	B	Y = A.B
0	0	0
0	1	0
1	0	0
1	1	1

Figure 8: Truth Table for AND gate

In Boolean algebra, multiplication symbol (Dot) is referred to as OR. The Boolean expression A.B = Y indicates that Y equals A AND B.

An AND gate can be realized by the electronic circuit, making use of two ideal p-n junction diode D_1 and D_2 as shown in figure.

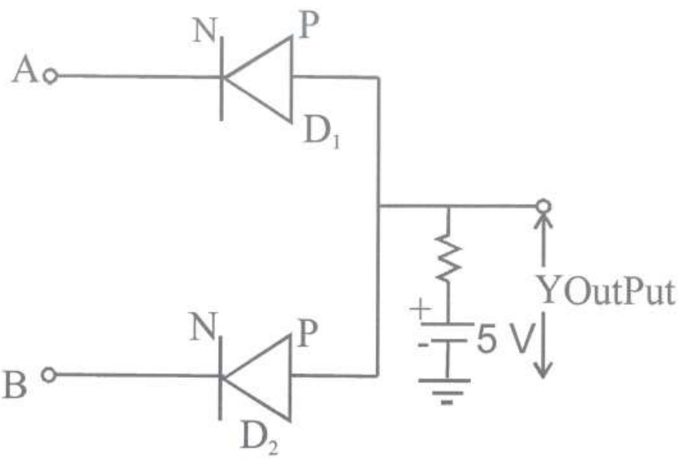


Figure 9: AND gate circuit diagram

Here negative terminal of the battery is grounded and corresponds to the 0 level, and the positive terminals of the battery (i.e. Voltage 5V in the present case) correspond to the level 1. The output Y is voltage at C w.r.t. earth. The operation of OR gate can be understood by the following our cases -

(i) When A and B are is connected to earth (A=0 and B=0) both the diodes gets forward biased hence conduct .The diodes being ideal , no voltage drop take place across either diode . Therefore a voltage drop of 5V takes place across resistance R, with C at zero potential w.r.t earth. Hence the output Y is 0(in level)

(ii) When A is connected to earth & B is connected to positive terminal of battery 5V (A=0 and B=1), the junction diode D_1 will conduct while D_2 will not conduct. Since diode D_1 is ideal , no voltage drop take place across D_1 .Therefore a voltage drop of 5V take place across R with C at +0V w.r.t earth. Hence the output Y is 0(in level)

(iii) When B is connected to earth & A is connected to positive terminal of battery 5V(A=0 and B=1), the junction diode D_2 will conduct while D_1 will not conduct .Since diode D_2 is ideal , no voltage drop take place across D_2 .Therefore a voltage drop of 5V take place across R with C at +0V w.r.t earth. Hence the output Y is 0 (in level)

(iv) When A & B are connected to positive terminal of battery 5V (A=0 and B=1), the junction diode D_1 & D_2 will not conduct. There will be no current through R. Now potential at C is at +5V w.r.t. earth .Hence the output Y is 1.

The analogue electric circuit having function similar to the AND gate is shown in figure.

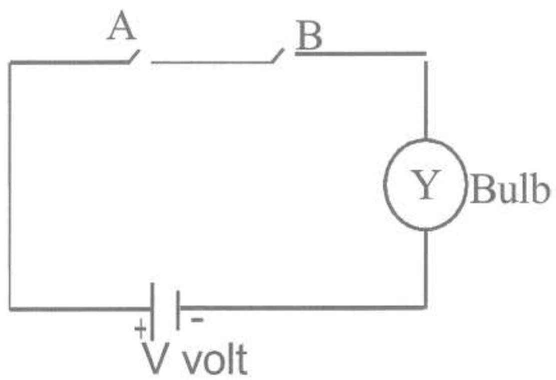


Figure 10: analogue electric circuit for AND gate

In this arrangement, off (or open) corresponds to 0 and on (or closed) corresponds to 1. The inputs are introduced through the switch A & B.

The lighting of the bulb is the output. Here we find that the bulb glow glows (i.e. output is 1) when both the switches A are B are closed (A=1, B=1). The bulb remains off (output is 0) when either switch A or switch B or both are open.

10.6.3 NOT Gate

The NOT Gate is the simplest of all logic gates as it has only one input & one output.
NOT gate is device which inverts input. If input is 1, the output is 0 & if input is 0 the output is 1.

Logic gate symbol is shown in figure. The bar symbol (¯) is referred to as NOT in Boolean algebra .The Boolean expression NOT operation is also called inversion.

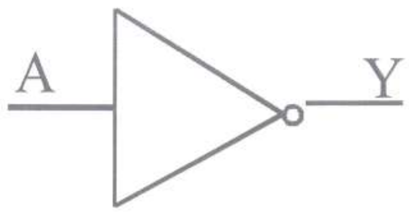


Figure 11: Symbol of NOT Gate

A NOT gate can be realized by the electronic circuit using n-p-n transistor as shown in figure (b).The base B of the transistor is connected to the input A through a resistance R_b and emitter E is earthed. The collector is connected to 5V battery. The output Y is the voltage at C w.r.t. Earth.

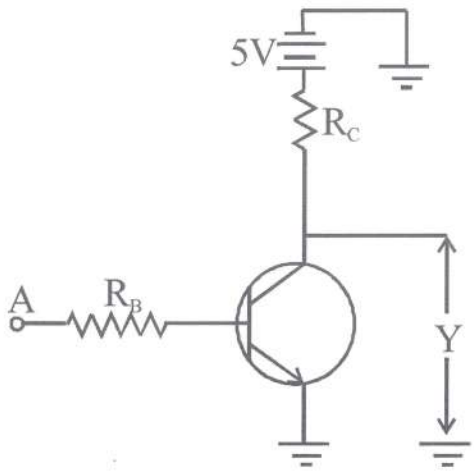


Figure 12: Circuit for NOT gate

The operation of the NOT gate can be understood by the following two cases.

(i) When A is earthed (A=0) then base-emitter junction is not in forward biased but base-collector junction is reversed biased. As the emitter current is zero, the base current is also zero and hence the collector will also be zero .Under these condition ,the transistor is said to be in cut off mode & voltage at C will be +5V w.r.t. earth due to the battery in collector circuit .Hence the output Y is 1

(ii) When A is connected to positive terminal of battery 5V (A=1), the base-emitter junction gets forward biased .There will be emitter current, base current & collector current. The value of resistor R_b & R_C are so adjusted that in this arrangement a large current flows. In this situation, the transistor is said to have gone to saturation state. The voltage drop across RC due to forward biasing of emitter is just equal to 5V.Hence the voltage at C is zero Volt. Therefore, the output Y is 0.

The truth table of NOT gate is shown in figure.

A	$Y = \overline{A}$
0	1
1	0

Figure 13: Truth Table for NOT Gate

The analogue electric circuit having function similar to the NOT gate is shown in figure.

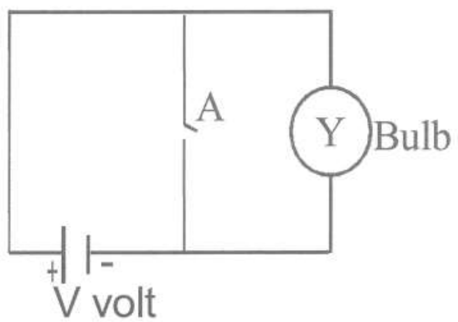


Figure 14: The analogue electric circuit for NOT gate

In this arrangement, off (or open) corresponds to 0 and on (or closed) corresponds to 1.The input is introduced through the switch A. The lighting of the bulb is output.

Here, we find that the bulb glows (I.e. output is 1) only when switch A is open or off (A=0) & bulb does not glow (i.e. output is 0) when switch A is closed (A=1)

10.6.4 NAND Gate (NOT +AND Gate)

If we connect output Y' of AND Gate to the i/p of NOT gate as shown in figure the gate so obtained is called NAND gate. The logic symbol of NAND Gate is shown in figure.

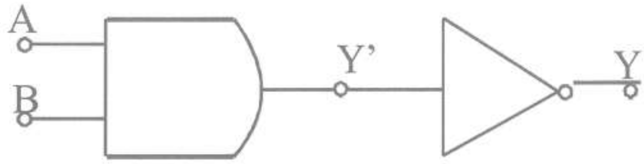


Figure 15: Connecting AND gate output to NOT gate input

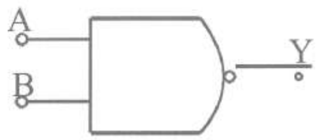


Figure 16: NAND Gate symbol

In Boolean expression, the NAND gate is expressed as, and is being read as 'A AND B negated'. The truth table of NAND gate is shown in figure.

A	B	Y'= A.B	$Y = \overline{A.B}$
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

Figure 17: Truth Table for NAND Gate

10.6.5 NOR Gate (NOT + OR Gate)

If we connect output Y' of OR Gate to the I/P of NOT gate as shown in figure the gate so obtained is not called NOR gate. The logic symbol of NOR Gate is shown in figure.

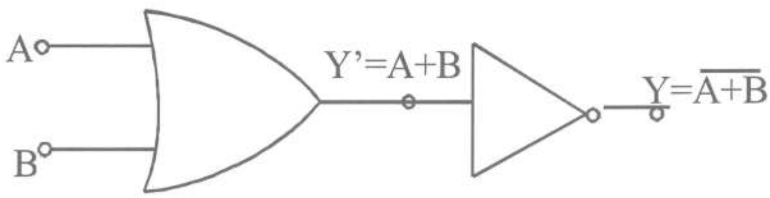


Figure 18: Connecting OR gate output to NOT gate input

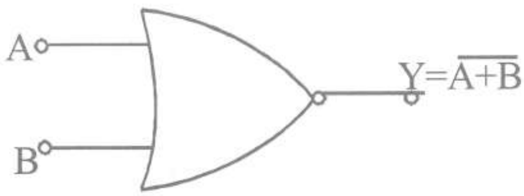


Figure 19: NOR Gate symbol

In Boolean expression, the NOR gate is expressed as $Y = (\overline{A+B})$, and is being read as 'A OR B negated'. The truth table of NOR gate is shown in figure.

A	B	$Y^1 = A+B$	$Y = \overline{A+B}$
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

Figure 20: Truth Table for NOR gate

NAND & NOR gate are known as universal gate.

10.6.6 Exclusive-OR gate (XOR Gate)

In an XOR gate, the output of a two-input XOR gate attains state 1 if one adds only input and attains state 1.

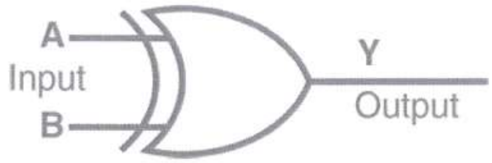


Figure 21: Logic symbol of Exclusive OR gate

The Boolean expression of the XOR gate is $Y= A.\bar{B}+\bar{A}B$ Or $Y=A\oplus B$

The truth table of an XOR gate is

A	B	$Y=A\oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

Figure 22: Truth Table for Exclusive OR gate

10.6.6 Exclusive-NOR Gate (XNOR Gate):

In the XNOR gate, the output is in state 1 when both inputs are the same, that is, both 0 and both 1.

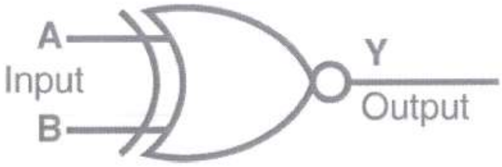


Figure 23: Logic symbol of Exclusive NOR Gate

The Boolean expression of the XNOR gate

$$Y = (\overline{A \oplus B}) = (A.B + \overline{A}.\overline{B})$$

The truth table of an XNOR gate is given below

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

Figure 24: Truth table for Exclusive NOR Gate

10.6 Application of Logic Gates

Logic gates have a lot of applications, but they are mainly based on their mode of operations or their truth table.

Basic logic gates are often found in circuits such as safety thermostats, push-button locks, automatic watering systems, light-activated burglar alarms and many other electronic devices.

One of the primary benefits is that basic logic gates can be used in various combinations if the operations are advanced. Besides, there is no limit to the number of gates that can be used in a single device.

However, it can be restricted due to the given physical space in the device. In digital integrated circuits (ICs), we will find an array of the logic gate area unit.

10.7 Summary

- As discussed in previous articles, the three basic logic gates are AND, OR and NOT.
- They perform basic multiplication (AND), addition (OR), and complement (NOT) functions.
- The NOT gate at the output of OR and AND gates form the logic NOR and NAND respectively.
- They output the complement of the original function performed by OR and AND gates. Moreover, NOR and NAND gates are classified as Universal Gates because basic logics i.e. AND, OR, and NOT can be built using either NOR or NAND gates only.
- The logic gates, discussed above, can be used to make a combinational logic circuit, and cascading them can also lead to having more than one input.

10.8 Keywords

- **Truth Table** – A truth table is a breakdown of all the possible truth values returned by a logical expression.
- **Boolean Algebra** – Boolean algebra is a branch of mathematics that deals with operations on logical values with binary variables
- **Boolean Expression** – A Boolean expression is a logical statement that is either TRUE or FALSE.

10.9 Review Questions

1. What is the use of basic logic gates?
2. What are the types of basic logic gates?
3. What are universal gates?
4. If a 3-input NOR gate has eight input possibilities, how many of those possibilities will result in a HIGH output?
5. Explain what is Boolean algebra?
6. Explain what are the basic logic elements?
7. Explain what is a truth table?
8. Explain what is pulse logic system?
9. What is an inverter?
10. Explain why is a two-input NAND gate called universal gate?

10.10 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication.
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education.
4. Millman, Electronics Devices and Circuits, TMH.
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi.
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India.

Contents

UNIT-11

Introduction to Communication Systems

Learning Outcomes

After studying this unit, the student will be able to:

1. Describe the basics of communications
2. Explain the concepts of frequencies in communications
3. Explain the significance and applications of various frequency bands
4. Explain IEEE spectrum of communication systems

The design and optimization of communication systems involve considerations of factors such as bandwidth, data rate, signal-to-noise ratio, modulation techniques, multiplexing, and transmission protocols. Communication engineers and researchers continuously work to improve system performance, increase transmission speeds, enhance reliability, and expand the capabilities of communication systems. In summary, communication systems form the backbone of our interconnected world, enabling the transfer of information across diverse mediums. These systems play a vital role in telecommunications, broadcasting, networking, and various other applications. Advances in technology continue to drive the evolution of communication systems, enhancing our ability to connect and communicate effectively.

11 Significance of frequencies in communication systems

Frequencies play a crucial role in modern communication systems, serving as the backbone for the transmission of information across various mediums. Communication systems rely on the use of specific frequency ranges to enable efficient and reliable transfer of data, voice, and video signals. These frequency bands, carefully allocated and regulated, provide the foundation for diverse communication technologies that shape our interconnected world. A wide range of frequencies is utilized in communication systems, each serving specific purposes and applications. The frequency spectrum encompasses an extensive span, ranging from extremely low frequencies in the hertz range to extremely high frequencies in the gigahertz range. Different bands within this spectrum are designated for various communication services and applications, including broadcasting, cellular networks, satellite communications, wireless networking, and more.

Various frequency bands have distinct characteristics, including propagation characteristics, transmission range, bandwidth availability, and susceptibility to interference. These factors influence the choice of frequencies for different communication systems based on factors such as coverage requirements, data transfer rates, signal quality, and regulatory considerations. The allocation and management of frequency bands are overseen by regulatory bodies and international agreements to ensure efficient spectrum utilization and minimize interference between different services. Spectrum planning and allocation are vital to ensure compatibility and coexistence of diverse communication systems. Understanding the various frequencies used in communication systems is essential for engineers, researchers, and professionals involved in designing, implementing, and maintaining communication networks. It enables them to make informed decisions regarding frequency selection, system design, interference mitigation, and overall performance optimization. In summary, the use of various frequencies in communication systems allows for the efficient transmission of data, voice, and video signals across diverse technologies and applications. These frequency bands are carefully allocated, regulated, and managed to ensure optimal performance, spectrum efficiency, and compatibility among different services. A comprehensive understanding of these frequencies is vital for the successful deployment and operation of modern communication systems. Various frequencies used in communication systems are given as follows with their frequency range:

Extremely Low Frequency (ELF): 3 Hz - 30 Hz

Super Low Frequency (SLF): 30 Hz - 300 Hz

Ultra Low Frequency (ULF): 300 Hz - 3 kHz

Very Low Frequency (VLF): 3 kHz - 30 kHz

Low Frequency (LF): 30 kHz - 300 kHz

Medium Frequency (MF): 300 kHz - 3 MHz

High Frequency (HF): 3 MHz - 30 MHz

Very High Frequency (VHF): 30 MHz - 300 MHz

Ultra-High Frequency (UHF): 300 MHz - 3 GHz

Super High Frequency (SHF): 3 GHz - 30 GHz

Extremely High Frequency (EHF): 30 GHz - 300 GHz

These frequency ranges are utilized for different types of communication systems, including radio broadcasting, television, cellular networks, satellite communications, Wi-Fi, Bluetooth, and many more. It's important to note that specific frequency bands within these ranges may be allocated for different purposes in different regions, countries, or organizations. Frequency allocation is managed by regulatory bodies to ensure efficient and interference-free communication. Please keep in mind that this is a general overview, and there may be additional specialized frequency bands used for specific applications or industries.

11.1 Extremely Low Frequency (ELF): 3 Hz - 30 Hz

Extremely Low Frequency (ELF) refers to the range of electromagnetic frequencies from 3 Hz to 30 Hz. It is the lowest frequency range in the electromagnetic spectrum. ELF signals have unique properties due to their extremely long wavelengths and are used in various applications for communication and research purposes.

Applications:

Submarine Communication: ELF waves can penetrate seawater to a significant depth, making them ideal for communication with submerged submarines. ELF signals can travel long distances through the water, allowing for reliable and secure communication between submarines and command centres.

Geophysical Research: ELF waves are used to study and monitor natural phenomena, such as lightning activity, atmospheric disturbances, and geomagnetic variations. By analysing ELF signals, scientists can gain insights into the Earth's electromagnetic environment and study the effects of natural events on the planet's atmosphere and geology. Power Transmission: ELF frequencies are utilized in power transmission systems to control and monitor the flow of electrical energy. These frequencies are used for power line carrier communication, enabling data transfer and control signals to be transmitted alongside the power lines.

Timekeeping: Highly precise atomic clocks, such as those used for global time synchronization, rely on ELF signals for long-range synchronization. ELF radio signals provide accurate timing references that are crucial for various applications, including telecommunications, satellite systems, and financial transactions.

Research and Experimentation: ELF frequencies are used in scientific research and experimentation, particularly in fields such as physics, astrophysics, and radio astronomy. Scientists use ELF waves to study the behavior of particles, plasma physics, and cosmic phenomena.

It's important to note that due to the extremely low frequency and long wavelength of ELF waves, the data rates achievable are very low. Therefore, ELF is not suitable for high-speed data communication but finds its niche in specialized applications where long-range propagation or unique properties of ELF waves are required.

In summary, Extremely Low Frequency (ELF) signals have applications ranging from submarine communication to geophysical research, power transmission, timekeeping, and scientific experimentation. The unique properties of ELF waves make them valuable in specific contexts where long-range propagation and penetration are desired.

11.2 Super Low Frequency (SLF): 30 Hz - 300 Hz

Super Low Frequency (SLF) refers to the range of electromagnetic frequencies from 30 Hz to 300 Hz. SLF signals have unique properties due to their extremely long wavelengths, and they find applications in specific areas where low-frequency communication and research are required.

Applications:

Submarine Communication: SLF signals are used for long-range communication with submarines. Due to their long wavelengths, SLF waves can penetrate seawater more effectively than higher frequency signals. This makes SLF communication valuable for maintaining contact with submerged submarines, enabling communication between command centres and submarines over great distances. Geophysical Research: SLF waves are utilized in geophysical research to study and monitor the Earth's ionosphere and magnetosphere. Scientists study the behaviour of SLF signals to gain insights into the effects of solar activity, geomagnetic storms, and atmospheric phenomena on the Earth's electromagnetic environment.

Lightning Research: SLF signals are employed in the study of lightning and related atmospheric phenomena. Researchers use SLF waves to investigate the characteristics and propagation of lightning discharges, helping to improve lightning detection systems, lightning protection measures, and atmospheric modelling. Power Transmission and Distribution: SLF frequencies are utilized in power systems for protective relaying and monitoring. These frequencies enable power utilities to detect faults, monitor power quality, and ensure stable and reliable operation of the electrical grid. Scientific Experiments: SLF waves are used in controlled scientific experiments, particularly in fields such as plasma physics, particle physics, and astrophysics. Researchers utilize SLF frequencies to study phenomena such as wave-particle interactions, plasma confinement, and cosmic ray detection.

It's important to note that due to the extremely low frequency of SLF waves, the data rates achievable are extremely low. Therefore, SLF is primarily used for specialized applications that require long-range propagation or specific characteristics of SLF waves. In summary, Super Low Frequency (SLF) signals find applications in submarine communication, geophysical research, lightning studies, power transmission and distribution, and scientific experiments. The unique properties of SLF waves make them valuable in these specialized contexts, where low-frequency communication and research are essential.

11.3 Ultra Low Frequency (ULF): 300 Hz - 3 kHz

Ultra Low Frequency (ULF) refers to the range of electromagnetic frequencies from 300 Hz to 3 kHz. ULF signals have extremely long wavelengths and find applications in various fields where low-frequency communication and research are required.

Applications:

Geophysical Research: ULF waves are extensively used in geophysical research to study and monitor natural phenomena such as Earth's magnetic field, ionospheric disturbances, and seismic activity. ULF signals are sensitive to changes in the Earth's electromagnetic environment, providing valuable insights into geophysical processes and contributing to the understanding of Earth's dynamics.

Submarine Communication: ULF waves are utilized for long-range communication with submarines. Due to their long wavelengths, ULF signals can penetrate seawater efficiently, allowing reliable communication with submerged submarines even at significant depths. ULF communication provides a means for maintaining contact between submarines and command centres.

Earthquake Precursors: ULF signals are studied for their potential as earthquake precursors. Researchers analyse ULF waves for anomalous variations or disturbances that could indicate the build-up or occurrence of seismic events. Monitoring ULF signals may contribute to the development of early warning systems for earthquakes. Magnetospheric Research: ULF waves play a crucial role in the study of Earth's magnetosphere and its interactions with the solar wind. ULF signals are used to investigate phenomena such as magnetospheric substorms, wave-particle interactions, and plasma dynamics in the near-Earth space environment.

Lightning Detection: ULF waves are employed in lightning detection systems. By monitoring ULF signals, it is possible to detect and locate lightning discharges, providing valuable data for severe weather monitoring, lightning safety, and atmospheric research. Medical Applications: ULF magnetic fields are utilized in medical imaging techniques such as magnetoencephalography (MEG) and magnetocardiography (MCG). These techniques capture the extremely weak magnetic fields generated by brain activity and heart functions, enabling non-invasive diagnostics and research in neurology and cardiology.

ULF signals possess unique properties that make them well-suited for specific applications in geophysical research, submarine communication, earthquake monitoring, magnetospheric studies, lightning detection, and medical imaging. In summary, Ultra Low Frequency (ULF) signals are utilized in geophysical research, submarine communication, earthquake precursors, magnetospheric studies, lightning detection, and medical applications. The long wavelengths of ULF waves enable them to penetrate various mediums and provide valuable insights into Earth's dynamics and other phenomena.

11.4 Very Low Frequency (VLF): 3 kHz - 30 kHz

Very Low Frequency (VLF) refers to the range of electromagnetic frequencies from 3 kHz to 30 kHz. VLF signals have long wavelengths and find applications in various fields, particularly in long-range communication and research.

Applications:

Submarine Communication: VLF waves are extensively used for long-range communication with submarines. Due to their ability to penetrate seawater and propagate over large distances, VLF signals allow reliable communication with submerged submarines. VLF communication systems enable command centres to maintain contact with submarines even when they are deep underwater. Navigation Systems: VLF signals are employed in various navigation systems, such as the Omega Navigation System and the Alpha Navigation System. These systems utilize VLF waves to determine the position of ships and aircraft accurately. VLF navigation is particularly useful in maritime applications where precise positioning is required.

Geophysical Research: VLF waves are used in geophysical research to study the Earth's ionosphere and magnetosphere. Researchers analyse VLF signals to investigate the behaviour of the ionosphere, the interaction between the Earth's magnetic field and the solar wind, and the propagation characteristics of VLF waves in the atmosphere. Time and Frequency Standards: VLF signals are utilized for time synchronization and frequency calibration. The Global Positioning System (GPS) employs VLF frequencies to synchronize the highly accurate atomic clocks used for global timekeeping. VLF time signals are also used by various organizations and laboratories as a reference for maintaining precise time and frequency standards.

Subsurface Imaging: VLF electromagnetic methods are used for subsurface imaging and exploration. By analysing the VLF response of the Earth's subsurface, geophysicists can detect and map geological structures, groundwater resources, and archaeological features. Lightning Research: VLF signals are utilized in the study of lightning and atmospheric phenomena. Researchers analyse VLF waves to investigate lightning discharges, study their characteristics, and monitor their occurrence. VLF data helps improve our understanding of thunderstorm dynamics and contributes to lightning detection and monitoring systems.

In summary, Very Low Frequency (VLF) signals find applications in submarine communication, navigation systems, geophysical research, time synchronization, subsurface imaging, and lightning research. The long wavelengths and propagation characteristics of VLF waves make them valuable in long-range communication, precise positioning, scientific research, and various practical applications.

11.5 Low Frequency (LF): 30 kHz - 300 kHz

Low Frequency (LF) refers to the range of electromagnetic frequencies from 30 kHz to 300 kHz. LF signals have relatively long wavelengths and find applications in various fields due to their unique propagation characteristics and suitability for specific communication and research purposes.

Applications:

Radio Broadcasting: LF frequencies are used for long-distance radio broadcasting. LF radio stations provide coverage over large areas by utilizing the ground wave propagation, which allows signals to follow the curvature of the Earth's surface. LF broadcasting is commonly used for transmitting navigation beacons, weather information, and emergency broadcasts. Industrial and Military Communications: LF signals are utilized for long-range communication in industrial and military applications. LF communication systems offer reliable and secure communication over large distances, making them suitable for applications such as military communications, remote monitoring, and control of industrial facilities. Geophysical Exploration: LF electromagnetic methods are employed in geophysical exploration to study subsurface structures and resources. By analysing the LF response of the Earth's subsurface, geophysicists can detect underground features such as mineral deposits, oil and gas reservoirs, and geological formations.

Wireless Power Transfer: LF frequencies are used for wireless power transfer applications. By utilizing resonant coupling or electromagnetic induction, LF signals can transfer power wirelessly over short distances. LF wireless power transfer finds applications in charging systems for electric vehicles, wireless charging pads, and biomedical devices. Underground Communication: LF signals can propagate through the ground, making them useful for underground communication systems. Applications include communication in mining operations, underground tunnels, and subterranean environments where conventional wireless signals may be obstructed. Research and Experiments: LF frequencies are utilized in scientific research and experiments, particularly in fields such as plasma physics, particle accelerators, and radio astronomy. Researchers use LF signals to investigate phenomena such as wave-particle interactions, plasma confinement, and astronomical radio observations.

In summary, Low Frequency (LF) signals find applications in radio broadcasting, industrial and military communications, geophysical exploration, wireless power transfer, underground communication, and scientific research. The unique propagation characteristics and long wavelengths of LF waves make them suitable for long-range communication, subsurface exploration, and specialized research purposes.

11.6 Medium Frequency (MF): 300 kHz - 3 MHz

Medium Frequency (MF) refers to the range of electromagnetic frequencies from 300 kHz to 3 MHz's MF signals have wavelengths that allow for efficient ground wave propagation as well as limited sky wave propagation, making them suitable for various applications in communication, broadcasting, and navigation.

Applications:

AM Radio Broadcasting: MF frequencies are widely used for AM (Amplitude Modulation) radio broadcasting. AM radio stations utilize MF signals to transmit audio content over long distances. The ground wave propagation of MF signals allows for reliable coverage within a region, making MF AM radio a popular medium for news, music, and talk shows.

Aviation Communication: MF frequencies are utilized for aviation communication, specifically for communication between aircraft and air traffic control (ATC) towers. MF signals provide reliable and clear communication over moderate distances, facilitating air traffic control instructions, weather updates, and flight coordination. Navigation Systems: Medium Frequency signals are used in various navigation systems, such as Non-Directional Beacons (NDBs) and LORAN-C (Long-Range Aid to Navigation). NDBs transmit MF signals that can be detected by aircraft and ships, providing navigational aids for determining position and course. LORAN-C systems utilize timed MF signals from multiple stations to enable accurate navigation over long distances. Two-Way Radio Communication: MF frequencies are employed in certain two-way radio communication systems, such as Citizens Band (CB) radio. CB radios operate within the MF range and are used for short-range communication in various applications, including trucking, outdoor activities, and emergency communication.

Marine Communication: MF frequencies are used for maritime communication, especially for ship-to-ship and ship-to-shore communication. MF radio systems enable vessels to communicate essential information, such as distress signals, safety messages, and navigational updates. Industrial Communication: MF frequencies find applications in industrial settings for communication within large facilities, such as manufacturing plants, warehouses, and construction sites. MF radio systems provide reliable communication over moderate distances, facilitating coordination, safety protocols, and efficient operations.

In summary, Medium Frequency (MF) signals are utilized in AM radio broadcasting, aviation communication, navigation systems, two-way radio communication, marine communication, and industrial communication. The propagation characteristics of MF waves allow for reliable communication over moderate distances, making them suitable for various applications in broadcasting, transportation, navigation, and industrial settings.

11.7 High Frequency (HF): 3 MHz - 30 MHz

High Frequency (HF) refers to the range of electromagnetic frequencies from 3 MHz to 30 MHz. HF signals are characterized by their ability to travel long distances through the Earth's ionosphere, allowing for reliable long-range communication in various applications. HF waves are widely used in radio communication, broadcasting, and other fields.

Applications:

Shortwave Broadcasting: HF frequencies are extensively used for international shortwave broadcasting. HF shortwave stations transmit signals that can be received globally, allowing for long-distance communication and broadcasting. HF shortwave broadcasts are utilized for international news, entertainment, and cultural exchange. Amateur Radio: HF bands are allocated for amateur radio operators, also known as ham radio operators. Amateur radio enthusiasts use HF frequencies to communicate with other operators around the world, participate in contests, and provide emergency communication support during disasters. Aviation Communication: HF frequencies are utilized for long-range communication in aviation. HF radios are installed on aircraft for communication over large expanses of oceanic airspace, where other communication options may be limited. HF communication enables pilots to maintain contact with air traffic control (ATC) and receive important weather updates.

Military Communication: HF signals play a vital role in military communication, particularly for long-range and secure communication. HF radios are used by military organizations to establish reliable communication links in various operational environments, including remote areas, maritime operations, and battlefield scenarios. Ionospheric Research: HF frequencies are employed in scientific research to study the Earth's ionosphere. Researchers use HF waves to probe and analyse the behaviour of the ionosphere, investigating phenomena such as ionospheric disturbances, propagation effects, and space weather impacts.

Over-The-Horizon Radar: HF signals are utilized in over-the-horizon (OTH) radar systems. OTH radar employs HF frequencies to detect and track objects beyond the line of sight, including aircraft, ships, and ballistic missiles. HF waves' ability to reflect off the ionosphere enables long-range radar surveillance. Point-to-Point Communication: HF frequencies are used for long-range point-to-point communication, particularly in remote areas where other communication infrastructure may be limited. HF radios provide reliable communication links for applications such as offshore platforms, remote monitoring stations, and expeditionary operations.

In summary, High Frequency (HF) signals find applications in shortwave broadcasting, amateur radio, aviation communication, military communication, ionospheric research, over-the-horizon radar, and long-range point-to-point communication. The unique propagation characteristics of HF waves allow for long-distance communication, making them valuable in various fields, including broadcasting, aviation, scientific research, and defence.

11.8 Very High Frequency (VHF): 30 MHz - 300 MHz

Very High Frequency (VHF) refers to the range of electromagnetic frequencies from 30 MHz to 300 MHz VHF signals have shorter wavelengths compared to lower frequency bands and are known for their ability to travel in a straight line and provide line-of-sight communication. VHF frequencies are widely used in various applications due to their favourable propagation characteristics and versatility.

Applications:

FM Radio Broadcasting: VHF frequencies are extensively used for FM (Frequency Modulation) radio broadcasting. FM radio stations operate within the VHF range, providing high-quality audio transmission with good signal clarity. VHF FM radio is popular for music, news, sports, and entertainment broadcasting. Television Broadcasting: VHF frequencies are utilized for over-the-air television broadcasting. VHF TV channels are allocated for transmitting television signals to home receivers. VHF TV broadcasts cover a wide area and are commonly used for local television stations.

Two-Way Radio Communication: VHF frequencies are widely used for two-way radio communication, including public safety communications, maritime communication, aviation communication, and business radio systems. VHF two-way radios provide reliable communication over moderate distances and are employed by various industries, emergency services, and organizations. Mobile Communication: VHF frequencies are used in certain mobile communication systems, such as maritime mobile service and land mobile radio systems. VHF mobile radios are used by maritime vessels, land-based vehicles, and public transportation systems to enable communication between mobile units and base stations.

Air Traffic Control: VHF frequencies are critical for air traffic control (ATC) communication. VHF radios are used by pilots, air traffic controllers, and ground services to ensure safe and efficient aircraft operations. VHF ATC communication facilitates instructions, clearance, and coordination between aircraft and ground control.

Weather Monitoring: VHF frequencies are utilized for weather monitoring and meteorological data transmission. VHF radiosondes are launched into the atmosphere to collect weather data and transmit it back to ground-based weather stations. VHF signals are also used for aircraft weather radar systems to detect and track weather patterns.

Wireless Microphones: VHF frequencies are commonly used for wireless microphone systems in various applications, including live performances, broadcasting, conferences, and events. VHF wireless microphones offer reliable and interference-free audio transmission within a limited range. In summary, Very High Frequency (VHF) signals find applications in FM radio broadcasting, television broadcasting, two-way radio communication, mobile communication, air traffic control, weather monitoring, and wireless microphone systems. The favorable propagation characteristics of VHF waves, including line-of-sight communication and reliable transmission, make them versatile and widely used in communication, broadcasting, and other fields.

11.9 Ultra-High Frequency (UHF): 300 MHz - 3 GHz

Ultra-High Frequency (UHF) refers to the range of electromagnetic frequencies from 300 MHz to 3 GHz. UHF signals have even shorter wavelengths compared to VHF frequencies and offer unique advantages in terms of signal penetration and bandwidth capacity. UHF is widely used in various applications due to its versatility, wide bandwidth availability, and suitability for both short-range and long-range communication.

Applications:

Television Broadcasting: UHF frequencies are extensively used for over-the-air television broadcasting. UHF TV channels provide higher bandwidth capacity compared to VHF, allowing for the transmission of high-definition (HD) and digital television signals. UHF TV broadcasts cover both local and regional television stations. Wireless Communication: UHF frequencies are widely utilized in wireless communication systems, including cellular networks, wireless data transmission, and wireless local area networks (WLANs). UHF bands are utilized by technologies such as 4G and 5G networks, enabling high-speed data transfer and mobile connectivity.

Two-Way Radio Communication: UHF frequencies are commonly used for two-way radio communication, including public safety communication, professional radio systems, and private radio networks. UHF two-way radios provide reliable communication over both short and moderate distances and are widely used by emergency services, law enforcement, construction sites, and event management. Wireless Microphones: UHF frequencies are commonly used for wireless microphone systems in various applications, including live performances, broadcasting, conferences, and events. UHF wireless microphones offer greater frequency flexibility, reduced interference, and longer transmission range compared to lower frequency bands.

Satellite Communication: UHF frequencies are used in satellite communication systems, including satellite broadcasting and satellite phones. UHF signals are utilized for uplink and downlink communication with satellites, providing global coverage for broadcasting, voice, and data transmission. Radio Astronomy: UHF frequencies play a crucial role in radio astronomy, allowing scientists to observe and study celestial objects and phenomena. UHF radio telescopes capture and analyse electromagnetic signals from space, providing valuable insights into the universe, such as the study of distant galaxies and cosmic microwave background radiation.

RFID and NFC Technology: UHF frequencies are employed in RFID (Radio Frequency Identification) and NFC (Near Field Communication) technologies. UHF RFID tags and readers are used for asset tracking, inventory management, access control, and contactless payment systems. UHF NFC technology enables secure and convenient communication between devices in close proximity.

In summary, Ultra High Frequency (UHF) signals find applications in television broadcasting, wireless communication, two-way radio communication, wireless microphone systems, satellite communication, radio astronomy, and RFID/NFC technology. The shorter wavelengths and wide bandwidth availability of UHF make it suitable for various communication and data transmission needs in both short-range and long-range applications.

11.10 Super High Frequency (SHF): 3 GHz - 30 GHz

Super High Frequency (SHF) refers to the range of electromagnetic frequencies from 3 GHz to 30 GHz. SHF signals have even shorter wavelengths compared to UHF frequencies and offer increased bandwidth capacity and data transfer rates. SHF is utilized in various applications that require high-speed and high-capacity communication, as well as for specialized scientific and industrial purposes.

Applications:

Satellite Communication: SHF frequencies are extensively used in satellite communication systems, including satellite broadcasting, broadband internet, and satellite phones. SHF signals allow for high-speed data transmission and enable global coverage for voice, video, and data services.

Wireless Communication: SHF frequencies are utilized in wireless communication systems, including Wi-Fi networks, point-to-point microwave links, and wireless backhaul for cellular networks. SHF bands provide higher data transfer rates and increased capacity, making them suitable for applications that require high-bandwidth connectivity.

Radar Systems: SHF frequencies are used in various radar systems, including weather radar, air traffic control radar, and military radar. SHF radar systems offer high-resolution imaging, accurate target detection, and enhanced performance in adverse weather conditions. Radio Astronomy: SHF frequencies are crucial in radio astronomy, enabling scientists to observe and study celestial objects and phenomena with high precision. SHF radio telescopes provide detailed insights into cosmic microwave background radiation, quasars, pulsars, and other astronomical phenomena.

Microwave Oven: Microwave ovens utilize SHF frequencies to generate electromagnetic waves that heat and cook food. The specific frequency used in microwave ovens is around 2.45 GHz, which allows for efficient absorption of energy by water molecules in the food. Remote Sensing: SHF frequencies are employed in remote sensing applications, including Earth observation satellites and ground-based remote sensing systems. SHF remote sensing provides valuable data for monitoring weather patterns, studying environmental changes, and mapping Earth's surface features. Automotive Radar: SHF frequencies are used in automotive radar systems for collision avoidance and adaptive cruise control. SHF radar sensors detect and track objects around vehicles, providing critical information for driver assistance and safety systems. Medical Imaging: SHF frequencies are utilized in medical imaging technologies such as magnetic resonance imaging (MRI). SHF waves are used to excite and manipulate the nuclei of atoms in the body, allowing for detailed imaging and diagnosis of medical conditions.

In summary, Super High Frequency (SHF) signals find applications in satellite communication, wireless communication, radar systems, radio astronomy, microwave ovens, remote sensing, automotive radar, and medical imaging. The shorter wavelengths and increased bandwidth capacity of SHF enable high-speed data transfer, precise imaging, and advanced communication capabilities in various fields, including telecommunications, astronomy, healthcare, and scientific research.

11.11 Extremely High Frequency (EHF): 30 GHz - 300 GHz

Extreme High Frequency (EHF) refers to the range of electromagnetic frequencies from 30 GHz to 300 GHz. EHF signals have extremely short wavelengths and offer exceptional bandwidth capacity, enabling high-speed data transfer and communication. EHF is utilized in various applications that require ultra-fast and high-capacity wireless communication, as well as in scientific and industrial sectors that rely on precise sensing and imaging.

Applications:

Wireless Communication: EHF frequencies are used in wireless communication systems, including Wi-Fi networks, 5G and beyond cellular networks, and wireless backhaul. EHF bands provide massive data transfer rates and enable ultra-low latency communication, supporting emerging technologies such as autonomous vehicles, Internet of Things (IoT), and high-definition streaming.

Terrestrial and Satellite Communication: EHF frequencies are utilized in point-to-point microwave links for long-distance communication, such as high-speed data transmission between remote locations and satellite communication. EHF signals allow for high-capacity connectivity and data transfer over long distances. Radio Astronomy: EHF frequencies play a vital role in radio astronomy, enabling scientists to study the universe with unprecedented detail. EHF radio telescopes provide insights into cosmic microwave background radiation, galaxy formation, and other astrophysical phenomena.

Remote Sensing and Imaging: EHF frequencies are employed in remote sensing applications, including Earth observation satellites and ground-based sensing systems. EHF remote sensing allows for precise imaging of Earth's surface, atmospheric monitoring, and the detection of atmospheric constituents.

Millimetre-Wave Scanners: EHF frequencies, particularly in the millimetre-wave range, are used in security and screening applications. Millimetre-wave scanners can detect concealed objects on individuals, making them valuable for airport security, border control, and high-security facilities.

Automotive Radar: EHF frequencies, specifically in the millimetre-wave range, are used in advanced automotive radar systems for collision avoidance, blind-spot detection, and autonomous driving. Millimetre-wave radar sensors provide high-resolution imaging and accurate detection of objects in the surroundings of a vehicle.

Industrial and Scientific Applications: EHF frequencies find applications in various industrial and scientific sectors, including spectroscopy, material characterization, and industrial process control. EHF waves can be used to analyse and measure the properties of materials, study chemical reactions, and perform non-destructive testing. Medical Imaging: EHF frequencies, particularly in the terahertz range, are explored for medical imaging applications. Terahertz imaging has the potential to provide detailed imaging and diagnosis of biological tissues while being non-ionizing and safe for medical use.

In summary, Extreme High Frequency (EHF) signals find applications in wireless communication, terrestrial and satellite communication, radio astronomy, remote sensing, millimetre-wave scanners, automotive radar, industrial and scientific applications, and medical imaging. The extremely short wavelengths and wide bandwidth capacity of EHF enable ultra-fast data transfer, precise sensing, and advanced communication capabilities in various fields, shaping the future of wireless connectivity and technological advancements.

11.12 Various Communication bands in Communication Systems

There are various bands for specific purpose communication having a range lower to higher frequency depending on the need of communications. They are following:

AM Broadcast Band: 535 kHz - 1605 kHz

FM Broadcast Band: 88 MHz - 108 MHz

Television VHF Band: 54 MHz - 88 MHz and 174 MHz - 216 MHz

Television UHF Band: 470 MHz - 806 MHz

GSM Cellular Band: 824 MHz - 960 MHz

3G Cellular Band: 1710 MHz - 2170 MHz

4G LTE Cellular Band: 700 MHz - 2600 MHz

Wi-Fi Band: 2.4 GHz and 5 GHz

Bluetooth Band: 2.4 GHz

GPS Band: 1.2276 GHz and 1.57542 GHz

Satellite Communication C Band: 3.7 GHz - 4.2 GHz

Satellite Communication Ku Band: 10.7 GHz - 14.5 GHz

Radar Band: Various frequency ranges, including X-band, S-band, C-band, and K-band

These are some of the commonly used frequency bands in communication systems, including radio broadcasting, television, cellular networks, satellite communications, wireless networking, and radar systems. Please note that the specific frequency ranges and band allocations can vary based on regional regulations, international agreements, and technological advancements. It's important to consult the relevant regulatory authorities or industry standards for the most accurate and up-to-date information regarding frequency bands.

11.13 IEEE Spectrum in Communication Systems

The IEEE Spectrum is a publication produced by the Institute of Electrical and Electronics Engineers (IEEE), one of the world's leading professional organizations dedicated to advancing technology. The IEEE Spectrum focuses on reporting and analysing the latest developments, trends, and advancements in various fields of technology, including communication systems.

The IEEE Spectrum serves as a valuable resource for professionals, researchers, and technology enthusiasts by providing in-depth articles, analysis, and insights into emerging technologies, industry trends, and scientific research in the field of communication systems. It covers a wide range of topics, including wireless communication, network protocols, signal processing, data transmission, and emerging communication technologies.

The publication aims to keep readers informed about the latest innovations, breakthroughs, and challenges in communication systems, offering a platform for experts to share their knowledge and expertise. It features articles, interviews, case studies, and technical papers written by industry experts, researchers, and thought leaders.

The IEEE Spectrum not only covers the theoretical aspects of communication systems but also provides practical applications and real-world implementations. It explores topics such as 5G and beyond, Internet of Things (IoT), wireless networking, optical communication, satellite communication, and many other areas relevant to modern communication systems.By staying up-to-date with the IEEE Spectrum, professionals and enthusiasts in the field of communication systems can gain valuable insights into the latest technologies, standards, research, and industry trends. It serves as a platform for knowledge exchange and helps foster collaboration and innovation in the field of communication systems. Overall, the IEEE Spectrum plays a vital role in disseminating information and promoting advancements in communication systems, contributing to the growth and development of the industry and enabling the creation of more efficient, reliable, and advanced communication technologies.

11.14 Summary

In this discussion, we explored the significance of various frequencies and bands used in communication systems. We started by introducing the concept of frequency bands and their importance in facilitating efficient and reliable communication. We covered a range of frequency bands, starting from Extremely Low Frequency (ELF) to Super High Frequency (SHF) and Extreme High Frequency (EHF). Each frequency band offers unique characteristics and finds applications in different fields. We discussed the applications of each band, highlighting their use in specific communication systems and scientific research. For example, ELF is used for submarine communication, while SHF is employed in satellite communication and radar systems. EHF, on the other hand, is utilized in wireless communication and industrial applications.

We also provided an introduction to communication bands, which are specific ranges of electromagnetic frequencies allocated for communication purposes. These bands are regulated and standardized to ensure coordinated use of the radio spectrum. The allocation of communication bands takes into account factors such as signal propagation, interference considerations, and available spectrum. Furthermore, we discussed the IEEE Spectrum, a publication by the Institute of Electrical and Electronics Engineers (IEEE), which focuses on reporting the latest developments and advancements in technology. The IEEE Spectrum serves as a valuable resource for professionals in the field of communication systems, providing insights into emerging technologies, industry trends, and scientific research.

Overall, understanding the various frequencies and bands used in communication systems is crucial for designing efficient and reliable communication networks. The allocation and utilization of these frequencies and bands play a vital role in shaping the capabilities and applications of communication systems in various sectors, including telecommunications, broadcasting, aviation, scientific research, and industrial applications.

11.15 Keywords

- Frequency bands
- Communication systems
- Electromagnetic frequencies
- Signal propagation
- Interference considerations
- Spectrum allocation
- Wireless communication
- Satellite communication
- Radar systems
- Radio astronomy
- Submarine communication
- Industry standards
- IEEE Spectrum
- Technology advancements
- Scientific research

11.16 Review Questions

- What are the main factors considered in allocating frequency bands for communication systems?
- How do different frequency bands, such as VHF and UHF, differ in their applications?
- What are the advantages and limitations of using higher frequency bands in communication systems?
- How do frequency bands affect the range and coverage of wireless communication systems?
- What are the specific applications of ELF, VLF, and MF bands in communication systems?
- How does the IEEE Spectrum contribute to the understanding and advancement of communication systems?
- What are the challenges and considerations in managing and coordinating the use of different frequency bands?
- How does the choice of frequency band impact the data transfer rate and capacity of a communication system?
- Can frequency bands be repurposed or shared between different communication systems? What are the implications?
- How does the allocation of frequency bands for communication systems vary across different countries or regions?

11.17 References

Rappaport, T. S. (2015). Wireless Communications: Principles and Practice (2nd ed.). Pearson.

This book provides comprehensive coverage of wireless communication principles, including discussions on frequency bands and their applications.

Federal Communications Commission (FCC). (n.d.). Frequency Allocation. Retrieved from https://www.fcc.gov/wireless/bureau-divisions/mobility-division/frequency-coordination/frequency-allocation

The FCC's website provides information on frequency allocation, including the frequency bands used in communication systems in the United States.

ITU-R. (2019). Recommendation ITU-R M.1732: Characteristics of systems operating in the frequency range 8.3 kHz to 3000 GHz. International Telecommunication Union.

This ITU-R recommendation provides detailed information on the characteristics and applications of various frequency bands used in communication systems.

IEEE Spectrum. (n.d.). Retrieved from https://spectrum.ieee.org/

The IEEE Spectrum is a valuable resource for staying up-to-date with the latest developments and advancements in communication systems, providing insights into emerging technologies and industry trends.

Goldsmith, A. (2005). Wireless Communications. Cambridge University Press.

This book offers a comprehensive overview of wireless communication systems, including discussions on frequency bands, spectrum allocation, and various communication technologies.

Contents

Learning Objectives

After studying this unit the student will be able to:

1 Describe Types of Communication.

2 Explain the concepts Communication.

3 Discuss Amplitude and frequency Modulation.

4 Explain equations of Amplitude and frequency Modulation.

Introduction

Communication systems are an essential aspect of modern life, allowing us to connect with each other across vast distances and share information in real-time. These systems rely on a combination of hardware, software, and network infrastructure to enable seamless communication between individuals and organizations around the world.

At their core, communication systems are designed to facilitate the transfer of information from one location to another. This can involve transmitting voice, video, or data signals, depending on the specific application. Communication systems can be broadly categorized into two main types: analog and digital.

Analog communication systems use continuous signals that vary in amplitude or frequency to transmit information. These systems have been used for many years and are still in use today in some applications. For example, analog radio broadcasts use variations in amplitude to transmit sound signals, while analog television broadcasts use variations in both amplitude and frequency to transmit video signals.

Digital communication systems, on the other hand, use discrete signals that represent information in a binary format, either as a 0 or a 1. These signals are transmitted over a network, and the receiving device decodes the binary signal to reconstruct the original information. Digital communication systems offer several advantages over analog systems, including improved signal quality, greater reliability, and the ability to transmit larger amounts of data.

The components of a communication system can be divided into three main categories: the transmitter, the channel, and the receiver. The transmitter is responsible for encoding the information into a signal that can be transmitted over the channel. The channel is the medium through which the signal is transmitted, and can be a physical medium such as a wire or fiber optic cable, or a wireless medium such as radio waves or infrared light. The receiver is responsible for decoding the signal and reconstructing the original information.

In addition to these basic components, communication systems also rely on a range of supporting technologies and infrastructure. For example, networks are used to connect multiple devices and allow them to communicate with each other. Network protocols and standards are used to ensure that devices can communicate with each other using a common language, and to ensure that data is transmitted securely and efficiently.

learning. The magazine provides in-depth analysis, reports, and interviews on these topics, helping readers understand the technical details and implications of the latest developments.

IEEE Spectrum also covers key issues and trends in communication technology. The magazine examines the social, political, and economic impact of new communication technologies, such as the potential benefits and risks of 5G, the challenges and opportunities presented by the proliferation of IoT devices, and the impact of cloud computing on network architecture and performance. This analysis helps readers understand the broader context of communication technology and make informed decisions about its adoption and use.

Another important aspect of IEEE Spectrum in communication is its role in promoting the work of the IEEE and its members. The magazine regularly features articles about the work of IEEE members in the area of communication, highlighting their achievements and contributions to the field. This helps to raise awareness of the important work done by IEEE members and to promote the IEEE as a leading professional organisation in communication.

In addition to its print and digital publications, IEEE Spectrum also provides a variety of online resources related to communication technology. The magazine offers webinars, podcasts, and videos that provide additional insights and information on key communication topics. These resources are designed to help readers stay informed and up-to-date on the latest developments in communication technology.

Overall, IEEE Spectrum is a valuable resource for anyone interested in communication technology. Its authoritative and insightful articles, along with its global reach and focus on cutting-edge research, make it an essential read for professionals in the field. Whether you are an engineer, researcher, policymaker, or just a curious reader, IEEE Spectrum in Communication is a publication that is well worth exploring.

12.2 Types of Communication Systems

12.2.1 Depending on signal specification or technology

The communication system is classified as follows:

Analogue

Analogue technology communicates data as electronic signals of varying frequency or amplitude. Broadcast and telephone transmission are common examples of analogue technology.

Digital

In digital technology, the data are generated and processed in two states: High (represented as 1) and low (represented as 0). Digital technology stores and transmits data in the form of 1s and 0s.

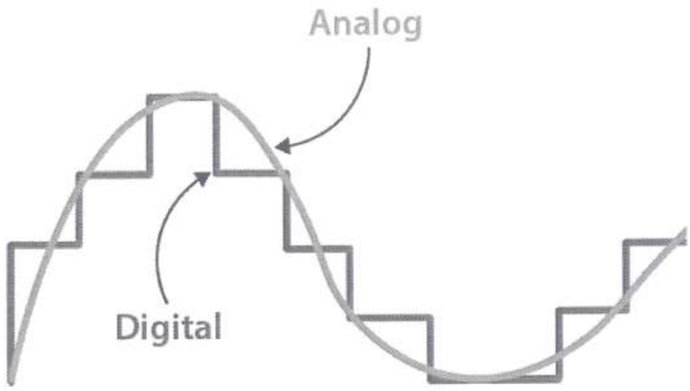


Figure 1: Analog and digital signal

12.2.2 Depending on the communication channel

The communication system is categorized as follows:

Wired (Line communication)

- Parallel wire communication
- Twisted wire communication
- Coaxial cable communication

- Optical fiber communication

Wireless (Space communication)

- Ground wave communication
- Sky wave communication
- Space wave communication
- Satellite communication

12.3 Modulation

Modulation is the phenomenon of superimposing the message signal of low frequency on a high frequency wave called, carrier wave) .The resulting wave is called the modulated wave, which is transmitted .
A message signal usually spreads over a range frequencies , called the signal band width .That is why message signal are also called base band signal, representing the band of frequencies of original signal.

Suppose we have to transmit an electrical signal of range 20 Hz to 20 kHz over a long distance. We cannot do it, as such following reasons –

1. Size of Antenna: An antenna is needed for both transmission & reception .Each antenna should have a size comparable to the wavelength of signal, at least λ/4 in size, so that time variation of signal is properly sensed by antenna.
   For an audio frequency signal of frequency 20 kHz
   $$\lambda = \frac{3x10^8}{15x10^3} = 20000m$$
   Therefore the length of antenna
   $$\frac{\lambda}{4} = \frac{2000}{4} = 5000m = 5km$$
   To set an antenna of vertical height 5 km which is practically impossible to construct & operate.
   If transmission frequency were raised to 1 MHz, then

$$\lambda = \frac{3x10^8}{10^6} = 300\text{m}$$

the length of antenna would be 300/4=75m, which is reasonable. Therefore, there is an urgent need of converting the low frequency message signal into high frequency signal before transmission.

2. Effective Power radiated by antenna: The power P radiated from liner antenna of length L is proportional to

$$(L/\lambda)^2 \quad \text{i.e.} \quad P \, \alpha \, \frac{1}{\lambda^2}$$

As the high power is needed for good transmission, therefore, for a given length, wavelength should be small or frequency should be high. Therefore high frequency signal is required for transmission.

3. Mixing up of signal from different transmitters: When many transmitters are transmitting baseband signal simultaneously, they get mixed & there is no way to distinguish between them. The possible solution is, communication at high frequency & allotting a band of frequencies to each transmitter so that there is no mixing.

All the three reasons explained above suggest that there is a need for transmission at high frequencies. This is done by modulation

Modulation of carrier wave can be done by three ways

12.3.1 Amplitude Modulation:

In amplitude modulation, the amplitude of carrier is varied in accordance with the instantaneous value of the audio frequency modulating signal (message signal), where as the frequency & phase angle remains same.

12.3.2 Frequency modulation:

In frequency modulation, frequency of carrier wave change according to instantaneous value of modulating signal, where as amplitude remains same.

12.3.3 Phase modulation:

In phase modulation the phase angle of carrier wave change according to phase angle of modulating signal , where as amplitude remain same.

12.4 Amplitude Modulation

Amplitude modulation is a process by which the wave signal is transmitted by modulating the amplitude of the signal. It is often called AM and is commonly used in transmitting a piece of information through a radio carrier wave. Amplitude modulation is mostly used in the form of electronic communication.

Currently, this technique is used in many areas of communication, such as in portable two-way radios, citizens band radios, VHF aircraft radios and in modems for computers. Amplitude modulation is also used to refer to medium wave AM radio broadcasting.

Amplitude modulation, or just AM, is one of the earliest modulation methods that is used in transmitting information over the radio. This technique was devised in the 20th century at a time when Landell de Moura and Reginald Fessenden were conducting experiments using a radiotelephone in the 1900s. After successful attempts, the modulation technique was established and used in electronic communication.

In general, amplitude modulation definition is given as a type of modulation where the amplitude of the carrier wave varies in some proportion with respect to the modulating data or the signal.

As for the mechanism, when amplitude modulation is used, there is a variation in the amplitude of the carrier. Here, the voltage or the power level of the information signal changes the amplitude of the carrier. In AM, the carrier does not vary in amplitude. However, the modulating data is in the form of signal components consisting of frequencies either higher or lower than that of the carrier. The signal components are known as sidebands, and the sideband power is responsible for the variations in the overall amplitude of the signal.

12.4.1 Expression for Amplitude Modulated Wave

We have carrier wave and modulating signals,

$$\left.\begin{aligned} m(t) &= A_m \sin \omega_m t \; and \\ c(t) &= A_c \sin \omega_c t \end{aligned}\right\} \rightarrow 1$$

.

m(t) = Modulating signal

c(t) = Carrier wave

A_m and Ac are the amplitude of modulating signal and carrier wave, respectively, in amplitude modulation. We are superimposing modulating signal into a carrier wave and also varying the amplitude of the carrier wave in accordance with the amplitude of the modulating signal, and the amplitude-modulated wave $C_m(t)$ will be

$C_m(t) = (A_c + A_m \sin \omega_m t) \sin \omega_c t$ 2

This is the general form of an amplitude-modulated wave.

$C_m(t)$ is the amplitude-modulated wave

Where,

$A = A_c + A_m \sin \omega_m t$ = Amplitude of the modulated wave

$\sin w_c t$ = Phase of modulated wave

$$\begin{aligned} C_m(t) &= A_c \left(1 + \tfrac{A_m}{A_c} \sin \omega_m t\right) \sin \omega_c t \\ &= A_c \sin \omega_c t + \tfrac{A_m}{A_c} A_c \sin \omega_m t \, \sin \omega_c t \end{aligned}$$

Where,

$\frac{A_m}{A_c} = \mu = \text{modulation index}$

$C_m(t) = A_c \sin \omega_c t + A_c \mu \sin \omega_m t \sin \omega_c t$

We can rewrite the above equation as

$$\begin{aligned} &\sin A \sin B = \tfrac{1}{2}[\cos(A-B) - \cos(A+B)] \\ &= A_c \sin \omega_c t + A_c \mu \tfrac{1}{2}[\cos(\omega_c - \omega_m) - \cos(\omega_c + \omega_m)] \\ &c_m(t) = A_c \sin \omega_c t + \tfrac{A_c \mu}{2} \cos(\omega_c - \omega_m) - \tfrac{A_c \mu}{2} \cos(\omega_c + \omega_m) \ldots\ldots 3 \end{aligned}$$

From equation 3, we can see amplitude modulated wave is the sum of three sine or cosine waves.

- Frequencies of Amplitude Modulated Wave

There are three frequencies in amplitude modulated wave – f_1, f_2 and f_3 – corresponding to ω_c, $\omega_c + \omega_m$ and $\omega_c - \omega_m$, respectively.

$\omega_1 = \omega_c \rightarrow$ it is corresponding $f_1 = f_c$

$\omega_2 = \omega_c + \omega_m \rightarrow$ it is corresponding $f_2 = f_c + f_m$

$\omega_3 = \omega_c - \omega_m \rightarrow$ it is corresponding $f_3 = f_c - f_m$

Where $f_c \rightarrow$ Carrier wave frequency

$f_c + f_m \rightarrow$ Upper side band frequency

$f_c - f_m \rightarrow$ Lower side band frequency

$f_m \rightarrow$ Modulating signal frequency

In general, $f_c >> f_m$

- Bandwidth(BW): it is the difference between the highest and lowest frequencies of the signal.

BW = Upper sideband frequency – Lower sideband frequency ($f_c - f_m$)

Or

$BW = f_{max} - f_{min}$

$BW = f_c + f_m - f_c + f_m = 2 f_m$

$BW = 2f_m$ = Twice the frequency of the modulating signal

- Modulation Index

It is the ratio of the amplitude of the modulating signal to the amplitude of the carrier wave.

$$\mu = \frac{A_m}{A_c} = \frac{\text{Amplitude of modulating signal}}{\text{Amplitude of carrier wave}}$$

12.4.2 Amplitude Modulated Waveform

The waveform representation of amplitude modulated wave is given below.

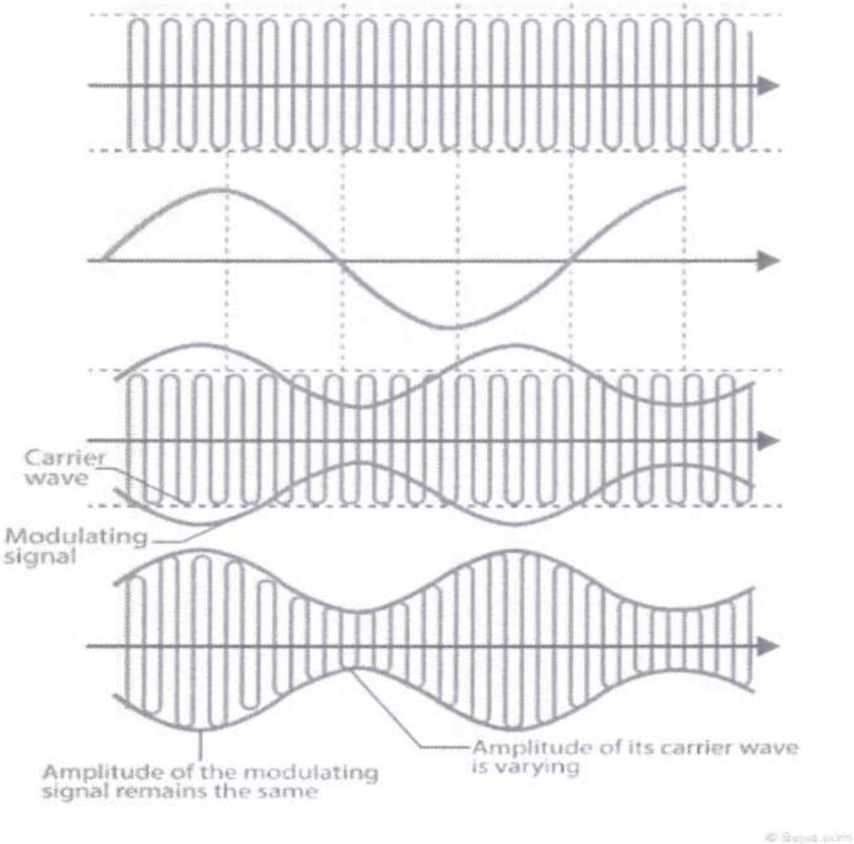


Figure 2: Amplitude modulation

1. Carrier wave

2. Modulating signal

3. Superposition of the carrier wave and modulating signal

4. Amplitude modulated wave

12.4.3 Applications of Amplitude Modulation

While amplitude modulation use has decreased over the years, it is still present and has several applications in certain transmission areas. We will look at them below.

- ***Broadcast Transmissions:*** AM is used in broadcasting transmission over the short, medium and long wavebands. Since AM is easy to demodulate, radio receivers for amplitude modulation are, therefore, easier and cheaper to manufacture.
- ***Air-band Radio:*** AM is used in VHF transmissions for many airborne applications, such as ground-to-air radio communications or two-way radio links, for ground staff personnel.
- ***Single Sideband:*** Amplitude modulation in this form is used for HF radio links or point-to-point HF links. AM uses a lower bandwidth and provides more effective use of the transmitted power.
- ***Quadrature Amplitude Modulation:*** AM is used extensively in transmitting data in several ways, including short-range wireless links, such as Wi-Fi to cellular telecommunications and others.

These are some of the important applications of amplitude modulation.

12.4.4 Demodulation Methods

The simplest AM demodulator is made up of a diode that acts as an envelope detector. The product detector, which is another type of demodulator, can offer better-quality demodulation but with a complex additional circuit.

12.5 Frequency modulation

Frequency modulation is a technique or a process of encoding information on a particular signal (analogue or digital) by varying the carrier wave frequency in accordance with the frequency of the modulating signal. As we know, a modulating signal is nothing but information or message that has to be transmitted after being converted into an electronic signal.

Much like amplitude modulation, frequency modulation also has a similar approach, where a carrier signal is modulated by the input signal. However, in the case of FM, the amplitude of the modulated signal is kept, or it remains constant.

The frequency modulation index is mostly over 1, and it usually requires a high bandwidth at a range of 200 kHz. FM operates in a very high-frequency range, normally between 88 to 108 Megahertz. There are complex circuits with an infinite number of sidebands that help in receiving high-quality signals with high sound quality.

Broadcast stations in the VHF portion of the frequency spectrum between 88.5 and 108 MHz often use large values of deviation (±75 kHz). This is known as wide-band FM (WBFM). Even though these signals support high-quality transmissions, they do occupy a large amount of bandwidth. Normally, 200 kHz is allowed for each wide-band FM transmission. On the other hand, communications use very little bandwidth. Whereas narrowband FM (NBFM) often uses deviation figures of around ±3 kHz. Besides, narrow-band FM is mostly used for two-way radio communication applications.

12.5.1 Frequency Modulation Equations

Frequency modulation equations mainly consist of a sinusoidal expression with the integral of the baseband signal that can be either a sine or cosine function.

It can be represented mathematically as;

$m(t) = A_m \cos(\omega_m t + \Theta)$ 1

$m(t) \rightarrow$ modulating signal

Where,

$A_m \rightarrow$ Amplitude of the modulating signal.

$\omega_m \rightarrow$ Angular frequency of the modulating signal.

$\Theta \rightarrow$ It is the phase of the modulating signal.

Same as amplitude modulation, when we try to modulate an input signal (information), we need a carrier wave, and we will experience

$C(t) = A_c \cos(\omega_c t + \Theta)$ 2

Angular modulation, which means ω_c (or) Θ of the carrier wave, starts varying linearly with respect to the modulating signal, like amplitude modulation.

Case I: Any Instant

The modulating signal at any instant of time.

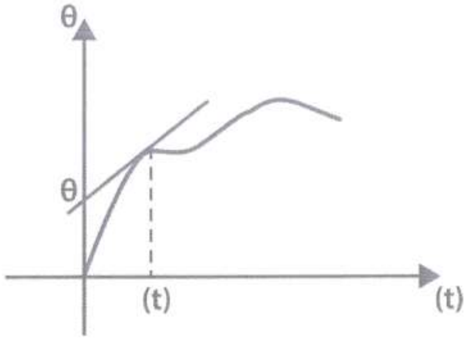


Figure 3: modulating signal at any instant of time

$C(t) = A_c \cos(\omega_c t + \Theta)$

For any particular instant $(\omega_c t + \Theta)$ is not varying with respect to time the Θ becomes so it Θ_0 = constant, then

ω_c = also constant

If we draw a tangent for the given signal at any instant of time, the slope of the tangent gives ω_c, and the tangent, when it cuts the Θ-axis, gives Θ_0 value.

Case II: For a Small Interval of Time

Now, let us consider a small interval of time $\Delta t = t_2 - t_1$

Time interval, $t_1 < t < t_2$, we will look this into two particular instant of t_1 and t_2

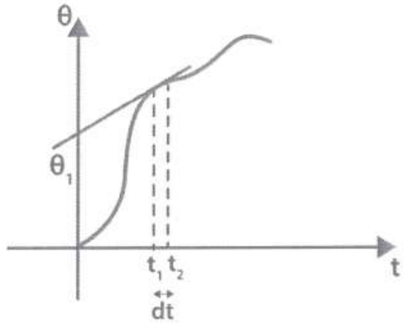


Figure 4: modulating signal a small Interval of Time

Let us consider an instant at (t_1); if we draw a tangent to a given signal at (t_1), the slope of the curve instantaneous frequency (w_i) at that particular instant.

Similarly, the intercept of the tangent with Θ – axis gives the instantaneous phase (Θ_i). Likewise, we can get w_i for any instant of the given curve. From this, one thing is clear,

$$\omega_i = \frac{d\theta}{dt}$$

Θ – is the phase at the instant

We write this as a function of time (t), and instantaneous frequency is

$$\omega(t) = \frac{d\theta(t)}{dt} \ldots\ldots\ldots\ldots\ldots 3$$

$$\int_{-\infty}^{t} \omega(t)dt = \theta \ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$

Equation 3 and 4 gives the fundamental understanding of phase and frequency. If we try to modulate this signal, let us see what's happening at any instant of time the signal phase is

$$\theta(t) = \omega_c t \int_{-\infty}^{t} k m(t) \ldots\ldots\ldots\ldots\ldots (5)$$

[this is known as phase modulation]

Where,

k – constant

$\omega_c \rightarrow$ frequency of the carrier wave

m(t) $\rightarrow$ modulating signal

The insert signal in this phase becomes,

$A \cos \Theta (t) = A_c \cos [\omega_c t + k\, m(t)]$

If we need the frequency of the wave,

$$\omega(t) = \frac{d\theta(t)}{dt} = \omega_c + km(t) \ldots\ldots\ldots\ldots 6$$

[This is known as frequency modulation]

Where,

$$\dot{m}(t) = \frac{d}{dt} m(t)$$

As we know, in the idea of frequency modulation, the frequency of the carrier wave must vary linearly with respect to a particular signal, as we can see in Equation 5. From this, we get,

$\omega(t) = \omega_c + km(t)$

If we do phase modulation, it is nothing but frequency modulation. When we do frequency modulation, we are differentiating the particular modulating signal, which then automatically depicts it as phase modulation.

- Expression for Frequency Modulated Wave

As we know, from amplitude modulation, we need two sine (or) cosine waves for modulation.

$m(t) = A_m \cos(\omega_m t)$ and $c(t) = A_c \cos(\omega_c t)$

or

$m(t) = A_m \cos(2\pi f_m t)$

$c(t) = A_c \cos(2\pi f_c t)$

Then frequency modulated wave will be:

$f_m(t) = f_c + k A_m \cos(2\pi f_m t)$

$f_m(t) = f_c + k\, m(t)$

Where,

fm(t) = is frequency modulated wave

$f_c \rightarrow$ frequency of the carrier wave

$m(t) \rightarrow$ modulating signal

$k \rightarrow$ proportionality constant

- Frequencies in Frequency Modulation

In FM, variation (or) deviation in frequency, the maximum deviation is Δf_{max}

$\Delta f_{max} = | f_m(t) - f_c |$

$= | KA_m \cos(2\pi f_m t) |$

The maximum deviation in frequency is K A_m

Generally, frequency deviation is defined as the measure of the change in a carrier frequency produced by the amplitude of the input-modulating signal.

- Modulation Index (μ)

The modulation index is the ratio of maximum deviation in frequency of the modulating signal.

$$\mu = \frac{\Delta f_{max}}{f_m} = \frac{K_{A_m}}{f_m}$$

12.5.2 Graphical Representation of Frequency Modulated Wave

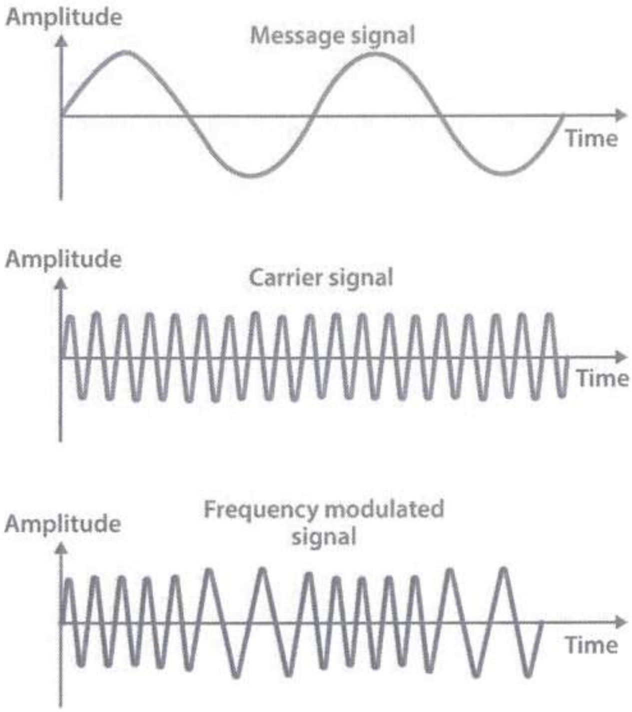


Figure 5: Frequency Modulation

If we observe the graph, we can notice that the frequency of a carrier increases when the amplitude of the input signal is increased. Here, the carrier frequency is maximum when the input signal is at its highest. Also, the frequency of a carrier decreases if the amplitude of the modulating signal goes down. What it means is that the carrier frequency is minimum when the input signal is at its lowest.

12.5.3 Frequency Demodulation

When there is modulation, usually, we need to successfully demodulate it and, at the same time, recover the original signal. In such cases, an FM demodulator, also known as an FM discriminator or FM detector, is used. While there are several types of FM demodulators, the main functionality of these devices is to convert the frequency variations of the input signal into amplitude variations of the output signal. The demodulators are used along with an audio amplifier or possibly a digital interface.

12.5.4 Applications of Frequency Modulation

If we talk about the applications of frequency modulation, it is mostly used in radio broadcasting. It offers a great advantage in radio transmission as it has a larger signal-to-noise ratio, which means that it results in low radio frequency interference. This is the main reason that many radio stations use FM to broadcast music over the radio.

Additionally, some of its uses are also found in radar, telemetry, seismic prospecting, and in EEG, different radio systems, music synthesis as well as in video-transmission instruments. In radio transmission, frequency modulation has a good advantage over other modulation. It has a larger signal-to-noise ratio, meaning it will reject radio frequency interferences much better than an equal power amplitude modulation (AM) signal. Due to this major reason, most music is broadcasted over FM radio.

12.6 Summary

- The communication system is a system which describes the information exchange between two points.
- The process of transmission and reception of information is called communication. The major elements of communication are the Transmitter of information, the Channel or medium of communication and the Receiver of information.
- Amplitude modulation is a process by which the wave signal is transmitted by modulating the amplitude of the signal. It is often called AM and is commonly used in transmitting a piece of information through a radio carrier wave.
- Amplitude modulation is mostly used in the form of electronic communication.
- Amplitude modulation is easier to implement.
- It requires a very high bandwidth that is equivalent to that of the highest audio frequency.
- Frequency modulation, commonly referred to as FM, is a common term that we hear in our daily lives. Today, Frequency modulation is used widely in radio communication and broadcasting..

12.7 Keywords

- **Information-** Message or information is the entity that is to be transmitted. It can be in the form of audio, video, temperature, picture, pressure, etc.
- **Signal-** The single-valued function of time carries the information. The information is converted into an electrical form for transmission.
- **Amplifier-** The electronic circuit or device that increases the amplitude or the strength of the transmitted signal is called an amplifier.
- **Transmitter** - It is the arrangement that processes the message signal into a suitable form for transmission and, subsequently, reception.
- **Antenna** - An antenna is a structure or a device that will radiate and receive electromagnetic waves. So, they are used in both transmitters and receivers.
- **Channel-** A channel refers to a physical medium such as wire, cables, or space through which the signal is passed from the transmitter to the receiver.

12.8 Review Questions

1. What are the elements of a communication system?
2. What are the types of modulation?
3. What are the types of communication systems?
4. Why are carrier waves of higher frequency compared to modulating signals?
5. Define modulation index.
6. What happens if $\mu > 1$?

7. Why do we need modulation?
8. How does frequency modulation work?
9. What is the advantage of frequency modulation over amplitude modulation?
10. How is the process of frequency modulation done?

12.8 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication.
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education.
4. Millman, Electronics Devices and Circuits, TMH.
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi.
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India.

Contents

Learning Objectives

After studying this unit the student will be able to:

1 Describe Transducers and it's working.

2 Explain applications of Transducers.

3 Discuss Thermocouple construction and its working.

4 Explain Thermocouple applications.

Introduction

The Transducers a device that can be used in various places to change physical activity into electric signal there are various types of transducers including Thermocouple, RTD, stain Gauge etc. A transducer is a device that converts one form of energy into another. In the context of digital electronics and instrumentation, transducers are often used to convert physical quantities, such as temperature, pressure, or force, into electrical signals that can be processed by electronic circuits. Transducers come in many different forms and can be based on a variety of physical phenomena, including mechanical, electrical, optical, and magnetic. Some examples of common transducers include temperature sensors, pressure sensors, strain gauges, accelerometers, and photoelectric sensors.

Mechanical transducers, such as pressure sensors and strain gauges, work by converting a mechanical force or displacement into an electrical signal. Pressure sensors, for example, use a diaphragm that deforms in response to changes in pressure, which causes a change in resistance or capacitance that can be measured by an electronic circuit. Strain gauges, on the other hand, use a resistive element that changes in resistance in response to changes in strain, which can be used to measure changes in force or deformation. Electrical transducers, such as thermocouples and photodiodes, work by converting an electrical signal into another electrical signal. Thermocouples, for example, use the Seebeck effect to generate a voltage that is proportional to the temperature difference between two junctions. Photodiodes, on the other hand, use the photoelectric effect to generate a current in response to light.

Optical transducers, such as laser displacement sensors and optical encoders, work by converting a physical quantity into a light signal that can be detected and measured by an electronic circuit. Laser displacement sensors, for example, use a laser beam to measure the distance between the sensor and a target object, while optical encoders use a series of light and dark lines to encode position information. Magnetic transducers, such as Hall-effect sensors and magneto resistive sensors, work by converting a magnetic field into an electrical signal. Hall-effect sensors, for example, use the Hall effect to measure the strength and direction of a magnetic field, while magneto resistive sensors use the magneto resistive effect to measure changes in resistance in response to changes in magnetic field.

13.1.2 Advantages of converting a physical quantity into an electrical signal

Here, we have listed the various advantages of converting a physical quantity into an electrical signal:

- Electrical signals are easily transmitted and processed for measurement.
- Electrical signals process less friction error.
- Small power is needed to control the electrical systems.
- Amplification and attenuation of electrical signals are easy.
- The measuring instrument used for measuring the electrical signal is very compact and accurate.

13.1.3 Parts of Transducer

A transducer consists of the following two important parts:

- Sensing element
- Transduction element

Transducers have other vital parts such as signal processing equipment, amplifiers and power supplies.

- **Sensing Element**

It is the part of a transducer that responds to the physical sensation. The response of the sensing element depends on the physical phenomenon.

- **Transduction Element**

The transduction element of the transducer converts the output of the sensing element into an electrical signal. The transduction element is also called the secondary transducer.

13.1.4 Types of Transducers

There are two types of transducers, as follows:

- Input Transducer
- Output Transducer
- **Input Transducer**

An input transducer or a sensor takes in physical energy and converts it into an electrical signal that can be read. A microphone, for example, converts physical sound waves into an electrical signal that can be transferred through wires.

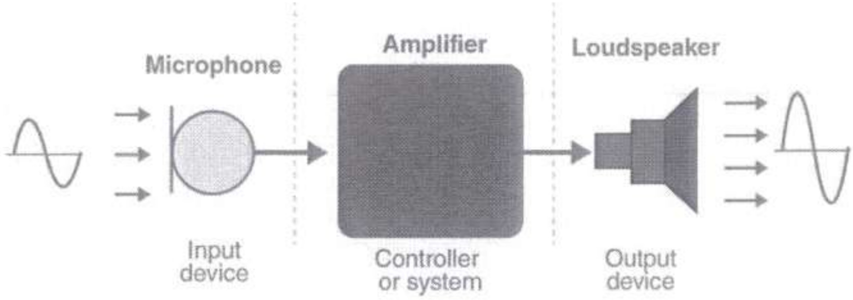


Figure 2: Input Transducer

- **Output Transducer**

An output transducer, or an actuator, takes in electrical signals and converts them into other forms of energy. A lamp converts electricity into light and a motor, on the other hand, converts electricity into motion.

13.1.5 Factors to consider while selecting a transducer

- Transducers should have high input impedance and low output impedance to avoid the loading effect.
- A transducer should be highly sensitive to desired signals and insensitive to unwanted signals.
- Transducers should be able to work in corrosive environments.
- The transducer circuit should have overload protection to withstand overloads.

13.1.6 Transducer Efficiency

Transducer efficiency is defined as the ratio of output power in the desired form to the total power input. Mathematically, the ratio is represented as follows:

$$E = \frac{Q}{P}$$

P represents the input in the above ratio, and Q represents the power output in the desired form. The efficiency of the transducer always falls between 0 and 1.

No transducer is 100% efficient; some power is always lost in the conversion process. This loss is manifested in the form of heat. In incandescent lamps of certain wattage, only a few watts are transformed into visible light. Most of the power is dissipated as heat. Due to this, an incandescent lamp is a bad transducer in terms of efficiency.

13.1.7 Applications of Transducer

- A transducer measures load on the engines
- They are used to detect the movement of muscles; this process is known as acceleromyograph.
- Transducers are used in an ultrasound machine.
- The transducers in a speaker convert electrical signals into acoustic sound.
- A transducer is used in the antenna to convert electromagnetic waves into an electrical signal.

13.2 Thermocouple

A thermocouple is a transducer that measures temperature. It consists of two different types of metals, joined together at one end. When the junction of the two metals is heated or cooled, a voltage is created that can be correlated back to the temperature.

A thermocouple is a simple, robust and cost-effective temperature sensor used in a wide range of temperature measurement processes.

Thermocouples are manufactured in a variety of styles, such as thermocouple probes, thermocouple probes with connectors, transition joint thermocouple probes, infrared thermocouples, bare wire thermocouple or even just thermocouple wire. Thermocouples are commonly used in a wide range of applications. Due to their wide range of models and technical specifications, but it is extremely important to understand its basic structure, functionality, ranges as to better determine the right thermocouple type and material of thermocouple for an application.

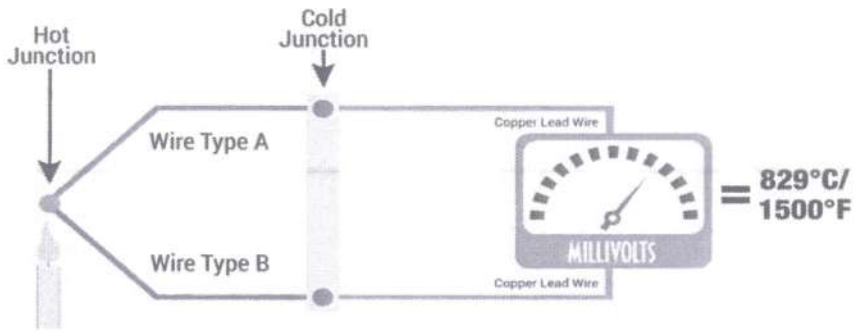


Figure 3: Thermocouple

13.2.1 Principle of operation

In 1821, the German physicist *Thomas Johann Seebeck* discovered that a magnetic needle held near a circuit made up of two dissimilar metals got deflected when one of the dissimilar metal junctions was heated. At the time, *Seebeck* referred to this consequence as thermo-magnetism.

The magnetic field he observed was later shown to be due to thermo-electric current. In practical use, the voltage generated at a single junction of two different types of wire is what is of interest as this can be used to measure temperature at very high and low temperatures.

The magnitude of the voltage depends on the types of wire being used. Generally, the voltage is in the microvolt range and care must be taken to obtain a usable measurement. Although very little current flows, power can be generated by a single thermocouple junction. Power generation using multiple thermocouples, as in a thermopile, is common.

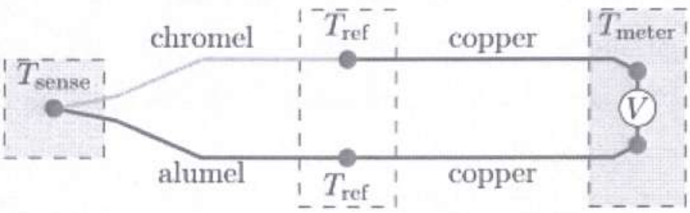


Figure 4: Thermocouple Principle of operation

The standard configuration for thermocouple usage is shown in the figure. Briefly, the desired temperature T_{sense} is obtained using three inputs—the characteristic function $E(T)$ of the thermocouple, the measured voltage V, and the reference junctions' temperature T_{ref}. The solution to the equation $E(T_{sense}) = V + E(T_{ref})$ yields T_{sense}. These details are often hidden from the user since the reference junction block (with T_{ref} thermometer), voltmeter, and equation solver are combined into a single product.

- **Seebeck effect**

The Seebeck effect refers to the development of an electromotive force across two points of an electrically conducting material when there is a temperature difference between those two points. Under open-circuit conditions where there is no internal current flow, the gradient of voltage (∇V) is directly proportional to the gradient in temperature (∇T)

$$\nabla V = -s(T)\nabla \mathrm{T},$$

Where S(T) is a temperature-dependent material property known as the Seebeck coefficient.

The standard measurement configuration shown in the figure shows four temperature regions and thus four voltage contributions:

1. Change from to , in the lower copper wire.
2. Change from to , in the alumel wire.
3. Change from to , in the chromel wire.
4. Change from to , in the upper copper wire.

The first and fourth contributions cancel out exactly, because these regions involve the same temperature change and an identical material. As a result, T_{meter} does not influence the measured voltage. The second and third contributions do not cancel, as they involve different materials.

The measured voltage turns out to be

$$V = \int_{T_{\text{ref}}}^{T_{\text{sense}}} \left(S_+(T) - S_-(T)\right)\, dT,$$

Where S_+ and S_- are the Seebeck coefficients of the conductors attached to the positive and negative terminals of the voltmeter, respectively (chromel and alumel in the figure).

- **Characteristic function**

The thermocouple's behavior is captured by a **characteristic function** E(T), which needs only to be consulted at two arguments:

$$V = E(T_{\text{sense}}) - E(T_{\text{ref}}).$$

In terms of the Seebeck coefficients, the characteristic function is defined by

$$E(T) = \int^{T} S_+(T') - S_-(T')dT' + \text{const}$$

The constant of integration in this indefinite integral has no significance, but is conventionally chosen such that $E\,(0^{\circ}C) = 0$.

Thermocouple manufacturers and metrology standards organizations such as NIST provide tables of the function E(T) that have been measured and interpolated over a range of temperatures, for particular thermocouple types

- **Reference junction**

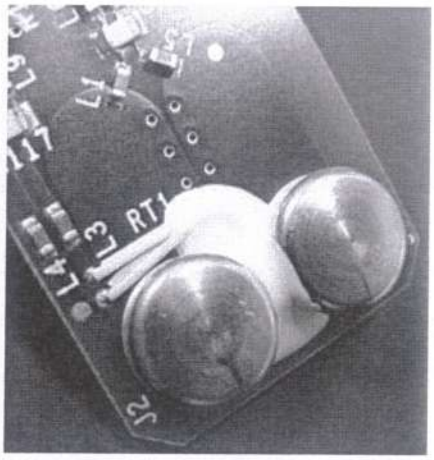


Figure 5: Reference junction block inside a Fluke CNX t3000 temperature meter.

To obtain the desired measurement of T_{sense}, it is not sufficient to just measure V. The temperature at the reference junctions T_{ref} must be already known. Two strategies are often used here:

- *"Ice bath" method:* The reference junction block is immersed in a semi-frozen bath of distilled water at atmospheric pressure. The precise temperature of the melting point phase transition acts as a natural thermostat, fixing T_{ref} to 0 °C.
- *Reference junction sensor (known as "cold junction compensation"):* The reference junction block is allowed to vary in temperature, but the temperature is measured at this block using a separate temperature sensor. This secondary measurement is used to compensate for temperature variation at the junction block. The thermocouple junction is often exposed to extreme environments, while the reference junction is often mounted near the instrument's location. Semiconductor thermometer devices are often used in modern thermocouple instruments.

In both cases the value $V+E(T_{ref})$ is calculated, then the function E(T) is searched for a matching value. The argument where this match occurs is the value of T_{sense}:

$$E(T_{\text{sense}}) = V + E(T_{\text{ref}}).$$

13.2.2 Applications of Thermocouple

Thermocouples are suitable for measuring over a large temperature range, from −270 up to 3000 °C (for a short time, in inert atmosphere.

Applications include temperature measurement for kilns, gas turbine exhaust, diesel engines, other industrial processes and fog machines. They are less suitable for applications where smaller temperature differences need to be measured with high accuracy, for example the range 0–100 °C with 0.1 °C accuracy. For such applications thermistors, silicon bandgap temperature sensors and resistance thermometers are more suitable.

- **Steel industry**

Type B, S, R and K thermocouples are used extensively in the steel and iron industries to monitor temperatures and chemistry throughout the steel making process.

Disposable, immersible, type S thermocouples are regularly used in the electric arc furnace process to accurately measure the temperature of steel before tapping. The cooling curve of a small steel sample can be analyzed and used to estimate the carbon content of molten steel.

- **Gas appliance safety**

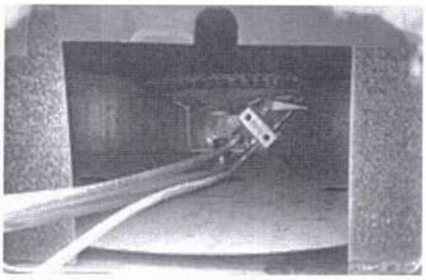


Figure 6: A thermocouple (the right most tube) inside the burner assembly of a water heater

Figure 7: Thermocouple connection in gas appliances

Many gas-fed heating appliances such as ovens and water heaters make use of a pilot flame to ignite the main gas burner when required. If the pilot flame goes out, unburned gas may be released, which is an explosion risk and a health hazard.

To prevent this, some appliances use a thermocouple in a fail-safe circuit to sense when the pilot light is burning. The tip of the thermocouple is placed in the pilot flame, generating a voltage which operates the supply valve which feeds gas to the pilot.

So long as the pilot flame remains lit, the thermocouple remains hot, and the pilot gas valve is held open. If the pilot light goes out, the thermocouple temperature falls, causing the voltage across the thermocouple to drop and the valve to close.

Where the probe may be easily placed above the flame, a rectifying sensor may often be used instead. With part ceramic construction, they may also be known as flame rods, flame sensors or flame detection electrodes.

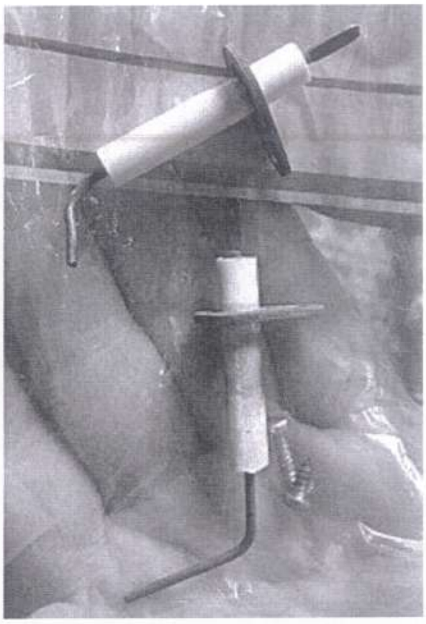


Figure 8: Flame-igniter (top)-and-flame-sensor

Some combined main burner and pilot gas valves reduce the power demand to within the range of a single universal thermocouple heated by a pilot (25 mV open circuit falling by half with the coil connected to a 10–12 mV, 0.2–0.25 A source, typically) by sizing the coil to be able to hold the valve open against a light spring, but only after the initial turning-on force is provided by the user pressing and holding a knob to compress the spring during lighting of the pilot.

These systems are identifiable by the "press and hold for x minutes" in the pilot lighting instructions. (The holding current requirement of such a valve is much less than a bigger solenoid designed for pulling the valve in from a closed position would require.)

Special test sets are made to confirm the valve let-go and holding currents, because an ordinary milliammeter cannot be used as it introduces more resistance than the gas valve coil. Apart from testing the open circuit voltage of the thermocouple, and the near short-circuit DC continuity through the thermocouple gas valve coil, the easiest non-specialist test is substitution of a known good gas valve.

Some systems, known as millivolt control systems, extend the thermocouple concept to both open and close the main gas valve as well. Not only does the voltage created by the pilot thermocouple activate the pilot gas valve, it is also routed through a thermostat to power the main gas valve as well.

Here, a larger voltage is needed than in a pilot flame safety system described above, and a thermopile is used rather than a single thermocouple. Such a system requires no external source of electricity for its operation and thus can operate during a power failure, provided that all the other related system components allow for this.

This excludes common forced air furnaces because external electrical power is required to operate the blower motor, but this feature is especially useful for un-powered convection heaters.

A similar gas shut-off safety mechanism using a thermocouple is sometimes employed to ensure that the main burner ignites within a certain time period, shutting off the main burner gas supply valve should that not happen.

Out of concern about energy wasted by the standing pilot flame, designers of many newer appliances have switched to an electronically controlled pilot-less ignition, also called intermittent ignition. With no standing pilot flame, there is no risk of gas buildup should the flame go out, so these appliances do not need thermocouple-based pilot safety switches.

As these designs lose the benefit of operation without a continuous source of electricity, standing pilots are still used in some appliances. The exception is later model instantaneous (aka "tankless") water heaters that use the flow of water to generate the current required to ignite the gas burner; these designs also use a thermocouple as a safety cut-off device in the event the gas fails to ignite, or if the flame is extinguished.

- **Thermopile radiation sensors**

Thermopiles are used for measuring the intensity of incident radiation, typically visible or infrared light, which heats the hot junctions, while the cold junctions are on a heat sink. It is possible to measure radiative intensities of only a few $\mu W/cm^2$ with commercially available thermopile sensors. For example, some laser power meters are based on such sensors; these are specifically known as thermopile laser sensor.

The principle of operation of a thermopile sensor is distinct from that of a bolometer, as the latter relies on a change in resistance.

- **Manufacturing**

Thermocouples can generally be used in the testing of prototype electrical and mechanical apparatus. For example, switchgear under test for its current carrying capacity may have thermocouples installed and monitored during a heat run test, to confirm that the temperature rise at rated current does not exceed designed limits.

- **Power production**

A thermocouple can produce current to drive some processes directly, without the need for extra circuitry and power sources. For example, the power from a thermocouple can activate a valve when a temperature difference arises. The electrical energy generated by a thermocouple is converted from the heat which must be supplied to the hot side to maintain the electric potential. A continuous transfer of heat is necessary because the current flowing through the thermocouple tends to cause the hot side to cool down and the cold side to heat up (the Peltier effect).

Thermocouples can be connected in series to form a thermopile, where all the hot junctions are exposed to a higher temperature and all the cold junctions to a lower temperature. The output is the sum of the voltages across the individual junctions, giving larger voltage and power output. In a radioisotope thermoelectric generator, the radioactive decay of transuranic elements as a heat source has been used to power spacecraft on missions too far from the Sun to use solar power.

Thermopiles heated by kerosene lamps were used to run battery less radio receivers in isolated areas. There are commercially produced lanterns that use the heat from a candle to run several light-emitting diodes, and thermoelectrically-powered fans to improve air circulation and heat distribution in wood stoves.

- **Process plants**

Chemical production and petroleum refineries will usually employ computers for logging and for limit testing the many temperatures associated with a process, typically numbering in the hundreds. For such cases, a number of thermocouple leads will be brought to a common reference block (a large block of copper) containing the second thermocouple of each circuit. The temperature of the block is in turn measured by a thermistor. Simple computations are used to determine the temperature at each measured location.

- **Thermocouple as vacuum gauge**

A thermocouple can be used as a vacuum gauge over the range of approximately 0.001 to 1 torr absolute pressure. In this pressure range, the mean free path of the gas is comparable to the dimensions of the vacuum chamber, and the flow regime is neither purely viscous nor purely molecular. In this configuration, the thermocouple junction is attached to the centre of a short heating wire, which is usually energized by a constant current of about 5 mA, and the heat is removed at a rate related to the thermal conductivity of the gas.

The temperature detected at the thermocouple junction depends on the thermal conductivity of the surrounding gas, which depends on the pressure of the gas. The potential difference measured by a thermocouple is proportional to the square of pressure over the low- to medium-vacuum range. At higher (viscous flow) and lower (molecular flow) pressures, the thermal conductivity of air or any other gas is essentially independent of pressure. The thermocouple was first used as a vacuum gauge by Voege in 1906. The mathematical model for the thermocouple as a vacuum gauge is quite complicated, as explained in detail by Van Atta, but can be simplified to:

$$P = \frac{B(V^2 - V_0^2)}{V_0^2},$$

Where P is the gas pressure, B is a constant that depends on the thermocouple temperature, the gas composition and the vacuum-chamber geometry, V_0 is the thermocouple voltage at zero pressure (absolute), and V is the voltage indicated by the thermocouple.

The alternative is the Pirani gauge, which operates in a similar way, over approximately the same pressure range, but is only a 2-terminal device, sensing the change in resistance with temperature of a thin electrically heated wire, rather than using a thermocouple.

13.3 Summary

- A transducer is an electronic device that converts energy from one form to another. The process of converting energy from one form to another is known as transduction.
- The primary function of transducers is to convert a physical force into an electrical signal so that it can be easily handled and transmitted for measurement.
- Electrical signals are easily transmitted and processed for measurement.
- The measuring instrument used for measuring the electrical signal is very compact and accurate.
- An input transducer or a sensor takes in physical energy and converts it into an electrical signal that can be read.
- An output transducer, or an actuator, takes in electrical signals and converts them into other forms of energy.
- Transducer efficiency is defined as the ratio of output power in the desired form to the total power input.
- A thermocouple is a sensor that measures temperature. It consists of two different types of metals, joined together at one end.
- Thermocouples are available in different combinations of metals or calibrations.

13.4 Keywords

- Sensing Element

It is the part of a transducer that responds to the physical sensation. The response of the sensing element depends on the physical phenomenon.

- Transduction Element

The transduction element of the transducer converts the output of the sensing element into an electrical signal. The transduction element is also called the secondary transducer.

- Seebeck effect

The Seebeck effect refers to the development of an electromotive force across two points of an electrically conducting material when there is a temperature difference between those two points. 10.9 Review Questions

13.5 Review Questions

1. What is a transducer?
2. What is the need of a transducer?
3. What are the important parts of a transducer?
4. What is an input transducer?
5. Give an example of a transducer.
6. How to use Thermocouple as vacuum gauge
7. What is Transducer Efficiency
8. What is Thermopile radiation sensors
9. How to use thermocouple in Steel industry
10. How to use thermocouple in Gas appliance safety

13.6 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication.
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education.
4. Millman, Electronics Devices and Circuits, TMH.
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi.
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India.

Contents

Learning Objectives

After studying this unit the student will be able to:

1 Describe Transducers and it's working.

2 Explain applications of Transducers.

3 Discuss RTD, Strain Gauges, Load Cell construction and its working.

4 Explain RTD, Strain Gauges, Load Cell applications.

Introduction

In this chapter, we discuss in detail regarding RTD, Strain Gauge and Load Cell. RTD (Resistance Temperature Detector) is a type of temperature sensor that operates on the principle of the change in electrical resistance of metals with changes in temperature. RTDs are commonly made of platinum, nickel, or copper and are highly accurate and reliable over a wide temperature range. They are commonly used in industrial and scientific applications where precise temperature measurement is required.

Strain gauge is a type of transducer that measures the strain or deformation of a material. It works by measuring the change in resistance of a metal wire or foil that is bonded to the surface of the material being measured. Strain gauges are commonly used in load cells, pressure sensors, and other applications where the measurement of strain is important. They are highly sensitive and can measure small changes in strain with high accuracy. Load cell is a type of transducer that measures the force or weight applied to it. It works by measuring the deformation or strain of a material under load, using one or more strain gauges bonded to the surface of the material. Load cells are commonly used in industrial and commercial weighing systems, such as scales, hoppers, and tanks. They are highly accurate and can measure forces ranging from a few grams to several tons, depending on the size and type of load cell used.

14.1 RTD (Resistance Temperature Detector)

A Resistance Temperature Detector (also known as a Resistance Thermometer or RTD) is an electronic device used to determine the temperature by measuring the resistance of an electrical wire.

This wire is referred to as a temperature sensor. If we want to measure temperature with high accuracy, an RTD is the ideal solution, as it has good linear characteristics over a wide range of temperatures. Other common electronics devices used to measure temperature include a thermocouple or a thermistor.

The variation of resistance of the metal with the variation of the temperature is given as,

$$R_t = R_0[1 + (t - t_0) + \beta(t - t_0)^2 + \cdots\cdots]$$

Where, R_t and R_0 are the resistance values at t°C and t_0°C temperatures. α and β are the constants depends on the metals.

This expression is for huge range of temperature. For small range of temperature, the expression can be,

$$R_t = R_0[1 + (t - t_0)]$$

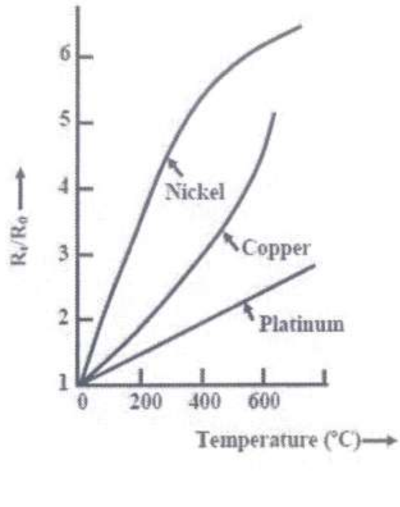


Figure 1: Temperature Vs Resistance graph

In RTD devices; Copper, Nickel and Platinum are widely used metals. These three metals are having different resistance variations with respective to the temperature variations. That is called resistance-temperature characteristics.

Platinum has the temperature range of 650°C, and then the Copper and Nickel have 120°C and 300°C respectively. The figure-1 shows the resistance-temperature characteristics curve of the three different metals. For Platinum, its resistance changes by approximately 0.4 ohms per degree Celsius of temperature.

The purity of the platinum is checked by measuring R_{100} / R_0. Because, whatever the materials actually we are using for making the RTD that should be pure. If it will not pure, it will deviate from the conventional resistance-temperature graph. So, α and β values will change depending upon the metals.

14.1.1 Construction of Resistance Temperature Detector or RTD

The construction is typically such that the wire is wound on a form (in a coil) on notched mica cross frame to achieve small size, improving the thermal conductivity to decrease the response time and a high rate of heat transfer is obtained. In the industrial RTD's, the coil is protected by a stainless steel sheath or a protective tube.

So that, the physical strain is negligible as the wire expands and increase the length of wire with the temperature change.

If the strain on the wire is increasing, then the tension increases. Due to that, the resistance of the wire will change which is undesirable. So, we don't want to change the resistance of wire by any other unwanted changes except the temperature changes.
This is also useful to RTD maintenance while the plant is in operation. Mica is placed in between the steel sheath and resistance wire for better electrical insulation. Due less strain in resistance wire, it should be carefully wound over mica sheet. The fig.2 shows the structural view of an Industrial Resistance Temperature Detector.

14.1.2 Signal Conditioning of RTD

We can get this RTD in market. But we must know the procedure how to use it and how to make the signal conditioning circuitry. So that, the lead wire errors and other calibration errors can be minimized. In this RTD, the change in resistance value is very small with respect to the temperature.

So, the RTD value is measured by using a bridge circuit. By supplying the constant electric current to the bridge circuit and measuring the resulting voltage drop across the resistor, the RTD resistance can be calculated.

Thereby, the temperature can be also determined. This temperature is determined by converting the RTD resistance value using a calibration expression. The different modules of RTD are shown in below figures.

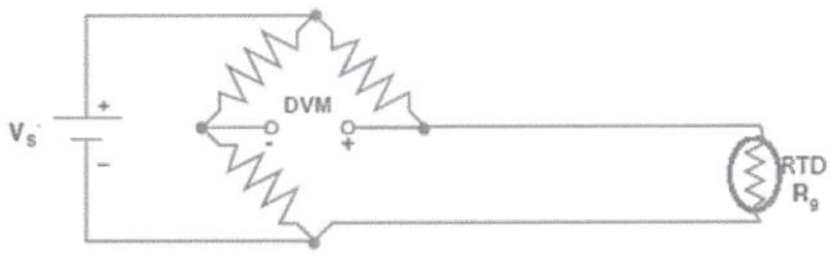


Figure 2: Two Wires RTD Bridge

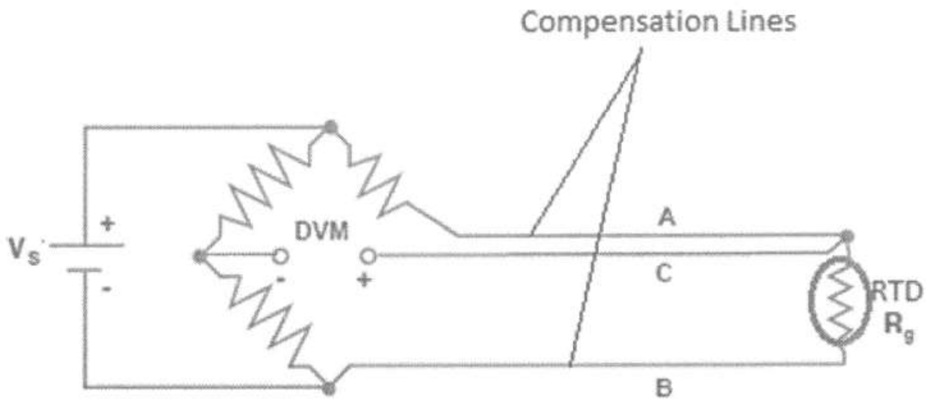


Figure 3: Three Wires RTD Bridge

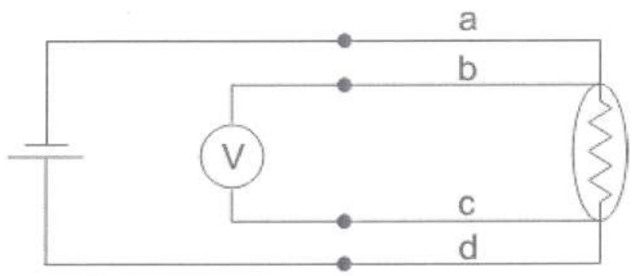


Figure 4: Four Wires RTD Bridge

In two wires RTD Bridge, the dummy wire is absent. The output taken from the remaining two ends as shown in fig.3. But the extension wire resistances are very important to be considered, because the impedance of the extension wires may affect the temperature reading. This effect is minimizing in three wires RTD bridge circuit by connecting a dummy wire C.

If wires A and B are matched properly in terms of length and cross section area, then their impedance effects will cancel because each wire is in opposite position. So that, the dummy wire C acts as a sense lead to measure the voltage drop across the RTD resistance and it carries no current. In these circuits, the output voltage is directly proportional to the temperature. So, we need one calibration equation to find the temperature.

14.1.3 Expressions for a Three Wires RTD Circuit

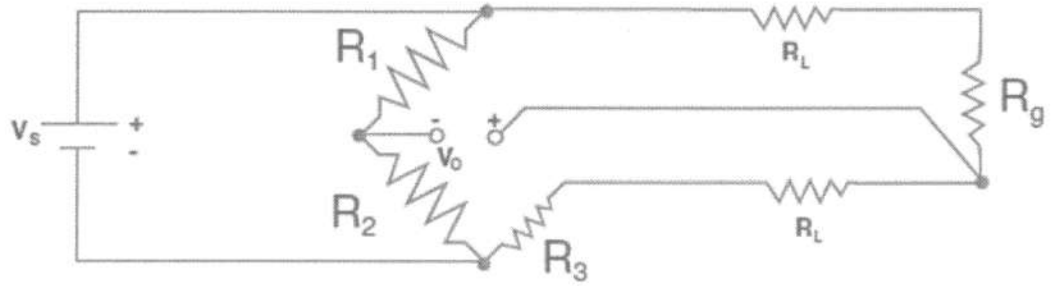


Figure 5: Circuitry view of Three Wires RTD Bridge

If we know the values of V_S and V_O, we can find R_g and then we can find the temperature value using calibration equation. Now, assume $R_1 = R_2$:

$$V_0 = V_S\left(\frac{R_3}{R_3+R_g}\right) - \left(\frac{V_S}{2}\right)$$

If $R_3 = R_g$; then $V_O = 0$ and the bridge is balanced.

This can be done manually, but if we don't want to do a manual calculation, we can just solve the equation 3 to get the expression for R_g.

$$R_g = R_3\left(\frac{V_S-2V_O}{V_S+2V_O}\right)$$

This expression assumes, when the lead resistance $R_L = 0$. Suppose, if R_L is present in a situation, then the expression of R_g becomes,

$$R_g = R_3\left(\frac{V_S-2V_O}{V_S+2V_O}\right) - R_L\left(\frac{4V_O}{V_S+2V_O}\right)$$

So, there is an error in the RTD resistance value because of the R_L resistance. That is why we need to compensated the R_L resistance as we discussed already by connecting one dummy line 'C' as shown in fig.4.

14.1.4 Limitations of RTD

In the RTD resistance, there will be an I^2R power dissipation by the device itself that causes a slight heating effect. This is called as self-heating in RTD.

This may also cause an erroneous reading. Thus, the electric current through the RTD resistance must be kept sufficiently low and constant to avoid self-heating.

14.2 Strain gauge

A **strain gauge** is used to measure strain. In 1938 Edward E. Simmons and writer C. Ruge invented the strain gauges. A metallic foil is supported by an insulating flexible backing that consists of a strain gauge.

When a force is applied, a strain gauge is used as a sensor for measuring variations in resistance, then converting those changes in electrical resistance into measurements. Strain gauges are made from long, thin pieces of metal conductor foil bonded to a flexible backing material known as a carrier.

14.2.1 What is a Strain Gauge?

A strain gauge is a sensor that translates force, pressure, tension, weight, etc. into a change in electrical resistance that can subsequently be measured. The resistance of a strain gauge varies with applied force. Stress and strain happen when outside forces are applied to a stationary item.

A strain gauge is one of the most crucial sensors used in the electrical measurement method used to assess mechanical quantities.

14.2.2 Strain Gauge Working Principle

Strain gauge is on the principle of electrical conductance and its dependence on the conductor's geometry. Whenever a conductor is stretched within the limits of its elasticity, it will not break, but it may be narrower and longer. Similarly, when it is compressed, it gets shorter and broader, and ultimately it changes resistance.

14.2.3 Types of Strain Gauge

Based on the position and arrangement of the strain gauge, it is classified into the following types.

- Linear strain gauge
- Rosette strain gauge
- Diaphragm strain gauge
- Shear strain gauge
- Double parallel strain gauge

Based on the type of resistance material, it is classified into the following four types.

- Fine wire strain gauge
- Metal foil strain gauge
- Semiconductor strain gauge
- Photo-electric strain gauge

Based on the construction method, it is classified into the following two types.

- Unbonded type
- Bonded type

14.2.4 Characteristics of Strain Gauge

The gauge dimensions, resistance, gauge factor, resistivity, temperature coefficient, and thermal stability mainly define the characteristics of a strain gauge. Gauge dimensions and shape are essential in choosing the right strain gauge type for a given application.

- A wide range of accessories are available along with the strain gauge
- It is highly precise and doesn't get influenced due to temperature changes. However, a thermistor is used for temperature corrections if they get affected by temperature changes.
- A strain gauge is ideal for long-distance communication as the output is an electrical signal.
- Strain gauges require low maintenance and have a long operating life.
- The production of strain gauges is easy because of their simple operating principle and a small number of components.
- Strain gauge suitable for long-term installation. However, Strain gauges require certain precautions during installation.
- All the strain gauges produced by Encardio-Rite are hermetically sealed and made up of stainless steel thus, it is waterproof.
- Strain gauges are fully encapsulated for protection against handling and installation damage.
- The remote digital readout resistivity is also possible in strain gauges.

14.2.5 Design Consideration of Strain Gauge

The strain gauges are made from metals. It can be further divided into wire-wounded and metal foil types. The earliest form of the device is Wire wounded type. Now-a-day, metal-foil strain gauges are the most common type. Metal foil strain gauge manufactured through photochemical etching or circuit printing. Some raw materials used for producing metal strain gauges are copper-nickel, nickel-chromium, and platinum-alloys.

14.2.6 Strain Gauge Applications

Strain gauges are extensively used in geotechnical monitoring to constantly check structures, dams, tunnels, and buildings so that accidents can be avoided well on time. The applications of strain gauges include:

- Airplane: Strain gauges are fixed to the structural load-bearing components to measure stresses along load paths for wing deformation in an airplane.
- Cable bridge: Strain gauge becomes important to monitor the bridges regularly for any kind of deformation as it might lead to fatal accidents. Strain gauge technology is used in the real-time monitoring of huge bridges, making the inspections precise.
- Rail Monitoring: Strain Gauges have a vast history in the safety of rails. The strain gauge is used to measure strain corresponding to stress on rails. Strain gauges measure axial tension or compression without impact on the rails. While in an emergency, the strain gauges can generate a warning. Hence, maintenance can be done early to minimize the impact on rail traffic.

14.2.7 Strain Gauge Rosette

We use the most common type of strain gauge to find the strain in a different direction.

- In order to measure strain, forces, and deformation in multiple axes, strain gauge rosettes are used. A strain gauge rosette with two strain gauges can be used if the direction of principal stresses is known. A combination of 3 strain gauges was placed to find the strain in any direction.
- Strains that occur in most engineering structures and machines are very small. For example, In a metallic tie rod, the maximum allowable axial strain will be less than the offset strain of 0.2%.
- A widely used method of strain gauge measurement is based on the electric resistance strain gauge. These strain gauges measure normal or longitudinal strain whether extension or contraction at a point on the surface of the deforming solid.
- When shear strain cannot be measured directly by one single strain gauge. An arrangement of three gauges mounted at the same point is used to define the state of strain of a point; such arrangement is known as a strain gauge rosette.

14.2.8 Strain Gauge Rosettes are further classified into three types:

- Delta strain gauge rosettes (the 2nd and 3rd strain gauges $60^0$and 120^0away from the 1st strain gauge)
- Rectangular strain gauge rosettes (the 2nd and 3rd strain gauges$45^0$and 90^0away from the 1st strain gauge)
- Y-type or Star strain gauge rosettes (the 2nd and 3rd strain gauge $120^0$and 240^0away from 1st strain gauge)

14.2.9 Strain Gauge Limitations

Each strain gauge has its limitations regarding temperature, fatigue, amount of strain, and the measurement environment. These limitations must be examined before using of strain gauge.

- Strain gauges measure strain in only one direction, a single strain gauge is often referred to as an axial gauge which is short for uniaxial gauge and reflects its sensitivity of the strain in one direction only.
- If the temperature is more, resistance will be more and vice versa. This is a common property for all the conductors. It can be resolved by using strain gauges that are self-temperature compensated (or) by a dummy strain gauge technique.

14.3 Load cell

A **load cell** converts a force such as tension, compression, pressure, or torque into an electrical signal that can be measured and standardized. It is a force transducer. As the force applied to the load cell increases, the electrical signal changes proportionally. The most common types of load cell are pneumatic, hydraulic, and strain gauges.

14.3.1 How Load Cells Work

When we use load cells, one end is usually secured to a frame or base, while the other end is free to attach the weight or weight-bearing element.

When force is applied to the body of the load cell, it flexes slightly under the strain. This is similar to what happens to a fishing rod when a fisherman hooks a fish.

The fisherman will secure the rod in their hands while the fish applies a pulling force on the other end of the fishing line. The result of this force is that the fishing rod bends, with a bigger, stronger fish, causing the bend to be more extreme.

When this action happens to a load sensor, the deformation is very subtle and not visible to the naked eye. To measure the deformation, strain gages are tightly bonded to the body of the load cell at pre-determined points, causing them to deform in unison with the body. The resulting movement alters the electrical resistance of the strain gages in proportion to the amount of deformation caused by the applied load.

Using signal conditioning electronics, the electrical resistance of the strain gages can be measured with the resulting signal being output as a weight or force reading.

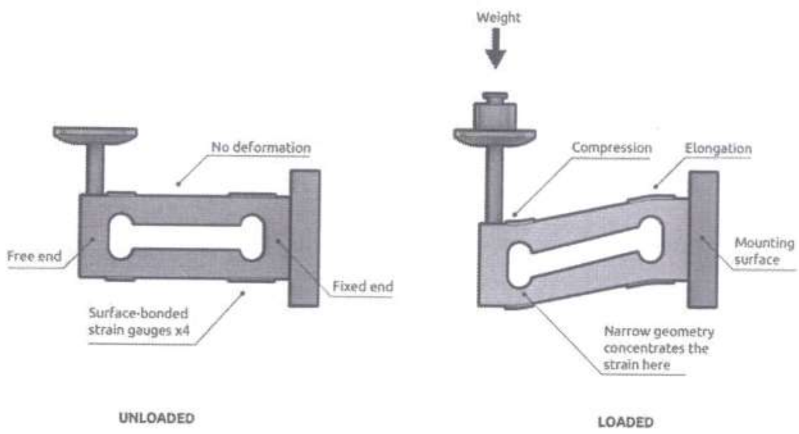


Figure 6: Load Cell Flexion

14.3.2 Technology of a Load Cell

A typical load cell consists of two parts: the main body and an attached electrical circuit. The main body is what bears the weight or force and accounts for most of the load cell's size. Typically, it is made from high-grade steel or aluminium, which ensures mechanical reliability, and predictable and uniform strain distribution.

The electrical circuit is housed within the load cell, tightly bonded to the main body. The circuit includes strain-gauges which are specialised parts of the circuit designed to sense the deformations of the main body.

These strain-gauges consist of thin, electrically conductive wire or foil arranged in a tight zig-zag pattern. This pattern makes them sensitive to stretch and compression along their length, but insensitive across their width. As such, they can be precisely positioned to sense forces that run along particular axes. For example, shear beam load cells have their strain gauges positioned at a 45-degree angle to the loading axis, so as to maximize the detection of the shear strain running through the load cell.

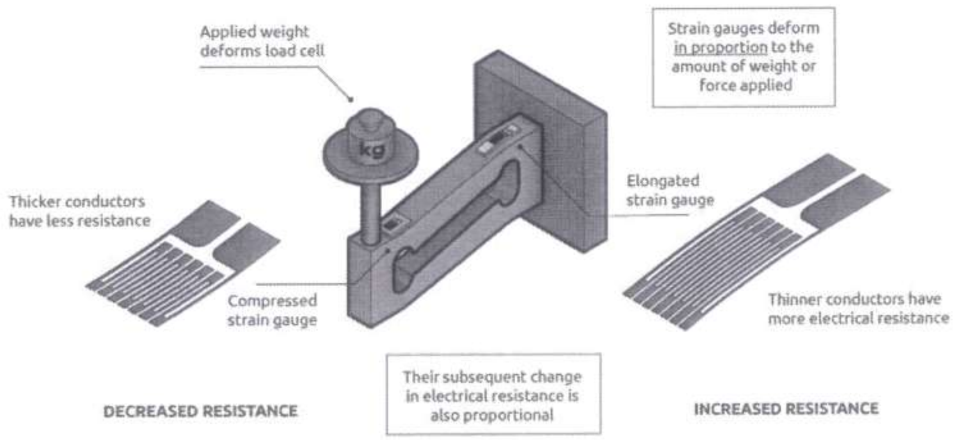


Figure 7: Strain Gauge De Formation

- **Strain gauge load cell**

Strain gauge load cells are the kind most often found in industrial settings. It is ideal as it is highly accurate, versatile, and cost-effective. Structurally, a load cell has a metal body to which strain gauges have been secured. The body is usually made of aluminum, alloy steel, or stainless steel which makes it very sturdy but also minimally elastic. This elasticity gives rise to the term "spring element", referring to the body of the load cell. When force is exerted on the load cell, the spring element is slightly deformed, and unless overloaded, always returns to its original shape. As the spring element deforms, the strain gauges also change shape. The resulting alteration to the resistance in the strain gauges can be measured as voltage. The change in voltage is proportional to the amount of force applied to the cell, thus the amount of force can be calculated from the load cell's output.

- **Strain Gauges**

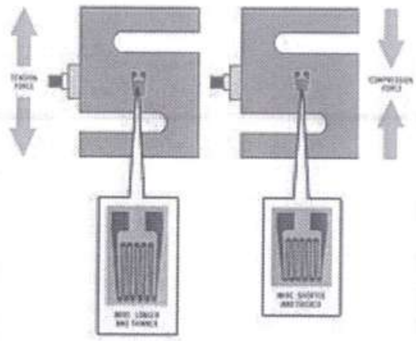


Figure 8: Strain Gauge

A strain gauge is constructed of very fine wire or foil, set up in a grid pattern and attached to a flexible backing. When the shape of the strain gauge is altered, a change in its electrical resistance occurs. The wire or foil in the strain gauge is arranged in a way that, when force is applied in one direction, a linear change in resistance results. Tension force stretches a strain gauge, causing it to get thinner and longer, resulting in an increase in resistance. Compression force does the opposite. The strain gauge compresses, becomes thicker and shorter, and resistance decreases. The strain gauge is attached to a flexible backing enabling it to be easily applied to a load cell, mirroring the minute changes to be measured.

Since the change in resistance measured by a single strain gauge is extremely small, it is difficult to accurately measure changes. Increasing the number of strain gauges applied collectively magnifies these small changes into something more measurable. A set of 4 strain gauges set in a specific circuit is an application of a Wheatstone bridge.

- **Wheatstone bridge**

A Wheatstone bridge is a configuration of four balanced resistors with a known excitation voltage applied as shown below:

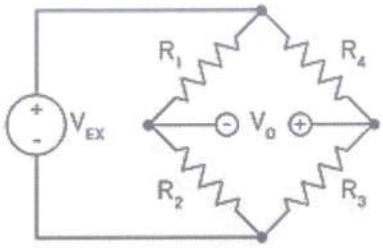


Figure 9: Wheat stone Bridge

Excitation voltage V_{EX} is a known constant and output voltage V_O is variable depending on the shape of the strain gauges. If all resistors are balanced, meaning $\frac{R1}{R2} = \frac{R3}{R4}$ then is V_O zero.

If the resistance in even one of the resistors changes, then V_O will likewise change. The change in V_O can be measured and interpreted using Ohm's law.

Ohm's law states that the current (I, measured in amperes) running through a conductor between two points is directly proportional to the voltage V across the two points.

Resistance (R, measured in ohms) is introduced as the constant in this relationship, independent of the current. Ohm's law is expressed in the equation $I = \frac{V}{R}$.

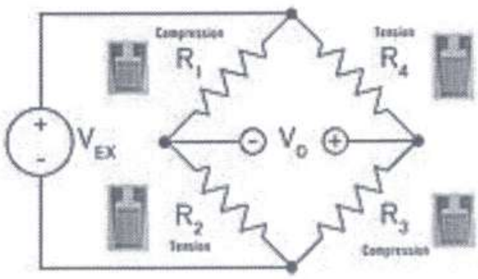


Figure 10: Wheat Stone Bridge circuit

When applied to the 4 legs of the Wheatstone bridge circuit, the resulting equation is:

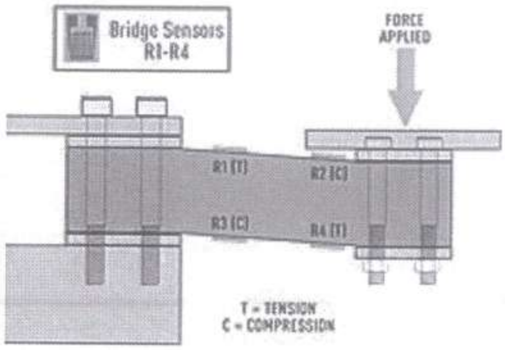


Figure 11: Bridge working

$$V_o = \left(\frac{R3}{R3 + R4} - \frac{R2}{R1 + R2}\right) V_{EX}$$

In a load cell, the resistors are replaced with strain gauges and arranged in alternating tension and compression formation. When force is exerted on the load cell, the structure and resistance of the strain gauges changes and V_O is measured. From the resulting data, V_O can be easily determined using the equation above.

14.3.3 Common types of load cells

There are several types of strain gauge load cells:

- Single Point load cells; used in small to medium platform scales with platform sizes of 200x200mm up to 1200x1200 mm.
- Planar Beam load cells; used in low profile solutions where space is limited, like medical scales and retail scales.
- Bending Beam load cells; used in pallet, platform and small hopper scales.
- Shear Beam load cells; used in low profile scale and process applications, available in capacities from 100kg up to 50t.
- Dual Shear Beam load cells; used in truck scales, tank and hopper applications.
- S-type load cells; used in tension applications where you will find static and dynamic loads.
- Compression load cells; used in truck scales, large platform scales, weighbridges and hopper scales.
- Ring Torsion load cells; used in high accuracy hoppers, silo's, platforms and pallet scales.

- Spoke Type load cells; used in low profile, high precision application. High forces varying from 1t-500t.
- Onboard load cells; used for onboard weighing systems on trucks, tractors and other vehicles.
- Load pins; used in applications for measuring dynamic, static or hoisting forces.
- Weigh pads; portable weigh pads for the weighing of cars and the measure the center of gravity of planes.
- Specials; all kind of special sensors.

- **Pneumatic load cell**

The pneumatic load cell is designed to automatically regulate the balancing pressure. Air pressure is applied to one end of the diaphragm and it escapes through the nozzle placed at the bottom of the load cell. A pressure gauge is attached to the load cell to measure the pressure inside the cell. The deflection of the diaphragm affects the airflow through the nozzle as well as the pressure inside the chamber.

- **Hydraulic load cell**

The hydraulic load cell uses a conventional piston and cylinder arrangement with the piston placed in a thin elastic diaphragm. The piston doesn't actually come in contact with the load cell. Mechanical stops are placed to prevent over strain of the diaphragm when the loads exceed certain limit.

The load cell is completely filled with oil. When the load is applied on the piston, the movement of the piston and the diaphragm results in an increase of oil pressure. This pressure is then transmitted to a hydraulic pressure gauge via a high pressure hose. The gauge's Bourdon tube senses the pressure and registers it on the dial. Because this sensor has no electrical components, it is ideal for use in hazardous areas.

Typical hydraulic load cell applications include tank, bin, and hopper weighing. By example, a hydraulic load cell is immune to transient voltages (lightning) so these types of load cells might be a more effective device in outdoor environments.

This technology is more expensive than other types of load cells. It is a more costly technology and thus cannot effectively compete on a cost of purchase basis.

14.3.4 Other types

- **Vibrating load cell**

Vibrating wire load cells, which are useful in geo-mechanical applications due to low amounts of drift, and capacitive load cells where the capacitance of a capacitor changes as the load presses the two plates of a capacitor closer together.

- **Piezoelectric load cell**

Piezoelectric load cells work on the same principle of deformation as the strain gauge load cells, but a voltage output is generated by the basic piezoelectric material – proportional to the deformation of load cell. Useful for dynamic/frequent measurements of force.

Most applications for piezo-based load cells are in the dynamic loading conditions, where strain gauge load cells can fail with high dynamic loading cycles. The piezoelectric effect is dynamic, that is, the electrical output of a gauge is an impulse function and is not static. The voltage output is only useful when the strain is changing and does not measure static values.

However, depending on conditioning system used, "quasi static" operation can be done. Using a charge amplifier with a long time constant allows accurate measurement lasting many minutes for small loads up to many hours for large loads. Another advantage of Piezoelectric load cells conditioned with a charge amplifier is the wide measuring range that can be achieved.

Users can choose a load cell with a range of hundred of kilo newtons and use it for measuring few newtons of force with the same signal-to-noise ratio; again this is possible only with the use of a charge amplifier for conditioning.

14.3.5 Uses of Load Cell

Load cells are used in several types of measuring instruments such as laboratory balances, industrial scales, platform scales and universal testing machines. From 1993 the British Antarctic Survey installed load cells in glass fiber nests to weigh albatross chicks. Load cells are used in a wide variety of items such as the seven-post shaker which is often used to set up race cars.

14.4 Summary

- A transducer is an electronic device that converts energy from one form to another. The process of converting energy from one form to another is known as transduction.
- The primary function of transducers is to convert a physical force into an electrical signal so that it can be easily handled and transmitted for measurement.
- A Resistance Temperature Detector (also known as a Resistance Thermometer or RTD) is an electronic device used to determine the temperature by measuring the resistance of an electrical wire
- In the RTD resistance, there will be an I^2R power dissipation by the device itself that causes a slight heating effect. This is called as self-heating in RTD.
- When a force is applied, a strain gauge is used as a sensor for measuring variations in resistance, then converting those changes in electrical resistance into measurements.
- When force is applied to the body of the load cell, it flexes slightly under the strain. This is similar to what happens to a fishing rod when a fisherman hooks a fish.

14.5 Keywords

- Sensing Element

It is the part of a transducer that responds to the physical sensation. The response of the sensing element depends on the physical phenomenon.

- Transduction Element

The transduction element of the transducer converts the output of the sensing element into an electrical signal. The transduction element is also called the secondary transducer.

14.6 Review Questions

1. What is a strain gauge?
2. What are the advantages and disadvantages of the strain gauge?
3. What are the types of strain gauges?
4. What is the formula for the Gauge factor?
5. Why strain gauge is important?
6. What is RTD?
7. What is Load Cell?
8. What are the types of Load Cell?
9. What is Wheatstone bridge how it constructed?
10. What is the use of load cell?

14.7 References

1. Basic Electrical and Electronics Engineering by Sukhija and Nagsarkar, Oxford Publication.
2. Basic Electrical & Electronics Engineering by Kothari, Nagrath, TMH
3. Electronic devices & circuits theory, R.L. Boylestad, Louis Nashelsky , Pearson education.
4. Millman, Electronics Devices and Circuits, TMH.
5. Basic Electronics Engineering by Vijay Baru et al, Dream Tech, New Delhi.
6. Fundamentals of Electrical and Electronics Engineering by Ghosh, Smarajit, PHI India.