Indira Gandhi National
Open University
School of Sciences

BPHET-143
DIGITAL AND ANALOG CIRCUITS AND INSTRUMENTATION

Block

1

PHYSICS OF SEMICONDUCTOR DEVICES

Course Design Committee

Prof. A. K. Ghatak, *Retd.*
IIT Delhi, New Delhi

Prof. Suresh Garg, *Retd.*
School of Sciences
IGNOU, New Delhi
Vice Chancellor,
Usha Martin University

Prof. R.M. Mehra, *Retd.*
Dept. of Electronics,
South Campus,
University of Delhi, Delhi

Dr. Ashok Goyal, *Retd.*
Dept. of Physics, Hansraj College
University of Delhi, Delhi

Dr. Parthasarathy
Dept. of Physics,
Maharaja Agrasen College,
University of Delhi, Delhi

Prof. M.S. Nathawat
Former Director,
School of Sciences,
IGNOU, New Delhi

Prof. Vijayshri
School of Sciences
IGNOU, New Delhi

Prof. Sudip Ranjan Jha
School of Sciences
IGNOU, New Delhi

Prof. Shubha Gokhale
School of Sciences
IGNOU, New Delhi

Dr. Sanjay Gupta
School of Sciences
IGNOU, New Delhi

Dr. Subhalakshmi Lamba
School of Sciences
IGNOU, New Delhi

Block Preparation Team

Prof. Vijayshri (Units 1-4)
School of Sciences
IGNOU, New Delhi

Prof. Shubha Gokhale (Unit 5)
School of Sciences
IGNOU, New Delhi

Course Coordinators: Prof. Shubha Gokhale, Prof. Vijayshri

Block Production Team

Mr. Rajiv Girhdar
Assistant Registrar,
MPDD, IGNOU, New Delhi

Mr. Hemant Kumar Parida
Section Officer,
MPDD, IGNOU, New Delhi

Acknowledgement: Parts of this block have been adapted from Units 3 and 4 of the B.Sc. Physics Elective course PHE-10 entitled "Electrical Circuits and Electronics", which were written by Dr. Shubha Gupta and edited by Prof. Abhai Mansingh.

July, 2022

ISBN:

Further information on the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068 or the official website of IGNOU at www.ignou.ac.in.

Printed and published on behalf of Indira Gandhi National Open University, New Delhi by the Registrar, MPDD, IGNOU.

Printed at : Chandra Prabhu Offset Printing Works Pvt. Ltd., Sector – 8, Noida 201301 (U. P.)

CONTENTS

DIGITAL AND ANALOG CIRCUITS AND INSTRUMENTATION: COURSE INTRODUCTION

Electronics plays a very important role is our everyday life. We come across it in a variety of equipment such as radio, television, electronically controlled home appliances, mobile phones, tabs, laptops, desk top computers, etc. Electronics is also widely used in defence, industrial sector, communication and entertainment industry, medical equipment, advanced research apparatus and other areas. It is a fast-developing branch of technology, and integrated circuits and microchips are able to perform and control many important tasks in miniature form.

Electronics has evolved over the last century from glass encapsulated bulky valve circuits to a hand-held powerful multi-tasking gadget like the smart phone that makes use of semiconductor technology. In order to understand and appreciate the wonderful tasks performed by electronic equipment, it is essential to have a thorough grounding in the subject fundamentals.

In this course, we will introduce you to the basic semiconductor devices like diodes, transistors with their construction, working principle and electrical characteristics. Then we will familiarise you with the applications of these devices in various circuits like amplifiers, oscillators, power supplies and the very important branch of digital electronics. This course is divided into four blocks detailed below:

Block 1: Physics of Semiconductor Devices

In this block, you will be learning about the semiconductors in intrinsic and extrinsic (doped) form, various types of diodes formed by junction of *p*- and *n*- type of semiconductors with their applications and transistors formed with two *p*-*n* junctions. You will learn about proper biasing of these devices for various applications. You will also be introduced to an analysis of transistor circuits using hybrid-parameters.

Block 2: Digital Circuits

Semiconductor diodes and transistors are the backbone of the most widely used branch of electronics, namely, digital electronics. Digital devices work only with two steps of input/ output, namely, 0 and 1 or ON and OFF. These circuits called logic gates form the basic part of any computer, laptop, calculator, or recently, the mobile phones. Logic circuits work on the binary number system. In this block, you will also learn the Boolean algebra governing these circuits.

Block 3: Analog Circuits

Transistors have many applications in day-to-day life in devices like amplifiers, oscillators, power supplies, etc. In this block we will discuss some such circuits.

Block 4: Operational Amplifier and Instrumentation

In this block, we will introduce you to an analog electronic circuit that performs mathematical operations like addition, subtraction, integration, differentiation of the input signal. It is called an Operational Amplifier. We will also discuss the CRO that you have used in the earlier physics laboratory course for viewing voltage waveforms.

We hope that you will enjoy studying this course.

BLOCK 1: PHYSICS OF SEMICONDUCTOR DEVICES

Semiconductors are present all around us. These are not directly sold as products but used in most electrical appliances and devices in our homes and surroundings. For example, if you use refrigerators, rice cookers, microwave ovens, these are the materials used in devices that work as temperature sensors. We are sure that you have started using less power consuming LED bulbs now – these use semiconductors. Think of any electronic appliance around you – Computers, laptops, laser printers, smartphones, washing machines, digital cameras, televisions, photocopiers, modern cars, smart homes, etc. and you will find that these work due to chips made of thousands of semiconductor devices. Semiconductor chips and devices are the backbone of modern life.

In this block entitled **Physics of Semiconductor Devices**, you will understand how semiconductor devices such as different kinds of semiconductor diodes and transistors that form the chips work. Essentially, you will learn about the basic physics governing their construction and working. We begin the discussion in Block 1 with an explanation of the basic physical concepts underlying semiconductor devices. The block contains five units.

In **Unit 1** entitled **Essentials of Semiconductor Physics**, we revise the basic concepts of semiconductors related to their structure, doping and the resulting *n*-type and *p*-type semiconductors, the concept of holes, and the transport of charges in them through drift and diffusion, which you know from school physics.

In **Unit 2** entitled **Junction Diodes**, we apply the concepts of Unit 1 to explain the construction and working of a number of junction diodes such as the *p*-*n* junction diode, zener diode, LED, solar cell and the photodiode. These are two-terminal devices that have numerous applications all around us in lighting, electronic appliances, energy production, etc. as you will learn in Unit 2.

Unit 3 is entitled **Transistors**. In this unit, you will learn about two types of transistors: the bipolar junction transistors and the field effect transistors. These devices are used as amplifiers and switches, for example, in satellites and communication systems, space vehicles, power systems, signal generators, logic gates, etc.

In **Unit 4** entitled **Bipolar Junction Transistor Biasing**, you will learn more about the bipolar junction transistors, their configurations, characteristics and biasing methods. These concepts help us understand how to use transistors for particular applications as amplifiers and switches.

In **Unit 5** entitled **Transistor Circuit Analysis**, you will learn how to analyse the transistor amplifier circuits in different configurations (CE, CB and CC) using the hybrid-parameter model of the transistor. You will be able to derive the amplifier parameters like current gain, input and output impedances using *h*-parameters.

In each unit, we have given a Study Guide to help you learn its concepts well.

We hope that you enjoy learning the physics of semiconductor devices and wish you success!

Silicon crystals are the most common semiconducting materials used in modern electronics. The block of silicon shown above hides within it so much of good semiconductor physics about which you will learn in this unit.
Source of picture:
https://en.wikipedia.org/wiki/Semiconductor

UNIT 1

ESSENTIALS OF SEMICONDUCTOR PHYSICS

Structure

STUDY GUIDE

You have learnt about semiconductors in your school physics courses. You know about *p*-type and *n*-type semiconductors. In Experiment 9 of the core laboratory course BPHCL-134, you have learnt the basic concepts about semiconductors, in brief. You have also obtained the *I-V* characteristics of a *p-n* junction diode experimentally, and interpreted them. So, you know the concepts of Secs. 1.2.1 and 1.2.2. We will be dealing with many basic concepts related to semiconductors, which may be new for you. We will refer to Blocks 1 and 3 of the course BPHCT-133 entitled Electricity and Magnetism, especially, Units 12 and 8, while discussing the concepts of drift velocity, drift current, electric field and electric potential in a semiconductor. You may like to revise these concepts before studying Sec. 1.4. We expect you to study all sections thoroughly and note your difficulties to ask your Counsellor. Solve all SAQs and Terminal Questions given in the unit. It will help you learn these concepts better.

"In science one tries to tell people, in such a way as to be understood by everyone, something that no one ever knew before."

P A M Dirac

1.1 INTRODUCTION

We begin the study of the physics of semiconductor devices by revising the basic concepts related to semiconductor physics. You know what semiconductors are from your school physics courses: These are materials having resistivity lower than insulators but higher than conductors. You may recall that conductors, for example, metals like aluminium, iron, silver and gold, conduct electric current readily and have low resistivity. And insulators, for example, materials like wood, paper and plastic, are poor conductors of electric current as these have high resistivity.

You may ask: Why is it important to learn about the physics of semiconductors? It is because these materials are used in all of electronics that forms the backbone of modern life: Electronic communication systems, computer and internet hardware, solar cells, laser diodes, counters, detectors, and so on. You will learn about a few of these devices in this block and the course. So, it is essential that you learn the concepts explained in this unit.

In Sec. 1.2, we quickly revise the basic concepts related to semiconductors that you know from school physics. We begin the unit by recalling the concepts of bonding in semiconductors, and explain what intrinsic and extrinsic semiconductors are (Sec. 1.2.1). You will also revisit the concept of hole in Sec. 1.2.2. In Sec.1.3, we discuss the energy band model of semiconductors, in brief.

In Sec. 1.4, we discuss transport of charge carriers in semiconductors due to **drift**. You are familiar with the concept of drift velocity of electrons in metals from Unit 12 of the course BPHCT-133. You will learn about the drift velocities and drift currents of electrons and holes in semiconductors in low and high electric fields, which help us explain the *I-V* characteristics of semiconductors. The operation of semiconductor devices in both these regions gives rise to interesting applications, which you will learn in Unit 2. There is another interesting way in which charge carriers move in semiconductors: through **diffusion** of electrons and 'holes' leading to **diffusion currents**, which also impact the operation of semiconductor devices. You will learn about this kind of charge carrier transport in Sec. 1.5.

In the next unit, you will learn about the first of these devices, namely, the *p-n* junction diode as also many other useful junction diodes such as the zener diode, LED, solar cell and photodiode.

Expected Learning Outcomes

After studying this unit, you should be able to:

- describe bonding in semiconductors and explain the concepts of intrinsic and extrinsic semiconductors on its basis;
- explain the concept of holes as charge carriers in semiconductors;
- discuss the concepts of energy bands and band gap in semiconductors;

- explain the transport of charge carriers due to drift for low and high electric fields applied to semiconducting materials;
- deduce the expressions for conductivity, resistivity and drift currents in semiconductors for low electric fields;
- draw the *I*-*V* characteristics of semiconductors and explain their main features like saturation and breakdown;
- explain the transport of charge carriers due to diffusion, and write the expressions for diffusion current densities in semiconductors; and
- solve simple numerical problems on calculating conductivity, resistivity, drift and diffusion current densities/currents in semiconductors.

1.2 UNDERSTANDING SEMICONDUCTORS

You have studied the basic physics of semiconductors in your school physics courses. You may like to recall what a semiconductor is. You know that semiconductors are materials that have resistivity values between the resistivity of conductors ($\sim 10^{-7}\ \Omega$m) and insulators ($\sim 10^{12} - 10^{24}\ \Omega$m).

You have learnt some basic physics of semiconductors in Experiment 9 of the core Laboratory course entitled Electricity and Magnetism: Laboratory (BPHCL-134). You know that there are two types of semiconductors: **Intrinsic** and **extrinsic**, and extrinsic semiconductors are of two types: *p*-**type** and *n*-**type**. You have also learnt about covalent bond model to explain why some materials behave as semiconductors. In this section, we will revise these basic concepts about semiconductors and the concept of hole.

Let us first learn: What kind of structure do these materials have that gives rise to this range of resistivity/conductivity and other features? An understanding of bonding in semiconductors answers such questions in a simple way.

1.2.1 Bonding in Semiconductors

You know that a **pure semiconductor is said to be an intrinsic semiconductor**. Examples of the most commonly used intrinsic semiconductors are crystalline silicon and germanium. In recent years, many more intrinsic semiconductors such as compound semiconductors, amorphous semiconductors and semiconducting polymers have been developed.

However, the most commonly used semiconducting materials are the elements silicon, germanium and the compound gallium arsenide. You also know that the *resistivity of intrinsic semiconductors is high*. *Therefore, they are of little use in electronics*.

However, as you know, the resistivity of semiconductors can be decreased (or their conductivity increased) if impurities of appropriate elements are added to them in small quantities. This process is called **doping**. Such doped semiconductors are termed **extrinsic** semiconductors.

Doping is a process of adding small quantities of another element, called impurity, in a pure/intrinsic semiconductor in order to increase its conductivity.

For example, if silicon and germanium are **doped** with a suitable pentavalent atom like phosphorus, antimony or arsenic, their resistivity decreases and conductivity increases by many orders. Thus, these become suitable for many uses in electronic circuits. Using the concept of covalent bonds, let us briefly explain how this is made possible.

Silicon (Si) (atomic number 14) has four valence electrons. The bonding structure of intrinsic silicon is shown in Fig. 1.1. Note from the figure that each silicon atom is sharing one valence electron each with four neighbouring atoms. This kind of sharing of electrons is called **covalent bonding**. This is what provides stability to silicon atoms in a crystal. This is true for germanium as well.

Germanium (Ge) (atomic number 32) too has four valence electrons. So, its structure is the same as that of silicon. Four valence electrons in each atom of Ge are shared by four neighbouring atoms as shown for silicon in Fig. 1.1.

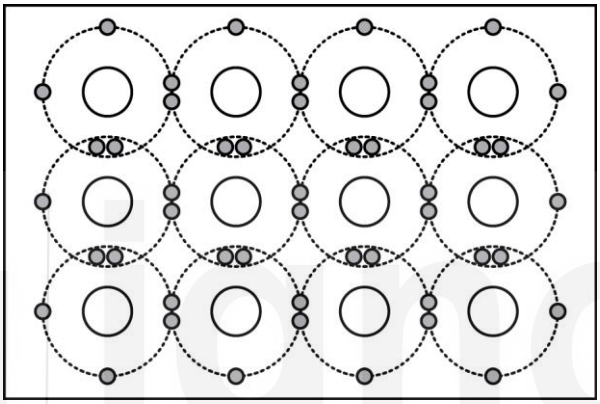


Fig. 1.1: **Covalent bonding in an intrinsic silicon semiconductor.**

When silicon (or germanium) is doped with a pentavalent impurity (that has five valence electrons), four valence electrons of the impurity form a covalent bond with four neighbouring silicon or germanium atoms. The fifth valence electron is not a part of the bond and is free to move in the crystal (see Fig. 1.2).

Thus, when a pentavalent impurity is added to a silicon or germanium crystal, it develops excess free electrons and is said to be an *n*-**type semiconductor**. Such an impurity which results in **excess free electrons** in an intrinsic semiconductor is known as a **donor impurity**. Some examples of donor impurities are: Arsenic (As), phosphorus (P), and antimony (Sb).

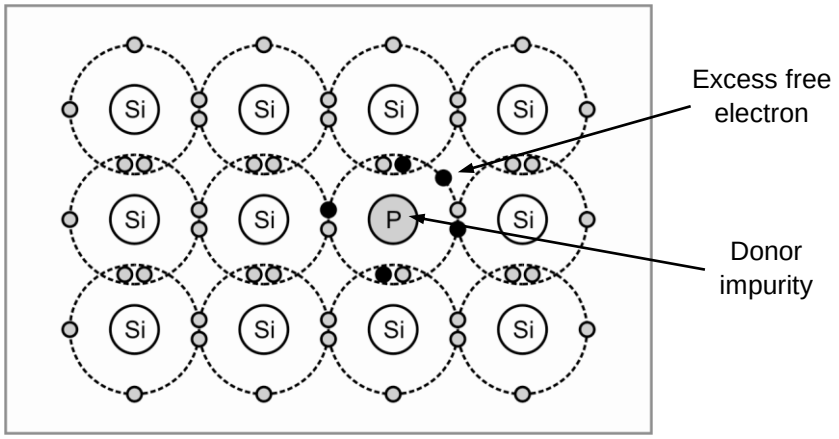


Fig. 1.2: **Structure of an *n*-type semiconductor doped with phosphorus.**

When silicon (or germanium) is doped with a trivalent impurity (that has three valence electrons), the three valence electrons of the impurity atoms form covalent bonds with three neighbouring silicon (or germanium) atoms. Thus, one **deficiency** (**of an electron**) is created per impurity atom doped in the silicon (or germanium) crystal (see Fig. 1.3). This deficiency (of an electron) is termed a **hole** (shown by a small unshaded, unfilled circle in Fig. 1.3).

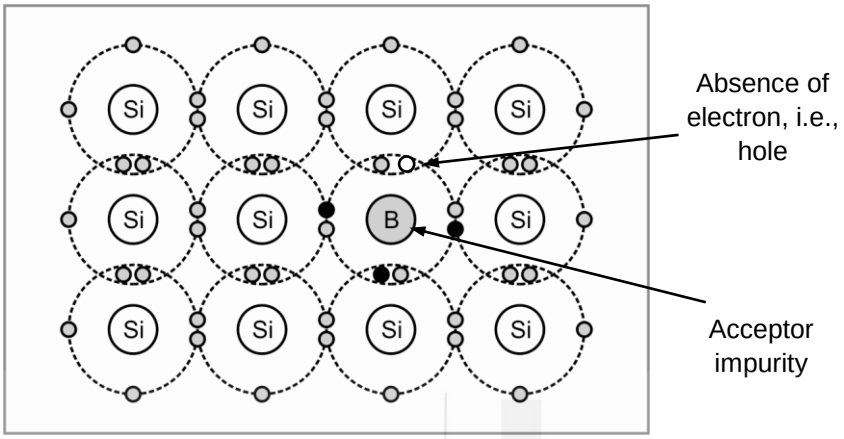


Fig. 1.3: **Structure of a *p*-type semiconductor doped with boron.**

Effectively, holes too act as charge carriers in semiconductors as you will learn in the next section. So, when a trivalent impurity is added to a silicon or germanium crystal, it develops deficiencies of electrons (called holes) and is said to be a *p*-**type semiconductor**. Such an impurity is known as an **acceptor impurity**. Boron (B), aluminium (Al), gallium (Ga) and indium (In) are some examples of acceptor impurities.

Let us now revise the concept of hole.

1.2.2 Concept of Hole

In Sec. 1.2.1, you have learnt about the **hole**, which is a **deficiency or the absence** (**of an electron**) in the covalent bond when the semiconductor is doped with a trivalent impurity.

Even in an intrinsic semiconductor, at room temperature, some of the covalent bonds in its atoms are broken because of the thermal energy supplied to them. Thus, electrons are free to move in the semiconductor crystal lattice. The **absence of electrons in the covalent bonds creates holes** (see Fig. 1.3). We say that electron-hole pairs are created. **The holes also serve as charge carriers in the semiconductor**. Let us explain how.

You have just learnt that there can be an absence of an electron in a covalent bond (due to breaking of the bond at thermal energies in an intrinsic semiconductor or due to doping in a *p*-type semiconductor). When a covalent bond is incomplete, that is, a hole exists, it is easy for an electron in a neighbouring atom to leave its covalent bond and fill the hole. This leaves a hole in the bond that the electron leaves. So, effectively, the hole moves in a direction opposite to that of the electron to a new position.

You can watch a video at the following link for visualising the concepts of Secs. 1.2.1 and 1.2.2:

https://www.youtube.com/watch?v=hE5erfDMXKc

This new position of the hole may be filled by an electron from another covalent bond. So, the hole will move again in a direction opposite to the motion of the electron.

Thus, the motion of an electron from one covalent bond to another in a semiconductor implies the motion of a hole in the opposite direction. In effect, we have a situation in which there is **movement of holes** in a direction opposite to electron motion. This results in a mechanism for the conduction of electric current in a semiconductor, which does not involve free electrons.

So, we say that, in a semiconductor, the **hole acts effectively as a positive charge with magnitude equal to the magnitude of the charge of the electron**. The holes are not actual physical *particles* but they do behave as *physical entities* which possess an effective mass and positive charge; their movement constitutes a flow of current. In sum, current flow in semiconductors results from the movement of two types of charge carriers: electrons and holes. This is a major difference between semiconductors and metals, which have only free electrons as charge carriers.

In an intrinsic semiconductor, the numbers of holes and electrons are equal. Every time an electron leaves a covalent bond, an **electron-hole pair** is created. Every time an electron moves into a hole and completes a bond, we say that the electron-hole pair **recombines** or that the hole disappears due to **electron-hole recombination**. This is a continuous process at room temperature as new electron-hole pairs are **created** due to thermal energies. At the same time, other electron-hole pairs **disappear** due to recombination. Therefore, the electron concentration n (number of electrons per unit volume) and the hole concentration p (number of holes per unit volume) are equal in an intrinsic semiconductor. So, we have:

$$n = p = n_i \tag{1.1}$$

where n_i is called the **intrinsic carrier concentration**.

This is a brief revision of the very basic concepts of covalent bonding in intrinsic semiconductors, extrinsic n-type and p-type semiconductors, and holes in a semiconductor. You may like to solve an SAQ based on them.

SAQ 1 – Bonding in semiconductors

a) What majority charge carriers are created when pentavalent and trivalent impurities are doped in an intrinsic semiconductor? What is the value of the charge that each of these charge carriers carries?

b) What is the intrinsic charge concentration of an intrinsic semiconductor, if the hole concentration in it is $2.5 \times 10^{19}\ \text{m}^{-3}$?

c) Is the hole a physical particle like the electron? Explain.

Now, we explain **the energy band model of intrinsic and extrinsic semiconductors**, in brief. It will help you understand the physics of semiconductor devices, which you will study later in this block.

1.3 ENERGY BAND MODEL

Let us first explain the energy band model for **intrinsic** semiconductors.

Intrinsic semiconductors

In your school physics and the fifth semester physics elective BPHET-141 entitled “Elements of Modern Physics”, you have learnt the quantum mechanical concept of energy level of various systems. You know that electrons in an atom occupy different **discrete** energy levels (see Fig. 1.4a).

Now, a semiconductor is made up of hundreds of thousands of atoms. As atoms in the semiconductor come closer (the inter-atomic spacing decreases), the energy levels of each atom split up, and tightly packed bands called **energy bands** are formed (see Fig. 1.4). For example, the energy levels of electrons in an **intrinsic semiconductor** are **continuous *bands of energy levels*** as shown in Fig. 1.4c rather than the single lines of Fig. 1.4a.

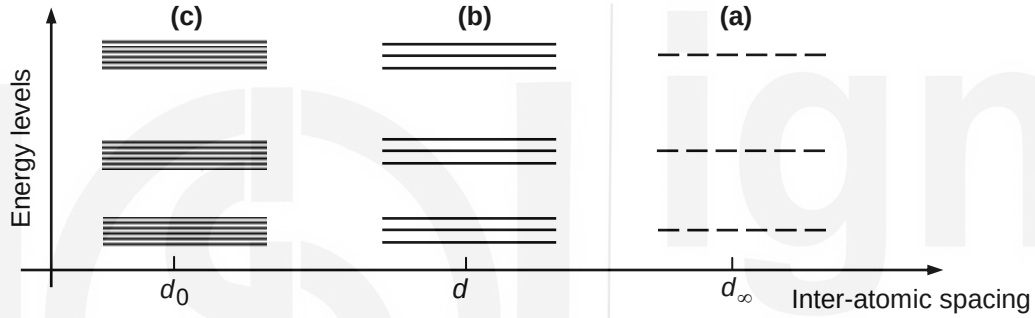


Fig. 1.4: **Schematic diagram showing formation of energy bands in an intrinsic semiconductor; a) discrete atomic levels of an electron; b) aggregates of energy bands are formed as the inter-atomic spacing decreases** $(d < d_\infty)$**; c) for** $d_0 << d$, **continuous energy bands are formed.**

For the purposes of studying this course, it is enough for you to know that there are two types of energy bands in semiconductors: **Valence Band** and **Conduction Band** as shown in Figs. 1.5a and b.

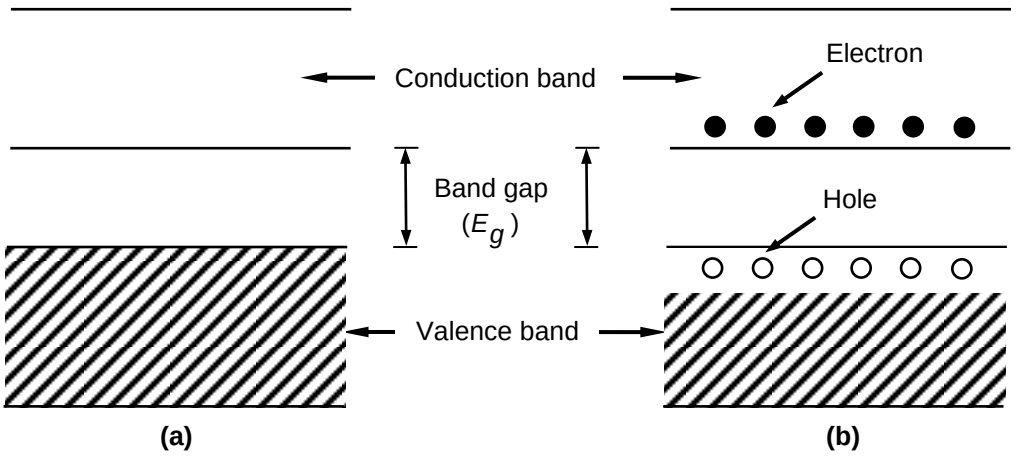


Fig. 1.5: **Energy bands for an intrinsic semiconductor. a) At absolute zero** $(T = 0\,\mathrm{K})$; **b) at** $(T > 0\,\mathrm{K})$**.**

Notice from Figs. 1.5a and b that there is an energy gap between the valence band and the conduction band; it is known as the **band gap**. It represents a

range of energies and **no electrons occupy energy levels in the band gap**. It is a **characteristic** of a semiconducting material and is different for different semiconducting materials.

The values of band gaps (E_g) for some semiconducting materials are given in Table 1.1 in the margin (read the margin remark also).

Table 1.1: Band gaps for a few semiconducting materials

Semiconducting material	E_g (eV)
Si	1.12
Ge	0.67
GaAs	1.42
CdTe	1.44
CdZnTe	1.6

The valence and conduction bands overlap in conductors. In insulators, the band gap is far more than that in semiconductors. It is typically about 15 eV.

Electrons can move into the conduction band from the valence band **on receiving external energy**. When **compared with insulators, the band gap in semiconductors is smaller**. So, electrons in the valence band can move readily into the conduction band on receiving energy equal to or more than the band gap.

At absolute zero temperature, all valence electrons occupy the energy levels in the valence band (Fig. 1.5a). So, the valence band is full and the conduction band is empty at $T = 0\text{K}$, and the semiconductor behaves like an insulator. At room temperature, a few electrons acquire enough energy (more than the band gap) to move into the **conduction band leaving behind holes in the valence band** (Fig. 1.5b). So, the current in a semiconductor at room temperature consists of free electrons moving in the conduction band and holes in the valence band. You should always remember that

Valence band is the energy band involving the energy levels of valence electrons of atoms in the semiconductor. It is **the highest energy band occupied** by charge carriers in a semiconductor.

Conduction band is the **lowest unoccupied band**. The conduction band is made up of high energy levels.

The conduction band and valence band are separated by a **band gap**.

We need to increase the number of electrons in the conduction band and holes in the valence bands to increase the conductivity of a semiconductor. This is done by doping the intrinsic semiconductors as you have learnt in Sec. 1.2.1. Let us understand the energy band model of both *n*-type and *p*-type extrinsic semiconductors.

n-type extrinsic semiconductors

Fig. 1.6 shows the energy band diagram for an *n*-type semiconductor. Recall from Sec. 1.2.1 that an *n*-**type** extrinsic semiconductor is doped by a pentavalent impurity. Now, four valence electrons of the donor impurity atom form covalent bonds with the atoms in the intrinsic semiconductor. So, they occupy energy levels in the valence band.

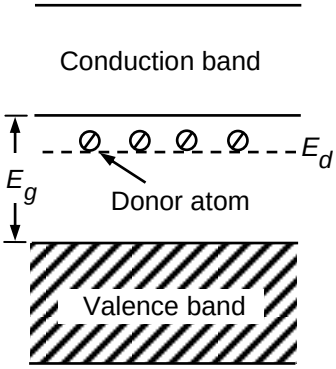


Fig. 1.6: **Energy band diagram for an *n*-type extrinsic semiconductor showing the donor energy level E_d.**

However, the fifth electron of the donor impurity is loosely bound to the donor atom. So, it needs very small thermal energy to become free and available for conduction. We depict this situation in the energy band diagram by introducing, in the band gap, an **allowed energy level** called the **donor energy level** (E_d) very near the conduction band.

At absolute zero temperature, the donor electrons remain bound to the donor impurity, that is, they remain in the donor level (E_d). The conduction band is, therefore, empty at $T = 0\text{K}$. As temperature increases, even at low temperatures, donor electrons move to the conduction band since $E_d < E_g$.

Thus, conduction electrons become available for current flow in an *n*-type extrinsic semiconductor at low thermal energies (0.01 eV for Ge and 0.05 eV for Si).

p-type extrinsic semiconductors

Fig. 1.7 shows the energy band diagram for a *p*-type semiconductor. You know from Sec. 1.2.1 that a *p*-type extrinsic semiconductor is obtained by doping an intrinsic semiconductor with a trivalent impurity. The three valence electrons of the acceptor impurity atom form covalent bonds with the atoms in the intrinsic semiconductor; but a vacancy or hole is created in one covalent bond that remains incomplete.

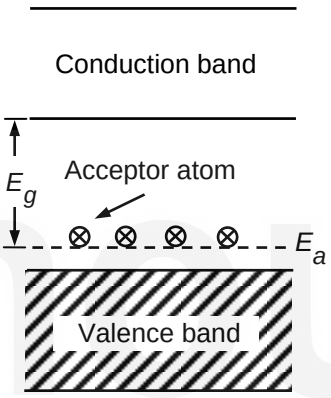


Fig. 1.7: **Energy band diagram for a *p*-type extrinsic semiconductor showing the acceptor energy level E_a.**

The holes can now be assumed to be loosely bound to the acceptor atom and even at very low thermal energies, they become available for conduction. We depict this situation in the energy band diagram by introducing, in the band gap, an **allowed energy level** called the **acceptor energy level** (E_a) very near the valence band.

At absolute zero temperature, the electrons remain bound to the acceptor impurity, that is, they remain in the valence band resulting in no movement of holes. There is, therefore, no current flow at $T = 0\text{K}$.

Note that $E_a < E_g$. Therefore, even at low temperatures or low thermal energies (0.01 eV for Ge and 0.05 eV for Si), electrons in the valence band move to the acceptor energy level leaving behind holes in the valence band for conduction.

So far, you have learnt that the doping of intrinsic semiconductor enhances the number of **majority charge carriers**: electrons in *n*-type semiconductors and holes in *p*-type semiconductors. The **holes are minority charge carriers in an *n*-type semiconductor** while the **electrons are minority charge carriers in a *p*-type semiconductor**.

For the time being, we will consider current flow due to only **majority charge carriers** in doped semiconductors. You should now solve SAQ 2 to find out whether you have understood the energy band model.

SAQ 2 – Energy band model

a) Will the band gap in insulators be more or less than that in semiconductors?

b) Draw schematic energy band diagrams of an intrinsic Ge semiconductor, an *n*-type Si semiconductor and a *p*-type GaAs semiconductor at $T = 0\text{K}$. Take E_a for GaAs to be 100 meV.

We can explain an important feature of semiconductors on the basis of these models: the temperature dependence of resistivity/conductivity. Let us do that before we move to the next section.

Temperature Dependence of Resistivity

What happens in a semiconductor when its temperature is increased? You have learnt in Secs. 1.2.1 and 1.2.2 that when the temperature of a semiconductor is increased, the covalent bonds in it are broken and a large number of free electrons become available for conduction. Or in the energy band model, a large number of electrons acquire sufficient energy to move from the valence band to the conduction band. Thus, many more electrons are free to move within the semiconductor material. This increase in the carrier concentration in the semiconductor leads to a **decrease in its resistivity** or an **increase in its conductivity as its temperature is increased** (Fig. 1.8).

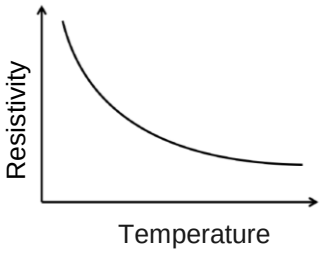


Fig. 1.8: **The resistivity of a semiconductor decreases with temperature for a wide range of temperatures. The graph here shows typical variation for intrinsic semiconductors.**

You know that the resistivity of a conductor increases when its temperature is increased. This is because free electrons in the conductor acquire greater thermal energies and collide continuously with each other. Also, the positive ions in the metal start vibrating about their positions due to the thermal energy. This causes more collisions of electrons with these ions. These factors give rise to greater resistance to the flow of current in the conductor, and a corresponding increase in its resistivity.

You can say that the collisions of these electrons should increase resistivity as in metals (read the margin remark). But the carrier concentration in semiconductors grows exponentially with temperature, and the effect of collisions can be neglected for a wide range of temperatures. Collisions become important at very high temperatures. Thus, we say that **semiconductors have a negative temperature coefficient for a wide range of temperatures**.

Let us briefly list the important differences between metals and semiconductors based on the discussion so far. These are as follows:

- In metals, free electrons are the only charge carriers whereas in semiconductors, there are two types of charge carriers: electrons and holes.
- The resistivity of semiconductors is much higher, and their conductivity much lower than that of metals.
- The temperature dependence of electrical resistivity of metals and semiconductors shows an opposite variation from each other. While the resistivity of metals increases with an increase in their temperature, that of the semiconductors decreases.
- Doping a semiconductor with appropriate impurities increases its conductivity and decreases its resistivity. However, addition of impurity in a metal increases its resistivity.

You should remember that increasing the conductivity by doping makes semiconductors the best materials to fabricate a variety of devices. The characteristics and applications of semiconductor devices can be controlled by doping them appropriately. You will study about some of these devices in the remaining units of this block and the course.

So far, you have learnt how semiconductors are different from metals and insulators. You have learnt about intrinsic and extrinsic semiconductors, covalent bonding in semiconductors and the energy band model. These concepts help us explain many features of semiconductor behaviour.

The next question we ask is: How do charge carriers move inside a semiconductor? What are the mechanisms of charge carrier transport in a semiconductor? What factors affect transport of charges (electrons and holes) in a semiconductor? It is important to understand the mechanisms for transport of charges to understand the working of semiconductor devices and their applications. The **transport of charge carriers in a semiconductor** takes place due to the following three processes:

a) **Temperature gradient**;

b) **Drift due to externally applied electric field**; and

c) **Diffusion due to the concentration gradient**.

We will not consider the first process here since **most semiconductor devices operate at constant temperature**. Therefore, charge transport due to temperature gradient is not relevant to our discussion. In the next two sections, we consider the two mechanisms of charge carrier transport in a semiconductor: Drift and diffusion.

1.4 TRANSPORT OF CHARGE CARRIERS: DRIFT

Do you recall the concepts of drift current and drift velocity of electrons in metals that you have learnt in Sec. 12.2 of Unit 12 of the second semester course entitled Electricity and Magnetism (BPHCT-133)? You can revise that section to understand this section better. We repeat them here, in brief.

You know that electrons move randomly in conductors resulting in zero average velocity. But when a voltage or external electric field is applied to a conductor, free electrons in it acquire a small velocity, in the direction opposite to the electric field. The net flow of electrons in one direction in the conductor in the **presence of an electric field** is called **drift**. The electrons' velocity is called the **drift velocity** and the resulting current**,** the **drift current**. The **drift velocity** of electrons is the **average velocity attained by them, in a conducting material due to an external electric field applied to it**. And it is in a direction **opposite** to the externally applied electric field.

We will use these concepts and basic definitions to study drift currents and drift velocities in semiconductors keeping in mind that there are two types of charges in a semiconductor: electrons and holes.

1.4.1 Drift Currents and Drift Velocities

Let us obtain the expression of the drift current for semiconductors. In Sec. 1.2.1, you have learnt that the majority charge carriers in *n*-type and *p*-type extrinsic semiconductors are electrons and holes, respectively. Let us now see how majority charge carriers are transported in a semiconductor when an external electric field is applied on it.

Suppose that a voltage *V* is applied across a block of semiconductor of area of cross-section *A* and length *L* (see Fig. 1.9 and read the margin remark).

Recall the relation between the applied voltage (potential difference) and the electric field from Unit 4 of BPHCT-133 whence, in this case:

$E = V/L$

Under the influence of the electric field, electrons drift in one direction and holes in the opposite direction. Remember that electrons are negatively charged and holes constitute positive charges. The convention is to indicate current flow in the direction opposite to that of electron flow. Hence, the direction of current flow indicates the direction of hole flow. The result is that the current due to electrons (I_e) and holes (I_h) flows in the **same direction**. This current is called the **drift current**.

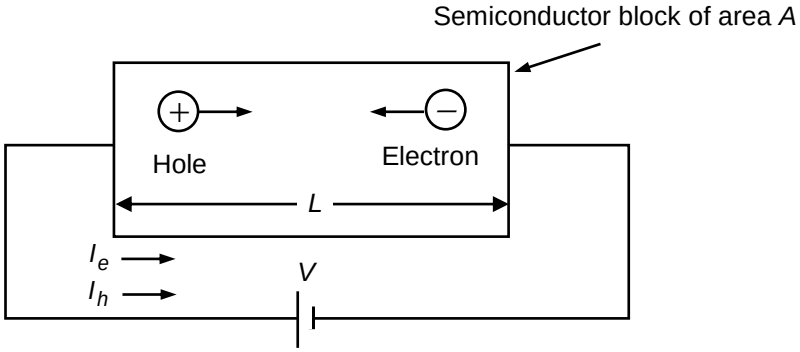


Fig. 1.9: **Drift of charge carriers in a semiconductor due to externally applied electric field.**

Recall Eq. (12.21) of Unit 12, BPHCT-133 for current density J in a metal:

$$J = qnv_d \quad \text{and} \quad I_d = qnv_d A \tag{1.2a}$$

where q is the charge, n, the charge carrier concentration (charge per unit volume), v_d, the drift velocity of electrons in metal. The drift current I_d written in Eq. (1.2a) is simply the product of the current density and area (A) of cross-section of the conductor.

We use the same equations keeping in mind that now there are two charge carriers in semiconductors, electrons and holes. Let the drift velocities of electrons and holes in the semiconductor be v_{de} and v_{dh}, respectively. Let their concentrations be n and p, respectively. So, the drift current due to electrons and holes in a semiconductor and the corresponding current density are:

$$I(\text{drift}) = \frac{V}{R} = JA = q(nv_{de} + pv_{dh})A \tag{1.2b}$$

$$J(\text{drift}) = q(nv_{de} + pv_{dh})E = \sigma E \tag{1.2c}$$

where E is the external electric field and σ, the electrical conductivity as you know from Eq. (12.22) of BPHCT-133. So, the drift current density is proportional to the applied electric field or the **electric potential gradient** (read the margin remark). From Eq. (1.2b), we get the expression for **electrical conductivity** of the semiconductor as:

$$\sigma = \frac{q(nv_{de} + pv_{dh})}{E} = q\left(\frac{nv_{de}}{E} + \frac{pv_{dh}}{E}\right) \tag{1.2d}$$

You know from Eq. (12.20) of BPHCT-133 that the drift velocity is proportional to the applied electric field. So, for semiconductors we can write:

$$v_{de} \propto E \quad \text{and} \quad v_{dh} \propto E$$

Recall from Eqs. (8.20 and 8.21) of Unit 8 of BPHCT-133 that the relation between the potential difference and the electric field is given as:

$$E = -\frac{dV}{dx}$$

for one-dimensional electric fields and

$$\vec{\mathbf{E}} = -\vec{\nabla}V,$$

for three-dimensional fields where $\frac{dV}{dx}$ and $\vec{\nabla}V$ are the electric potential gradients.

The constant of proportionality is known as the **mobility** of the charge carrier And we have:

$$v_{de} = \mu_{de} E \qquad (1.3a)$$

and $$v_{dh} = \mu_{dh} E \qquad (1.3b)$$

whence $$\mu_{de} = \frac{v_{de}}{E} \qquad (1.4a)$$

and $$\mu_{dh} = \frac{v_{dh}}{E} \qquad (1.4b)$$

where μ_{de} and μ_{dh} are called the **drift mobility of the electrons** and **the drift mobility of the holes**, respectively. Thus, by definition, the **charge carrier mobility is drift velocity per unit electric field**. Then we can write Eq. (1.2d) as:

$$\sigma = q(n\mu_{de} + p\mu_{dh}) \qquad (1.5a)$$

The **resistivity of the semiconductor** is given by:

$$\rho = \frac{1}{\sigma} = [q(n\mu_{de} + p\mu_{dh})]^{-1} \qquad (1.5b)$$

In terms of mobility, the drift current density for electrons and holes are, respectively, given by:

$$J_e \text{ (drift)} = q\, n\, \mu_{de}\, E \qquad (1.6a)$$

$$J_h \text{ (drift)} = q\, p\, \mu_{dh}\, E \qquad (1.6b)$$

and the **total drift current density** in a semiconductor is given by:

$$J \text{ (drift)} = J_e \text{ (drift)} + J_h \text{ (drift)} = (q\, n\, \mu_{de} + q\, p\, \mu_{dh})E \qquad (1.7)$$

Let us now consider an example for calculating the conductivity/resistivity of semiconductors.

EXAMPLE 1.1 : CONDUCTIVITY OF SEMICONDUCTORS

Calculate the conductivity and resistivity of intrinsic silicon at 300 K. If an acceptor impurity is added to the extent of 1 part in 10^9 silicon atoms, calculate its conductivity and resistivity. It is given that at 300 K, the intrinsic carrier concentration for silicon is $1.5\times10^{16}\ m^{-3}$ and electron and hole drift mobilities are $0.13\ m^2\ V^{-1}s^{-1}$ and $0.05\ m^2\ V^{-1}s^{-1}$, respectively. The number of silicon atoms per unit volume is given as $5.0\times10^{28}\ m^{-3}$.

SOLUTION ■ We use Eqs. (1.1, 1.5a and b) for these calculations.

For calculating the conductivity and resistivity of intrinsic silicon, we use $n_i = n = p$ from Eq. (1.1) in Eqs. (1.5a and b).

So, we get:

$$\sigma = q n_i (\mu_{de} + \mu_{dh}) = [1.6\times10^{-19}\times1.5\times10^{16}\times(0.13+0.05)]\Omega^{-1}\text{m}^{-1}$$

or $\sigma = 4.3\times10^{-4}\,\Omega^{-1}\text{m}^{-1}$ and $\rho = 1/[4.3\times10^{-4}\,\Omega^{-1}\text{m}^{-1}] = 2.3\times10^{3}\,\Omega\text{m}$

When an acceptor impurity is added to the extent of 1 part in 10^9 silicon atoms, then there are $N_A = 5.0\times10^{28}\,\text{m}^{-3}/10^9 = 5.0\times10^{19}\,\text{m}^{-3}$ acceptor atoms and the hole concentration is $p = 5.0\times10^{19}\,\text{m}^{-3}$

The electron concentration is

$$n = n_i^2/N_A = [(1.5\times10^{16})^2/5.0\times10^{19}]\text{m}^{-3} = 4.5\times10^{12}\,\text{m}^{-3}$$

So, $n << p$ and we can neglect the terms containing n in Eqs. (1.5a and b):

$$\sigma = q p \mu_{dh} = [1.6\times10^{-19}\times5.0\times10^{19}\times(0.05)]\Omega^{-1}\text{m}^{-1} = 0.4\,\Omega^{-1}\text{m}^{-1}$$

and $\rho = 1/[0.4\,\Omega^{-1}\text{m}^{-1}] = 2.5\ \Omega\text{m}$

Note that adding 1 acceptor atom in 10^9 silicon atoms has increased silicon's conductivity by about 1000 times and decreased its resistivity by the same magnitude.

SAQ 3 – Conductivity and resistivity of a semiconductor

An intrinsic semiconductor has charge carrier concentration of $3.5\times10^{16}\,\text{m}^{-3}$. Calculate its conductivity and resistivity if the drift mobilities of electrons and holes are $0.30\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$ and $0.15\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$, respectively.

The results that we have obtained so far hold for relatively low applied electric fields. We now ask: **What happens when the external electric field applied to the semiconductor is high**? It is important for you to understand the answer as it is relevant for understanding the physics of semiconductor devices, their characteristics and applications.

1.4.2 High Field Conduction in Semiconductors

As the external electric field increases, the average energy of the charge carriers increases. At sufficiently high electric fields, a number of physical phenomena occur, such as **saturation of drift velocity**, **breakdown** due to impact ionization, that is, generation of electron-hole pairs, band-to-band tunnelling, etc. These phenomena have important implications for the design and operation of semiconductor devices. We discuss the basic physics of the first two phenomena here as the rest are beyond the scope of this course.

These are:

1. **Saturation of drift velocity**; and
2. **Breakdown**.

We explain both these phenomena briefly.

1. Saturation of drift velocity

You have learnt in this section that **charge carrier mobility is a constant** but this is so **only at low electric fields and for small drift velocities**. At **low electric fields, drift velocity varies linearly with the electric field** as given by Eqs. (1.3a and b). Do remember always that this linear relationship holds only for **low electric fields**.

As the **electric field is increased, the drift velocity departs from the linear relationship**. When a very high electric field is applied to a semiconductor, the drift velocity becomes high, and approaches the thermal velocity of free electrons, which is of the order of 10^5 ms^{-1} or 100 km per second at room temperatures. (Compare it with the speed of the fastest fighter aircraft so far, which is about 1 km per second!) At such high drift velocities, Eqs. (1.3a and b) do not hold.

For such high electric fields, the **charge carrier mobility is no longer constant**. **It becomes dependent on the electric field and decreases due to various scattering mechanisms in the semiconductor**. These mechanisms cause the carrier drift velocity to **saturate** with increasing electric field. (The details of these mechanisms are taught in higher level courses.)

So, at high electric fields, **the charge carrier mobilities and drift velocities are not constant. Mobility of charge carriers decreases, and drift velocity attains a maximum value called the saturation velocity** (v_{sat}). Fig. 1.10 shows a typical variation of drift velocity with electric field for a semiconductor.

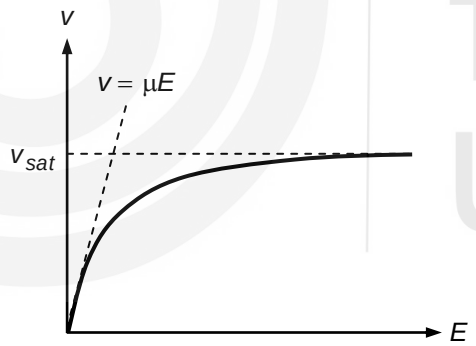


Fig. 1.10: **Typical graph of drift velocity versus electric field in a semiconductor.**

Note in Fig. 1.10 that initially (for low electric fields), the drift velocity varies linearly with the electric field. As electric field increases, the drift velocity does not vary linearly with it. At high electric fields, the curve flattens, i.e., the **drift velocity attains saturation**.

Both electrons and holes exhibit similar behaviour though their saturation velocities are different. Due to the saturation of drift velocity, **drift current becomes independent of the voltage, i.e., the applied electric field**. This happens at typical electric fields in excess of 10^6 V m^{-1}.

The concept of **saturation of drift velocity** at high electric fields is very important in understanding the *I-V* characteristics of semiconductor devices, and, therefore, in their design and operation.

2. Breakdown

When the external electric field is very high and is increased above a certain critical value, it frees electrons from the atoms/molecules of the semiconductor, creating electron-hole pairs. These free electrons are accelerated by the electric field and gain enough energy to ionize other atoms through collisions. This process is called **impact ionization**. The electrons freed in this process collide with more atoms and ionise them creating more free electrons, and so on. Thus, more and more free electrons are generated in such collisions and cause further impact ionization.

So, when high electric field is applied to a semiconductor, electrons and holes are created in large numbers and carrier multiplication takes place continuously in it. This results in a **sudden large increase in the current in the semiconductor**. This sudden increase of current with voltage above a certain critical voltage is called **breakdown**. The critical voltage at which breakdown occurs is called the **breakdown voltage**.

Fig. 1.11 shows a schematic representation of the variation of current density with applied electric field in the ohmic (linear), saturation and breakdown regions. Note the features of the curve shown in the figure. At low electric fields, the variation is linear, i.e., ohmic. So, the straight line *OA* lies in the ohmic region.

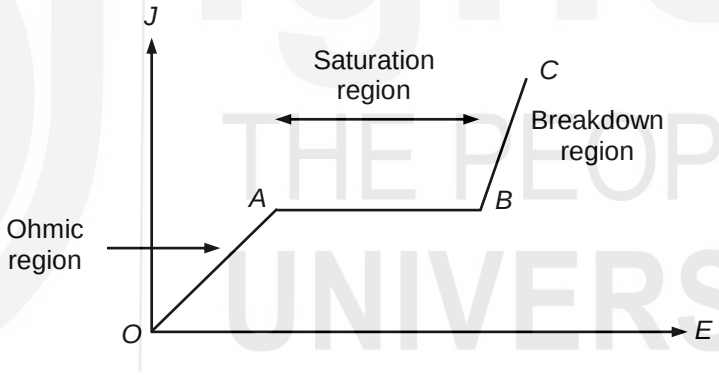


Fig. 1.11: **Typical graph of current density versus electric field in a semiconductor.**

At higher electric fields, the drift velocity is saturated and we get the saturation region from *A* to *B*. In both ohmic and saturation regions, the carrier concentration is constant. The variation in current density is only due to the variation of drift velocity with electric field. As the applied electric field attains a value higher than a critical value, the number of charge carriers increases rapidly and there is a sudden increase in the current density from *B* to *C* in the breakdown region.

An understanding of the phenomenon of breakdown is important in the operation of semiconductor devices like the junction diodes and transistors, and their applications in voltage regulation, power amplification, etc. You will learn about this in the forthcoming units.

We will next discuss the second mechanism of charge transport in a semiconductor. But you should check if you have learnt the concepts of this section well enough. Solve SAQ 4.

SAQ 4 – High field conduction in semiconductors

a) State, giving reasons, whether the following statements are true or false:

i) Drift velocity is independent of applied electric field for all its values.

ii) Breakdown can occur even at low electric fields.

b) The electron concentration in an *n*-type semiconductor is $5.0\times10^{19}\,\text{m}^{-3}$. Calculate the drift velocity and drift current density for electrons for an applied electric field of $1.0\,\text{mV m}^{-1}$ given that $\mu_{de} = 0.20\,\text{m}^2\text{V}^{-1}\text{s}^{-1}$.

Let us now revise the main ideas of this section.

Recap

CARRIER TRANSPORT IN SEMICONDUCTORS: DRIFT

- **Drift of charge carriers** in a semiconductor occurs due to an electric potential gradient (or an external electric field) across the semiconductor.
- **Electrical conductivity** of a semiconductor is given by:

$$\sigma = \frac{q(n v_{de} + p v_{dh})}{E} = q\left(\frac{n v_{de}}{E} + \frac{p v_{dh}}{E}\right) \qquad (1.2d)$$

$$\sigma = q(n\mu_{de} + p\mu_{dh}) \qquad (1.5a)$$

The **resistivity of the semiconductor** is given by:

$$\rho = \frac{1}{\sigma} = [q(n\mu_{de} + p\mu_{dh})]^{-1} \qquad (1.5b)$$

- The **drift current density** is proportional to the applied electric field:

$$J(\text{drift}) \propto E, \frac{dV}{dx}$$

or $J(\text{drift}) = J_e(\text{drift}) + J_h(\text{drift}) = (q n \mu_{de} + q p \mu_{dh}) E \qquad (1.7)$

or $J(\text{drift}) = \sigma E$

For high electric fields, two important phenomena occur in semiconductors, namely, **saturation of drift velocity** and **breakdown**. **Saturation** occurs when the drift velocities acquire large values, carrier mobilities decrease and drift currents due to electrons or holes become independent of the applied voltage. **Breakdown** occurs when the number of charge carriers becomes large due to ionization by high electric field and impact ionization. Then the drift current shows sudden increase with voltage.

Before you begin studying the next section, you should remember that Fig. 1.11 also represents the *I*-*V* characteristics of an intrinsic semiconductor. If you replace the current density by the current ($I = JA$) and electric field by the voltage applied ($E = V / L$) you will obtain a similar curve with an initial

linear (ohmic) region, followed by saturation region and breakdown region. You may like to check if you have grasped this idea. Try SAQ 5.

SAQ 5 – *I-V* characteristics of a semiconductor

Draw the *I-V* characteristics of an intrinsic semiconductor showing the linear, saturation and breakdown regions.

1.5 CHARGE CARRIER TRANSPORT: DIFFUSION

In addition to **drift,** charge transport in semiconductors takes place due to another mechanism called **diffusion**. You have learnt in Sec. 1.4.1 that drift of charge carriers occurs when an electric field is applied. The transport of charge carriers (holes and/or electrons) also takes place due to **non-uniform concentration of electrons and holes in a semiconductor**.

The region in which a greater number of charge carriers is present is called **higher concentration region** and the region in which a smaller number of charge carriers is present is called **lower concentration region**. The change in the concentration of the carrier particles in the material creates a **concentration gradient** in it. Let us learn more about how current flows in a semiconductor due to a concentration gradient.

1.5.1 Concentration Gradient Driven Current Flow

A concentration gradient is set up in a semiconductor when the concentration of charges in one region in it is more than in other regions. Due to this gradient, an electric field is produced in the semiconductor. As a result, the charge carriers move from regions of high concentrations to regions of low concentrations.

This diffusion process is analogous to that which occurs in gases. For example, what happens when you spray perfume or a deodorant in a corner of a room? After a while, don't you smell it in other parts of the room? This is because of the diffusion of these particles in the room.

The process, by which charge carriers (electrons or holes) in a semiconductor move from a region of higher concentration to a region of lower concentration is called **diffusion** (see Fig. 1.12 and read the margin remark).

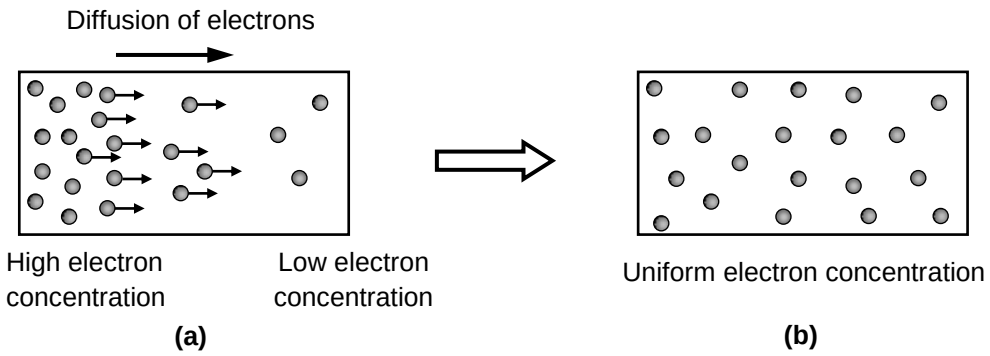


Fig. 1.12: **Diffusion of electrons in an *n*-type semiconductor.**

The flow of charge carriers due to diffusion results in a **diffusion current** in the semiconductor. **Diffusion process occurs in a semiconductor that is**

non-uniformly doped. For example, consider an *n*-type semiconductor that is non-uniformly doped as shown in Fig. 1.12a.

Note that due to non-uniform doping, more electrons are present in the left region of the *n*-type semiconductor initially than in the right region (Fig. 1.12a). The electrons in the left region diffuse to the right region, until a **uniform concentration of electrons** is reached (Fig. 1.12b). The diffusion of holes occurs in the same way in *p*-type semiconductors.

Thus, current flows in the semiconductor due to diffusion of electrons and holes driven by the concentration gradient in the semiconductor. This current is called the **diffusion current**.

Remember that no external electric field across the semiconductor is required for diffusion current to flow in it. This is because diffusion takes place due to the **difference** in concentration of the carrier particles and not the concentrations themselves.

The total current in the semiconductor is the sum of the drift current and the diffusion current. To find the total current in a semiconductor, we have to know the expression for the diffusion current or diffusion current density. Let us do that now.

1.5.2 Diffusion Current Densities and Total Current Density

The diffusion current density is directly proportional to the concentration gradient. Recall that concentration gradient is the difference in concentration of electrons or holes in a given area. So, if the concentration gradient is high, then the diffusion current density is also high. Similarly, if the concentration gradient is low, then the diffusion current density is also low. Therefore, we can write the diffusion current densities for *n*-type and *p*-type semiconductors, respectively, as:

$$J_e(\text{diffusion}) \propto \frac{dn}{dx} \qquad (1.8a)$$

$$J_h(\text{diffusion}) \propto \frac{dp}{dx} \qquad (1.8b)$$

where $J_e(\text{diffusion})$ is the diffusion current density due to electrons, and $J_h(\text{diffusion})$, the diffusion current density due to holes. Replacing the proportionality sign with a constant of proportionality, we get:

$$J_e(\text{diffusion}) = +qD_n\frac{dn}{dx} \qquad (1.9a)$$

and

$$J_h(\text{diffusion}) = -qD_p\frac{dp}{dx} \qquad (1.9b)$$

where D_n and D_p are the constants of proportionality known as the **diffusion coefficients of electrons and holes**, respectively. Note that the negative sign in Eq. (1.9b) appears because the concentration gradient for holes is in

the negative *x*-direction, opposite to that of electrons, which we have taken in the positive *x*-direction. The diffusion coefficients give us a measure of the ease of diffusion of charge carriers. These are related to the mobility of charge carriers through the Einstein relation:

$$\frac{D}{\mu} = \frac{k_B T}{q} \tag{1.10a}$$

For semiconductors:

$$\frac{D_n}{\mu_n} = \frac{k_B T}{q} = \frac{D_p}{\mu_p} \tag{1.10b}$$

The total diffusion current density in a semiconductor is the sum of diffusion current densities of electrons and holes:

$$J(\text{diffusion}) = J_e(\text{diffusion}) + J_h(\text{diffusion}) = (+q D_n \frac{dn}{dx} - q D_p \frac{dp}{dx}) E \tag{1.11}$$

The **total current density** due to electrons and holes is the sum of their respective drift and diffusion current densities:

$$J = J_e + J_h \tag{1.12a}$$

where $$J_e = J_e(\text{drift}) + J_e(\text{diffusion}) = qn\mu_n E + q D_n \frac{dn}{dx} \tag{1.12b}$$

and $$J_h = J_h(\text{drift}) + J_h(\text{diffusion}) = qp\mu_p E - q D_p \frac{dp}{dx} \tag{1.12c}$$

You may like to work out SAQ 6 based on the concepts of Sec. 1.5.

SAQ 6 – Diffusion current in semiconductors

Calculate the diffusion coefficient for electrons in a semiconductor at 300 K if the mobility of electrons is $\mu_n = 0.20\,\text{m}^2\text{V}^{-1}\text{s}^{-1}$. What would be the diffusion coefficient for holes in this semiconductor at 300 K if hole mobility was $\mu_p = 0.10\,\text{m}^2\text{V}^{-1}\text{s}^{-1}$?

With this discussion on diffusion current densities and diffusion currents in semiconductors, we end this unit. Let us now summarise what you have learnt in it.

1.6 SUMMARY

Concept	Description
Intrinsic and extrinsic semiconductors	■ Semiconductors are materials • with **conductivity** much less than that of conductors and more than that of insulators; and with **resistivity** much more than that of conductors but less than that of insulators.

- whose electrical resistivity decreases with an increase in temperature.
- with two types of charge carriers: **electrons** and **holes**.

- **Holes** are created due to the absence of electrons in covalent bonds in the semiconducting material and constitute an effective positive charge.
- **Intrinsic semiconductors** do not contain impurities and the intrinsic carrier concentration n_i is equal to electron and hole concentrations (n and p, respectively) in them:

$$n_i = n = p$$

- Doping a semiconductor with appropriate impurities increases its conductivity and decreases its resistivity, and such semiconductors are called **extrinsic semiconductors**. A semiconductor doped with **pentavalent impurities** has excess electrons as charge carriers and is called n-**type** extrinsic semiconductor. A semiconductor doped with **trivalent impurities** has excess holes as charge carriers and is called p-**type** extrinsic semiconductor.

Energy band model

- According to the **energy band model** of the semiconductor, there are two types of energy bands in semiconductors: **Valence Band** and **Conduction Band** separated by the **band gap** (E_g).
  - In conductors, the valence and the conduction bands overlap and in insulators, the band gap is much larger than that of semiconductors.
  - In n-type semiconductors, donor energy levels (E_d) are created near the conduction band. Thus, electrons require much less energy to move to the conduction band and, therefore, electrons are the majority charge carriers in n-type semiconductors.
  - In p-type semiconductors, acceptor energy levels (E_a) are created near the valence band. Thus, electrons require much less energy to move to the acceptor energy levels, creating holes in the valence band for conduction. Therefore, holes are the majority charge carriers in p-type semiconductors.

Transport of charge carriers

- The **transport of charge carriers in a semiconductor** takes place due to the following processes:
  - **Temperature gradient**;
  - **Drift** due to externally applied electric field; and
  - **Diffusion** due to the concentration gradient.

Drift

- **Drift of charge carriers** in a semiconductor occurs when an electric field is applied to it, that is, due to an **electric potential gradient** across it.
  - The **drift current density** is proportional to the applied electric field:

$$J(\text{drift}) \propto E, \frac{dV}{dx}$$

and is related to conductivity as:

$$J(\text{drift}) = q(n v_{de} + p v_{dh})E = \sigma E$$

- **Conductivity** $\sigma = \dfrac{q(n v_{de} + p v_{dh})}{E} = q\left(\dfrac{n v_{de}}{E} + \dfrac{p v_{dh}}{E}\right)$

- Also, since $v_{de} = \mu_{de} E$ and $v_{dh} = \mu_{dh} E$,

  conductivity $\sigma = q(n\mu_{de} + p\mu_{dh})$

  resistivity $\rho = \dfrac{1}{\sigma} = [q(n\mu_{de} + p\mu_{dh})]^{-1}$

- In terms of mobility, the drift current density for electrons and holes are, respectively, given by:

  $$J_e(\text{drift}) = q n \mu_{de} E$$

  $$J_h(\text{drift}) = q p \mu_{dh} E$$

  and the total drift current density in a semiconductor is given by:

  $$J(\text{drift}) = J_e(\text{drift}) + J_h(\text{drift}) = (q n \mu_{de} + q p \mu_{dh})E$$

- For high electric fields, two important phenomena occur in semiconductors, namely, **saturation of drift velocity** and **breakdown**. **Saturation** occurs when the drift velocities acquire large values, carrier mobilities decrease and drift currents due to electrons or holes becomes independent of applied voltage. **Breakdown** occurs when the number of charge carriers becomes large due to ionization by high electric field and impact ionization. Then the drift current shows sudden increase with voltage.

Diffusion

■ **Diffusion current** flows in a semiconductor due to diffusion of electrons and holes caused by the **concentration gradient** in it due to non-uniform doping.

- The diffusion current density is proportional to the concentration gradient:

  $$J(\text{diffusion}) \propto \frac{dn}{dx}, \frac{dp}{dx}$$

  and $J_e(\text{diffusion}) = +q D_n \dfrac{dn}{dx}$

  $$J_h(\text{diffusion}) = -q D_p \frac{dp}{dx} \quad \text{where} \quad \frac{D_n}{\mu_n} = \frac{k_B T}{q} = \frac{D_p}{\mu_p}$$

The total diffusion current in a semiconductor is:

$$J(\text{diffusion}) = J_e(\text{diffusion}) + J_h(\text{diffusion})$$

Total current density

- The **total current density** due to electrons and holes, respectively, is the sum of their respective drift and diffusion current densities, which are as follows:

$$J_e = qn\mu_n E + qD_n \frac{dn}{dx}$$

and $$J_h = qp\mu_p E - qD_p \frac{dp}{dx}$$

The total current density due to electrons and holes is given by:

$$J = J_e + J_h$$

1.7 TERMINAL QUESTIONS

1. List four factors that affect the conductivity of semiconductors.

2. Fill in the blanks in the following sentences:

   a) In intrinsic semiconductors, the number of electrons …….the number of holes.

   b) The majority charge carriers are ……..in *n*-type extrinsic semiconductors and ……..in *p*-type extrinsic semiconductors.

   c) In *p*-type semiconductors, the electron concentration is …….than the hole concentration.

   d) In *n*-type semiconductors, the electron concentration is …….than the hole concentration.

   e) At room temperatures, conduction occurs in the …….. band in the *n*-type extrinsic semiconductors in and in the ……..band in *p*-type extrinsic semiconductors.

   f) The ……. and ……..energy levels exist in the energy band diagrams of *n*-type semiconductors and *p*-type semiconductors, respectively.

3. The intrinsic carrier concentration in a semiconductor is $n_i = 2.2\times10^{19}\ \text{m}^{-3}$ at a given temperature. Calculate its conductivity given that the electron and hole mobilities are $0.35\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$ and $0.15\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$, respectively.

4. The intrinsic carrier concentration of a semiconductor is $2.0\times10^{16}\ \text{m}^{-3}$. If the mobilities of electrons and holes are $0.24\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$ and $0.06\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$, respectively, calculate its resistivity.

5. The intrinsic carrier concentration in a semiconductor is $3.0\times10^{19}\ \text{m}^{-3}$. It is doped with a pentavalent impurity and the density of donor atoms is 5.0×10^{23} atoms m^{-3}. The electron and hole mobilities are

$0.45\,m^2\,V^{-1}s^{-1}$ and $0.25\,m^2\,V^{-1}s^{-1}$, respectively. Calculate its conductivity before and after the donor impurity is added.

6. Calculate the resistance of an intrinsic semiconductor piece of length 1.0 cm, width 1.0 cm and thickness 1.0 cm at 300 K. It is given that the intrinsic carrier concentration at 300 K is $3.5\times10^{16}\,m^{-3}$ and the mobilities of electrons and hole are $0.40\,m^2\,V^{-1}s^{-1}$ and $0.18\,m^2\,V^{-1}s^{-1}$, respectively.

7. Distinguish between ohmic, saturation and breakdown regions of the *I-V* characteristics of a semiconductor.

8. Electrons are injected at room temperature into one end of a semiconductor of length 2.25 μm. It is observed that the electrons take one-hundredth of a second to reach the opposite end of the semiconductor when the sample is subjected to a voltage of 0.5 μV. What are the electron drift velocity and electron mobility for this semiconductor?

9. At 300 K, the diffusion constants for electrons and holes in a semiconductor are $D_n = 5.5\times10^{-3}\,m^2s^{-1}$ and $D_p = 5.0\times10^{-4}\,m^2s^{-1}$, respectively. What are the mobilities of the electrons and holes at 300 K?

10. The electron concentration in a semiconductor is given by

$$n(x) = 2.0\times10^{16}\exp(-100x)m^{-3}$$

where *x* is measured in mm. It is given that the electron diffusion coefficient is $D_n = 6.4\times10^{-3}\,m^2s^{-1}$. Determine the electron diffusion current density as a function of *x*.

1.8 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. a) Electrons are created when pentavalent impurities are doped in an intrinsic semiconductor. Holes are created when trivalent impurities are doped in an intrinsic semiconductor. The charge on electron is -1.6×10^{-19} C and that on hole is 1.6×10^{-19} C.

   b) In an intrinsic semiconductor, the intrinsic charge concentration is equal to the hole concentration [see Eq. (1.1)]. So, it is $2.5\times10^{19}\,m^{-3}$.

   c) The hole is not a physical particle like the electron. But it may be regarded as a physical entity that behaves like a positively charged carrier in a semiconductor since it moves opposite to the electron in it.

2. a) The band gap in insulators is more than that in a semiconductor.

   b) Schematic diagrams (not to scale) are shown in Figs. 1.13a to c. For intrinsic Ge semiconductor, $E_g = 0.67\,eV$ (Table 1.1). For *n*-type Si semiconductor, $E_g = 1.12\,eV$ and the donor level E_d lies near the

conduction band. For *p*-type GaAs semiconductor, $E_g = 1.42\,\text{eV}$ and the acceptor level E_a lies near the valence band.

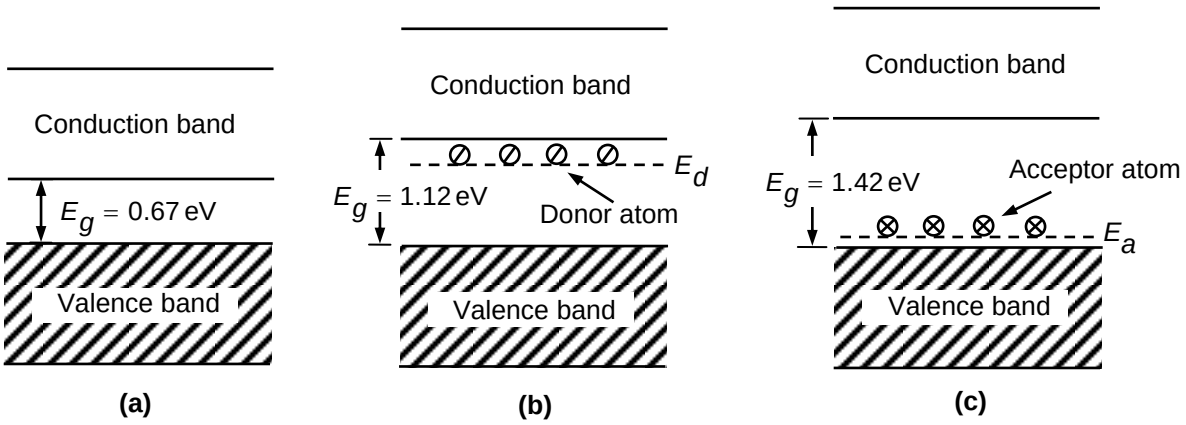


Fig. 1.13: **Energy bands for a) an intrinsic** Ge **semiconductor; b)** *n*-**type** Si **semiconductor; c)** *p*-**type** GaAs **semiconductor at** $T = 0\,\text{K}$.

3. We use Eqs. (1.1, 1.5a and b). For an intrinsic semiconductor, we get from Eq. (1.1) that $n = p = n_i = 3.5\times10^{16}\ \text{m}^{-3}$. From Eq. (1.5a), the conductivity of the semiconductor is:

   $\sigma = [1.6\times10^{-19}\times3.5\times10^{16}\times(0.30+0.15)]\Omega^{-1}\text{m}^{-1} = 2.5\times10^{-3}\,\Omega^{-1}\text{m}^{-1}$

   and from Eq. (1.5b), the resistivity is: $\rho = 1/[2.5\times10^{-3}\Omega^{-1}\text{m}^{-1}] = 400\,\Omega\text{m}$

4. a) i) It is false. This is because only at a high applied electric field, the drift velocity attains saturation and is independent of it. At low electric field, drift velocity varies linearly with it.

   ii) It is false. Breakdown cannot occur at low electric fields as electrons do not acquire enough energy to break free of the covalent bonds.

   b) We use Eqs. (1.3a and 1.5a) to calculate the drift velocity and drift current density for electrons given that $n = 5.0\times10^{19}\ \text{m}^{-3}$, $E = 1.0\,\text{mV m}^{-1}$ and $\mu_{de} = 0.20\,\text{m}^2\text{V}^{-1}\text{s}^{-1}$. Substituting the values of mobility and electric field in Eq. (1.3a and 1.5a), we get:

   $$v_{de} = \mu_{de} E = 0.20\times1.0\times10^{-3}\ \text{ms}^{-1} = 2.0\times10^{-4}\ \text{ms}^{-1}$$

   $$\sigma_e = qn\mu_{de} = 1.6\times10^{-19}\times5.0\times10^{19}\times0.2 = 1.6\,\Omega^{-1}\text{m}^{-1}$$

5. The *I*-*V* characteristic curve for an intrinsic semiconductor is shown in Fig. 1.14:

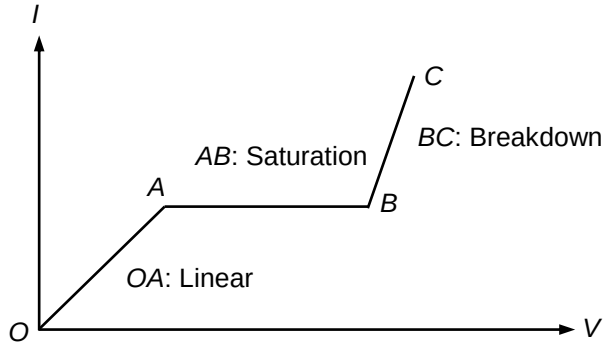


Fig. 1.14: **The** *I*-*V* **characteristic curve for an intrinsic semiconductor.**

6. We use Eq. (1.10b) for calculating the diffusion coefficient for electrons with $T = 300\text{K}$ and $\mu_n = 0.20\text{m}^2\text{V}^{-1}\text{s}^{-1}$. Therefore,

$$D_n = \mu_n \frac{k_B T}{q} = 0.20 \times \frac{1.38\times10^{-23}\times 300}{1.6\times10^{-19}}\text{m}^2\text{s}^{-1} = 5.2\times10^{-3}\,\text{m}^2\text{s}^{-1}$$

We use Eq. (1.10b) and substitute the value of hole mobility to calculate the diffusion coefficient for holes in the semiconductor:

$$D_p = \mu_p \frac{k_B T}{q} = 0.10 \times \frac{1.38\times10^{-23}\times 300}{1.6\times10^{-19}}\text{m}^2\text{s}^{-1} = 2.6\times10^{-3}\,\text{m}^2\text{s}^{-1}$$

Terminal Questions

1. The number of free charge carriers and their concentration, applied electric field, temperature and doping by impurities affect the conductivity of a semiconductor.

2. a) Equals; b) Electrons, holes; c) Less; d) Greater; e) Conduction band, valence band; f) Donor, acceptor.

3. We use Eq. (1.5a) with $n_i = 2.2\times10^{19}\,\text{m}^{-3}$, $\mu_{de} = 0.35\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$ and $\mu_{dh} = 0.15\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$. For an intrinsic semiconductor we put $n = p = n_i$ in Eq. (1.5a) and get:

$$\sigma = q n_i (\mu_{de} + \mu_{dh}) = 1.6\times10^{-19}\times 2.2\times10^{19}\times(0.35+0.15)\,\Omega^{-1}\text{m}^{-1}$$

$$= 1.76\,\Omega^{-1}\text{m}^{-1} \approx 1.8\,\Omega^{-1}\text{m}^{-1}$$

4. We use Eq. (1.5b) with $n_i = 2.0\times10^{16}\,\text{m}^{-3}$, $\mu_{de} = 0.24\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$ and $\mu_{dh} = 0.06\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$. For an intrinsic semiconductor we put $n = p = n_i$ in Eq. (1.5b) and get:

$$\rho = \frac{1}{q n_i (\mu_{de} + \mu_{dh})} = \left(\frac{1}{1.6\times10^{-19}\times 2.0\times10^{16}\times(0.24+0.06)}\right)\Omega\text{m} = 1042\,\Omega\text{m}$$

5. We follow Example 1.1 and use Eq. (1.5a) with $n_i = 3.0\times10^{19}\,\text{m}^{-3}$, $\mu_{de} = 0.45\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$ and $\mu_{dh} = 0.25\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$. Before the donor impurity is added, the semiconductor is intrinsic and therefore, we put $n = p = n_i$ in Eq. (1.5a). So, we get:

$$\sigma = q n_i (\mu_{de} + \mu_{dh}) = 1.6\times10^{-19}\times 3.0\times10^{19}\times(0.45+0.25)\,\Omega^{-1}\text{m}^{-1}$$

$$= 3.4\,\Omega^{-1}\text{m}^{-1}$$

After the donor impurity is added, the conductivity will be due to electrons, and electron concentration is equal to the number of donor atoms per unit volume. Therefore, we substitute $n = 5.0\times10^{23}\ \text{m}^{-3}$ in $\sigma = q n \mu_{de}$ and get:

$$\sigma = 1.6\times10^{-19}\times 5.0\times10^{23}\times 0.45\ \Omega^{-1}\text{m}^{-1} = 3.6\times10^{4}\,\Omega^{-1}\text{m}^{-1}$$

Notice the tremendous increase in conductivity on the addition of impurity.

6. We first calculate the resistivity using Eq. (1.5b) with $n_i = 3.5\times10^{16}\,\text{m}^{-3}$, $\mu_{de} = 0.40\ \text{m}^2\,\text{V}^{-1}\text{s}^{-1}$ and $\mu_{dh} = 0.18\ \text{m}^2\,\text{V}^{-1}\text{s}^{-1}$. For an intrinsic semiconductor we put $n = p = n_i$ in Eq. (1.5b) and get:

$$\rho = \frac{1}{q\,n_i\,(\mu_{de} + \mu_{dh})} = \left(1.6\times10^{-19}\times 3.5\times10^{16}\times(0.40 + 0.18)\right)^{-1}\Omega\text{m} = 308\,\Omega\text{m}$$

You know from your school physics that resistance is related to resistivity as: $R = \frac{\rho L}{A}$, where L is the length of the semiconductor and A, its area of cross-section. We substitute $L = 1.0\,\text{cm} = 1.0\times10^{-2}\,\text{m}$ and $A = 1\text{cm}^2 = 1.0\times10^{-4}\,\text{m}$ and get:

$$R = \frac{\rho L}{A} = \frac{308\times1.0\times10^{-2}}{1.0\times10^{-4}}\,\Omega = 3.08\times10^{4}\,\Omega$$

Note that we have converted the unit of cm into metres everywhere.

7. The ohmic region of the *I-V* characteristics of a semiconductor shows a linear dependence of current on the applied voltage. In the saturation region, the current is constant for the voltage applied. The current rises rapidly in the breakdown region with voltage.

8. Since electrons take $t = 10^{-2}\,\text{s}$ to travel a length of 2.25 μm or $L = 2.25\times10^{-6}\,\text{m}$, their average drift velocity is:

$$v_{de} = \frac{L}{t} = \frac{2.25\times10^{-6}}{10^{-2}}\,\text{ms}^{-1} = 2.25\times10^{-4}\,\text{ms}^{-1}$$

From Eq. (1.3a), we have: $v_{de} = \mu_{de}E$ or $\mu_{de} = \frac{v_{de}}{E}$

We are given the applied voltage and E is the electric field. It is related to the voltage by $E = \frac{V}{L}$ and therefore,

$$\mu_{de} = \frac{v_{de}}{E} = \frac{v_{de}L}{V} = \frac{2.25\times10^{-4}\times 2.25\times10^{-6}}{0.5\times10^{-6}} = 1.0\times10^{-3}\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$$

9. We use Eq. (1.10b) for calculating the mobility of electrons with $T = 300\text{K}$ and $D_n = 5.5\times10^{-3}\,\text{m}^2\text{s}^{-1}$. Therefore,

$$\mu_n = D_n\left(\frac{q}{k_B T}\right) = 5.5\times10^{-3}\times\frac{1.6\times10^{-19}}{1.38\times10^{-23}\times 300}\,\text{m}^2\text{s}^{-1}$$

$$= 0.21\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$$

For hole mobility in the semiconductor, we use Eq. (1.10b) and get:

$$\mu_p = D_p\left(\frac{q}{k_B T}\right) = 5.0\times10^{-4}\times\frac{1.6\times10^{-19}}{1.38\times10^{-23}\times 300}\,\text{m}^2\text{s}^{-1}$$

$$= 1.9\times10^{-2}\,\text{m}^2\,\text{V}^{-1}\text{s}^{-1}$$

10. From Eq. (1.9a), the electron diffusion current density is:

$$J_e\,(\text{diffusion}) = +\,q\,D_n\,\frac{dn}{dx}$$

Since the electron concentration is given by

$$n(x) = 2.0\times10^{16}\exp(-100x)\,\text{m}^{-3}$$

$$\therefore \quad \frac{dn}{dx} = 2.0\times10^{16}\,\frac{d}{dx}\left[\exp(-100x)\right] = 2.0\times10^{16}\,(-100)\exp(-100x)$$

$$= -\,2.0\times10^{18}\,\text{m}^{-4}\exp(-100x)$$

Therefore,

$$J_e\,(\text{diffusion}) = -1.6\times10^{-19}\times 6.4\times10^{-3}\times\ 2.0\times10^{18}\exp(-100x)$$

or $$J_e(\text{diffusion}) = -2.0\times10^{-3}\exp(-100x)\,\text{C}\,\text{m}^{-2}\,\text{s}^{-1}$$

UNIT 2

JUNCTION DIODES

Junction diodes are used all around us. Their applications touch our lives in many ways: In ordinary DC power supplies in our homes, digital displays, chips inside computers and communication systems, light emitting diodes that light up our festivals, solar cells in solar panels, etc. In this unit, you will learn the basic physics of junction diodes that makes them so all pervading.

Source of pictures: https://en.wikipedia.org/
The individual attributions are given below.

Structure

STUDY GUIDE

You have learnt about the *p-n* junction diode in your school physics courses and Experiment 9 of the core Laboratory course BPHCL-134. You have also obtained the *I-V* characteristics of a *p-n* junction diode experimentally, and interpreted them. So, you know the concepts of Secs. 2.2.1 and 2.2.2. You will learn how *p-n* junction diode is used as a rectifier, and other junction diodes such as the zener diode LED, solar cell and photodiodes in Secs. 2.2.3 and 2.3, which maybe new for you. While studying this unit, focus on the way each diode is doped and constructed, and the physics underlying its working. These are what make the *I-V* characteristics of the diodes suitable for specific applications. You should understand their similarities and their differences. Make your own notes, answer all SAQs and Terminal Questions on your own to learn the concepts of this unit well.

Source of pictures above: All pictures in this unit have been taken from https://commons.wikimedia.org/
Individual attributions for the above pictures are: Digital display: Beyond silence; LED TV: Rico Shen; Computer chip: JulianVilla 26; Solar panel: AleSpa - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=29290121.

"A scientist is happy, not in resting on … attainments, but in the steady acquisition of fresh knowledge."

Max Planck

2.1 INTRODUCTION

In Unit 1, you have learnt the basic physics of intrinsic and extrinsic semiconductors. In this unit, you will learn about semiconductor devices called **junction diodes**, of which the *p*-*n* junction diode is the most well known to you. You have performed Experiment 9 in BPHCL 134 for drawing and interpreting the *I*-*V* characteristics of a *p*-*n* junction diode. You have learnt the basic physics of the *p*-*n* junction diode in Secs. 9.3 and 9.4 of Experiment 9.

Junction diodes have a mind-boggling variety of applications in diverse fields. For example, these are used in digital displays (in watches, LED TVs or computer panels), solar panels, voltage regulators, complex chips inside domestic appliances, computer and internet hardware, laser diodes, counters, detectors, communication systems, smoke detectors, remote controls of TVs, and so on.

While studying this unit, you will realise that the working and *I*-*V* characteristics of semiconductor devices are determined by how these are constructed. It is indeed remarkable how so many semiconductor devices can be constructed to have specific *I*-*V* characteristics for desired applications. Continuous research and innovation lead to newer materials for junction diodes, and widens the scope of their applications.

In this unit, you will first revise the basic physics of *p*-*n* junction diodes. We begin the unit by explaining the formation of *p*-*n* junction and the concept of barrier potential (Sec. 2.2.1), which is a revision of Sec. 9.3 of Experiment 9. We also explain the *I*-*V* characteristics of *p*-*n* junction diodes in Sec. 2.2.2 and briefly discuss the application of *p*-*n* junction diodes as rectifiers in Sec. 2.2.3. You will learn about this application in detail in Unit 12.

In Sec. 2.3, we explain the construction, working and the *I*-*V* characteristics of some important junction diodes such as the zener diode, light emitting diode (LED), solar cell and photodetector. We also acquaint you with their numerous applications around us. In the next unit, you will learn about another all-pervasive semiconductor device, the junction transistor.

Expected Learning Outcomes

After studying this unit, you should be able to:

- explain the formation of the *p*-*n* junction and the concepts of depletion region and barrier potential;
- describe the working of a *p*-*n* junction diode, draw its *I*-*V* characteristics and interpret them;
- explain briefly how the *p*-*n* junction diode acts as a rectifier; and
- discuss the construction (in brief), working and *I*-*V* characteristics of zener diode, LED, solar cell and photodiode, and list their applications.

2.2 THE *p-n* JUNCTION DIODE

A *p-n* junction diode was the first semiconductor device constructed with *n*-type and *p*-type semiconductors. It is based on the discovery of the *p-n* barrier by R S Ohl, an American engineer, in 1939. All diodes (including LEDs, solar cells, photodiodes, etc.) followed from the pioneering work done by Ohl.

Understanding the **physics of a *p-n* junction** is important because it is the basic building block of most semiconductor devices, for example, junction diodes like zener diode, LEDs, solar cells, photodiodes, transistors, and integrated circuits. Actually, the electronic action of a semiconductor device takes place at the sites of the *p-n* junctions in it. So, you must learn what actually happens at and near the *p-n* junction.

A *p-n* junction is formed from a single intrinsic semiconducting crystal when controlled amounts of donor and acceptor impurities are added to it. The boundary or interface between the *n*-type and *p*-type parts of the crystal is called the *p-n* **junction** (Fig. 2.1a). A *p-n* **junction diode** is **a two-terminal device formed by this kind of doping** (Fig. 2.1b).

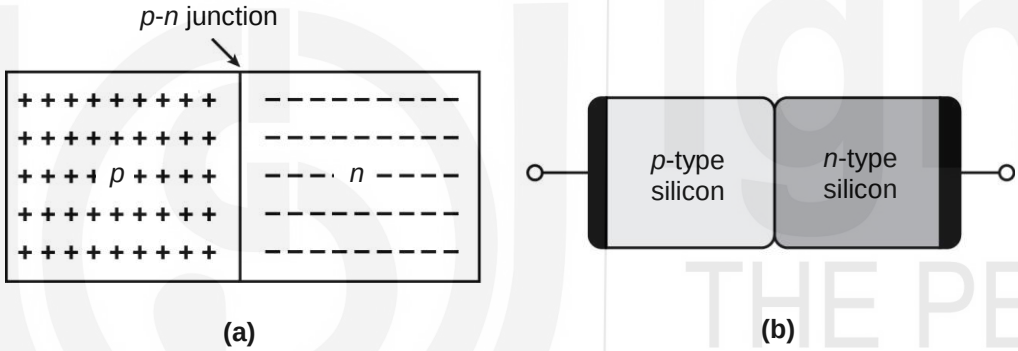


Fig. 2.1: **a) A *p-n* junction is the interface formed when a single intrinsic semiconducting crystal is doped with controlled amounts of donor and acceptor impurities so that holes are the majority charge carriers on the *p*-side and electrons on the *n*-side; b) *p-n* junction diode.**

In Fig. 2.1a, we show only the majority charge carriers on both sides. However, remember that minority charge carriers (holes in *n*-side and electrons in *p*-side) do exist in the doped crystal.

2.2.1 *p-n* Junction Formation and Barrier Potential

Note from Fig. 2.1a that the *n*-region of the semiconducting crystal has a greater concentration of electrons and its *p*-region has a greater concentration of holes. So, a **concentration gradient exists in the semiconductor**. You have learnt in Sec. 1.4 of Unit 1 that this leads to diffusion of electrons and holes from regions of respective higher carrier concentration to lower carrier concentrations. Therefore, electrons *diffuse* from the *n*-region to the *p*-region and holes *diffuse* from the *p*-region to the *n*-region (see Fig.2.2a).
REMEMBER: Such an exchange of mobile carriers occurs mainly in a narrow region around the junction.

Now think: Will this process continue till such time as the number of electrons or holes in both regions becomes equal? The answer is, no, it will not. Why?

To answer this question, we ask: What happens when electrons diffuse from the *n*-region to the *p*-region? Note from Fig. 2.2b that positively charged donor ions are left behind near the *p*-*n* junction on its *n*-side. Similarly, when holes diffuse from the *p*-region to the *n*-region, negatively charged acceptor ions are left behind near the junction on its *p*-side. Thus, the diffusion of electrons and holes leads to the accumulation of positive and negative ions near the junction (Fig. 2.2b).

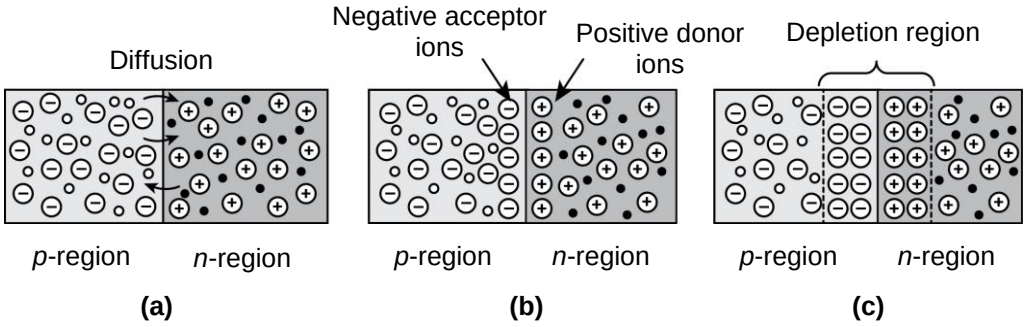


Fig. 2.2: **a) Diffusion of electrons and holes across the *p*-*n* junction; b) accumulation of negative acceptor ions and positive donor ions near the *p*-*n* junction; c) depletion region.**

What happens when positive donor ions and negative acceptor ions accumulate near the junction? We see that an electric field is established near the *p*-*n* junction. **This electric field prevents further movement of majority charge carriers across the junction**: electrons from the *n*-side to the *p*-side and holes from the *p*-side to the *n*-side.

The electrostatic potential associated with this electric field is known as the **barrier potential**. It is called the **barrier potential** because it prevents further movement of majority charge carriers across the *p*-*n* junction **when there is no external electric field**. Barrier potential is a characteristic of the semiconductor material. It is typically 0.7 V for Si and 0.3 V for Ge. It cannot be measured with the help of a voltmeter.

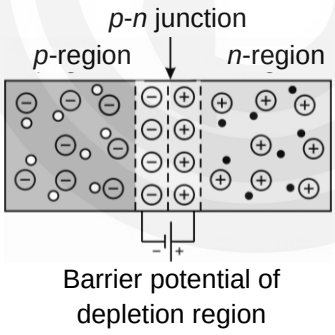


Fig. 2.3: **Barrier potential due to depletion region.**

As a result, a narrow region near the junction is depleted of mobile charge carriers. It is called the **depletion region** or **space-charge region** as **no mobile charge carriers are present in it** (Fig. 2.2c). The width of the depletion region depends on how heavily each side is doped with donor and acceptor impurities, respectively. Typically, it is about 0.5 μm thick. So, the **depletion region acts like a barrier that opposes the flow of electrons from the *n*-region and holes from the *p*-region**.

What is the polarity of the barrier potential? Refer to Fig. 2.3. The barrier potential is of such polarity that it opposes the diffusion of electrons from *n* to *p*-region and the diffusion of holes from *p* to *n*-region. Note from the figure that the barrier potential makes the *p*-side negative with respect to the *n*-side.

However, the barrier potential helps the movement of minority charge carriers. So, holes from *n*-region can move to *p*-region and electrons from *p*-region to the *n*-region. The movement of minority charge carriers constitutes a small drift current depending only on the numbers of minority charge carriers. It is almost independent of the value of the barrier potential (Fig. 2.4a).

So, as the barrier potential builds up, the diffusion current due to majority charge carriers decreases until thermal equilibrium is reached. At thermal equilibrium, **the minority carrier drift current equals the diffusion current** and the **net current across the junction is zero**.

The barrier potential then reaches its steady state value and does not increase any more (see Fig. 2.4b). Let us just give you an idea of the corresponding electric field: For a typical barrier potential (V) of the order of 0.7 V and the barrier width (d) of the order of 0.5 micron, the magnitude of the electric field E is of the order of $10^6\ \mathrm{V\,m^{-1}}$ since $E = V/d$.

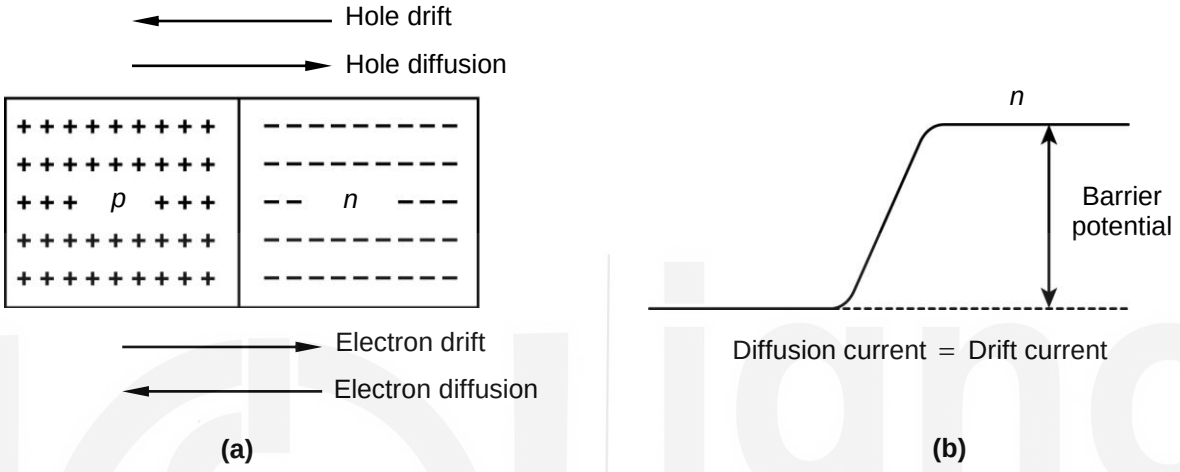


Fig. 2.4: **a) Hole and electron diffusion and drift across the *p-n* junction; b) at thermal equilibrium, the diffusion current equals the drift current and the barrier potential reaches a steady state value.**

An actual *p-n* junction diode is shown in Fig. 2.5 along with its symbol.

The question now is: **How do we overcome the barrier potential so that current can flow across the junction**? In the next section, you will revise how the *p-n* junction diode is biased for current to flow through it and its *I-V* characteristics.

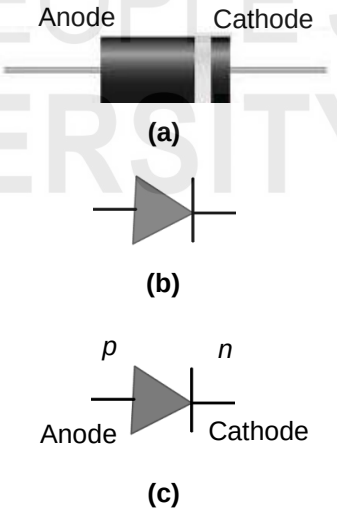


Fig. 2.5: a) An actual *p-n* junction diode with its anode (*p*-side) and cathode (*n*-side) shown; b) circuit symbol of a *p-n* junction diode; c) anode and cathode in a *p-n* junction diode.

2.2.2 *I-V* Characteristics of a *p-n* Junction Diode

You have drawn the *I-V* characteristics of the *p-n* junction diode in Experiment 9 of BPHCL 134. Recall what you did to obtain the curve. You applied an external voltage to the diode to forward bias it, measured the current with increasing voltage. Then you reverse biased the diode and measured the current with increasing voltage. In this section, we explain the physical processes occurring in the forward and reverse biased *p-n* junction diode, which result in its *I-V* characteristic curve.

Forward biasing of the *p-n* junction diode

In Fig. 2.6, we show a *p-n* junction diode with an external battery/power supply connected to it. Notice that the positive terminal of the battery is connected to its *p*-side and the negative terminal to its *n*-side. This is the **forward biasing of the diode**. Such a biasing helps to decrease the barrier potential as the external battery opposes it. Compare Figs. 2.3 and 2.6 and note the opposite polarities of the barrier potential and forward biasing voltage. When the external voltage exceeds the barrier potential, it becomes easier for the

electrons and holes to cross the *p*-*n* junction. Electrons in the *n*-region recombine with the positive ions, and holes recombine with the negative ions near the junction. This reduces the width of the depletion region, which results in majority charge carrier flow across the junction. Electrons on the *n*-side and holes on the *p*-side of the material encounter a reduced barrier potential at the junction due to the decreased width of the depletion region. Further, electrons are attracted to the positive voltage applied to the *p*-side, and holes to the negative voltage applied to the *n*-side of the *p*-*n* junction diode.

As the applied voltage is increased, the width of the depletion region decreases further until a large number of electrons flow through the junction from the *n*-region to the *p*-region and holes from the *p*-region to the *n*-region. This gives rise to a current in the diode from the *p*-region to the *n*-region called the **forward current**. The forward current in the diode is typically of the order of a few milliamperes for applied voltages of less than 1 V. The forward current rises exponentially [see Eq. (2.1a) on the next page]. The drift current due to minority charge carriers also flows across the junction.

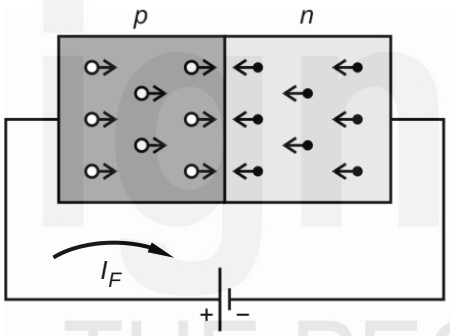


Fig. 2.6: **Forward biasing of a *p*-*n* junction diode.**

In forward bias, the *p*-*n* **junction diode offers low resistance to the flow of current**. The value of the junction resistance, called the **forward resistance**, is in the range 10 Ω to 30 Ω.

Reverse biasing of the *p*-*n* junction diode

When the terminals of the battery are reversed, that is, we connect the *p*-end to the negative terminal of the battery and the *n*-end to its positive terminal as in Fig. 2.7, the *p*-*n* junction diode is said to be **reverse biased**.

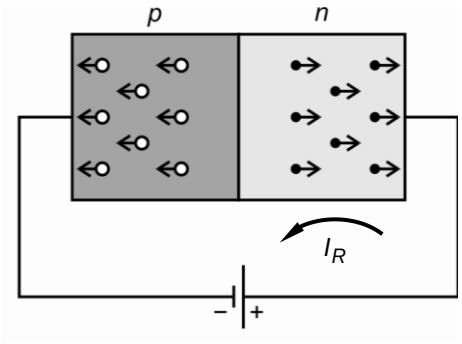


Fig. 2.7: **Reverse biasing of a *p*-*n* junction diode.**

In this case, the electrons in the *n*-region and the holes in the *p*-region move away from the junction, which means that no current should flow across the

junction. But is it really so? Does no current flow at all when the junction is reverse biased? Actually, a small current does flow across the junction. This is the drift current due to **minority charge carriers**, which remains unaffected. The minority charge carrier drift current is called the **reverse saturation current** or **leakage current**. It is of the order of a few nanoamperes to microamperes. So, in reverse bias, the *p-n* junction diode offers high resistance to the flow of current in the range 10 kΩ to 100 kΩ. Let us write the expressions for the forward current and reverse saturation current through the diode.

The forward current is given by:

$$I_f = I_S \exp\left(\frac{qV}{k_B T}\right), \qquad (2.1a)$$

where I_S is the reverse saturation current, V, the applied voltage, k_B, the Boltzmann constant and T, the temperature.

Under reverse biasing, the diode current is just the reverse saturation current flowing in the opposite direction:

$$I_r = -I_S \qquad (2.1b)$$

Combining Eqs. (2.1a and b), we can write the expression for the diode current as:

$$I = I_S \exp\left(\frac{qV}{k_B T} - 1\right) \qquad (2.2)$$

The derivation of these equations is beyond the scope of this course (it is done in the basic solid state physics courses). Fig. 2.8 shows the *I-V* characteristics of a *p-n* junction diode.

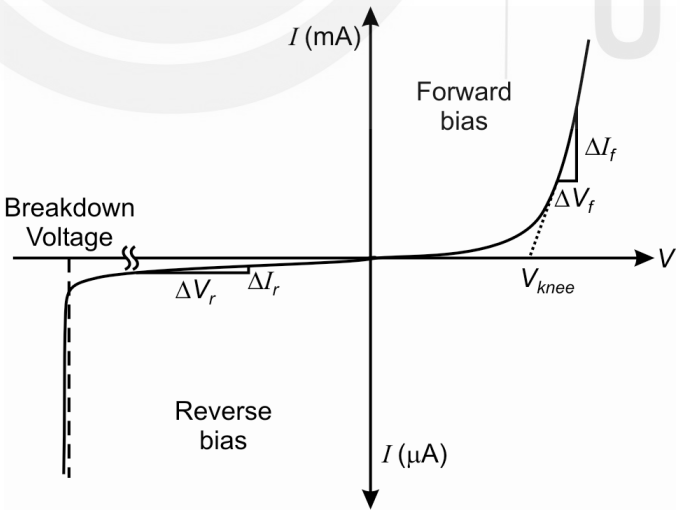


Fig. 2.8: *I-V* **characteristics of a** *p-n* **junction diode.**

Note that for the first few tenths of a volt in forward bias, the current through the diode is very small; the diode does not conduct. When the applied voltage becomes greater than the barrier potential, (0.7 V for Si and 0.3 V for Ge), the current through the diode starts increasing rapidly. The **forward voltage at**

which the flow of the current through the *p-n* junction of the diode increases rapidly is called the knee voltage. When we extrapolate the linear part of the forward bias characteristic curve to meet the *x*-axis, it intercepts the *x*-axis at a point. The intercept gives the value of **knee voltage**.

Note that the *I-V* characteristic curve of the *p-n* junction diode is not a straight line. We say that the *p-n* junction diode is a **nonlinear device**, that is, the current through it does not vary linearly with applied voltage.

The slope of the curve in the region $V > 0$ gives the **forward resistance**.

When the diode is reverse biased, the current is very small (~ microamperes) for voltages less than the breakdown voltage. At **breakdown voltage**, which is the maximum reverse bias voltage that can be applied to a *p-n* junction diode, the current increases rapidly. You have learnt about breakdown due to impact ionisation in a semiconductor in Sec. 1.4.2 of Unit 1. This is called the **avalanche breakdown**. Note that at this point of the curve, the breakdown voltage remains almost constant for a large change in the current. You will learn more about this in the next section. Let us now briefly discuss the application of *p-n* junction diode as a rectifier. You will learn about it in detail in Unit 12 of this course.

2.2.3 The *p-n* Junction Diode as a Rectifier

A large number of electronic circuits require dc voltage for operation. Usually, the alternating voltage or ac from the mains supply is converted into direct voltage or dc. The **rectifier** is a major component of such instruments (Fig. 2.9), which converts an alternating current (ac) into direct current (dc). Fig. 2.9 shows an ideal dc, which is obtained after filter circuits filter the output of a *p-n* junction diode which is shown later in Fig. 2.10.

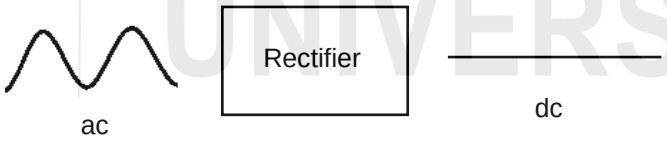


Fig. 2.9: A rectifier with *p-n* junction diodes and filters converts ac into dc.

One important application of a *p-n* **junction diode is as a rectifier**. You may ask: How does a *p-n* junction diode act as a rectifier?

You have learnt in Sec. 2.2.2 that the *p-n* junction diode allows electric current to flow in a circuit **only in one direction when it is forward biased** but not in the opposite direction when it is reverse biased (the reverse saturation current is extremely small). Here we consider an ideal diode for which the reverse current is zero and the forward resistance is also zero. This unique property of the *p-n* junction diode (stemming from its *I-V* characteristics) allows it to act like a rectifier.

So, suppose a sinusoidal alternating voltage source is connected in the circuit shown in Fig. 2.10. Then, **during the positive half cycle** of the sinusoidal voltage, the *p-n* junction diode is forward biased and electric current flows through the circuit.

During the **negative half cycle** of the sinusoidal voltage, the diode is reverse biased and no current flows through the circuit.

Thus, current flows across the diode and in the circuit only for the half positive cycle of ac input and the output of the diode is not an alternating signal (see Fig. 2.10). Rather, it is a truncated signal. The output is not a direct current or voltage. But this is just to give you an idea of how a *p-n* junction diode acts as a rectifier. You will learn the details about this in Unit 12.

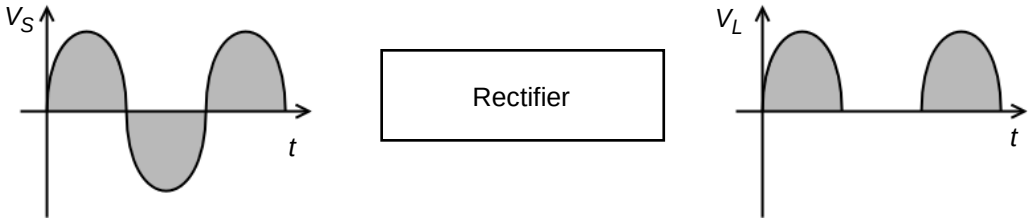


Fig. 2.10: **Action of *p-n* junction diode as a rectifier; a sinusoidal input voltage is converted to a unidirectional pulsating output voltage.**

You may now like to revise the concepts of Sec. 2.2 before you study about other diodes. Solve SAQ 1.

SAQ 1 – *p-n* Junction diode

1. What happens to the depletion region when the i) forward biasing voltage, and ii) the reverse biasing voltage across a *p-n* junction diode is increased?
2. A forward voltage of 0.5 V is applied across a Ge *p-n* junction diode. Will current flow in the diode?
3. The knee voltage of GaAs is 1.2 V. When will a diode made of GaAs start conducting?

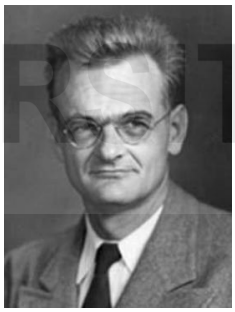
Let us now study some other types of junction diodes.

2.3 OTHER JUNCTION DIODES

In this section, you will learn about some other special purpose junction diodes, which have applications in many fields. These are: Zener diode, LED, solar cell and photodetector. For each one of these diodes, we explain their construction, working and *I-V* characteristics.

2.3.1 The Zener Diode

The ordinary *p-n* junction diode about which you have studied in Sec. 2.2 **does not operate in the breakdown region**. But the **zener diode is designed for operation in the breakdown region**. It is the mainstay of voltage regulators as you will learn in Unit 12. Zener diodes are actually **very heavily doped** *p-n* junction diodes made of silicon. In a zener diode, both *p*-type and *n*-type silicon are doped more heavily than in an ordinary *p-n* junction diode. Its construction is like the *p-n* junction diode shown in



The zener diode is named after **Clarence M. Zener** (1905 –1993) who was an American theoretical physicist. He first described the breakdown of electrical insulators. His research on breakdown was used later by Bell Labs to develop the zener diode.

Fig. 2.3 except that the depletion region in the zener diode is very thin (<1 micron) due to heavy doping. So, when a small reverse bias voltage of about 5 to 6V is applied, the corresponding electric field in the depletion region of width *d* is very high: $E = \frac{V}{d} = \frac{6\,\text{V}}{10^{-6}\,\text{m}} = 6 \times 10^{6}\,\text{V m}^{-1}$.

The *I-V* characteristics of a zener diode are shown in Fig. 2.11. Let us understand the working of a zener diode that gives rise to these characteristics.

In the forward region, the zener diode conducts current from around 0.7 V (knee voltage for silicon) like a conventional *p-n* junction diode. At reverse voltages between zero and breakdown voltage (the leakage region), a small leakage current flows through the diode.

A zener diode is designed to operate in the breakdown region of its *I-V* characteristics. Manufacturers vary the doping levels of silicon diodes to manufacture zener diodes with breakdown voltages ranging from 2 V to 200 V.

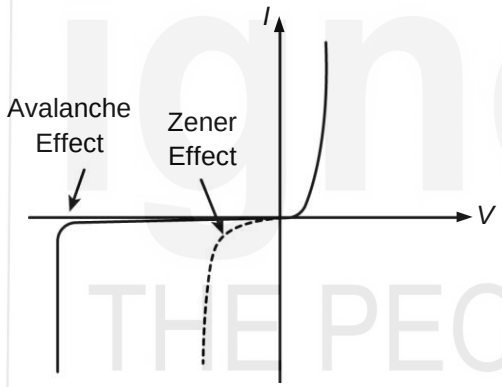


Fig. 2.11: *I-V* **characteristics of a zener diode.**

Note that Fig. 2.11 shows two breakdown regions: **Zener breakdown** and **avalanche breakdown**. These are different effects corresponding to different breakdown voltages that result from the differences in doping levels. You have learnt about the phenomenon of breakdown in Sec. 1.4.2 of Unit 1. The basic physics is the same.

Zener breakdown

Let us understand: What happens when the *p*-side and *n*-side of the *p-n* junction diode are doped heavily? Heavy doping of the *p-n* junction diode results in **a relatively thin depletion region**. Why is it so?

This is because due to heavy doping, a large number of electrons and holes diffuse and so, a **large number of positively charged donor ions and negatively charged acceptor ions** accumulate near the junction. This results in a **strong electric field across the junction**. Since the electric field is very strong, *only a thin depletion region is enough to prevent further diffusion of electrons and holes across the junction*.

Now, if we reverse bias the diode, the electric field in the depletion region becomes much stronger, and generates electron-hole pairs in large numbers

in the diode. This causes a rapid increase in the diode current and results in breakdown in the diode. You have learnt about this kind of breakdown due to high electric fields in Sec. 1.4.2. So, even at lower reverse voltages of 2 or 3 V, minority charge carriers flow across the junction and **increase the reverse current**.

This phenomenon of **breakdown at lower reverse voltages** is called **zener breakdown**. Zener breakdown voltages are typically 2 V to less than 6 V. Remember that zener breakdown takes place **due to very strong electric fields and a thin depletion region** caused by heavy doping. However, the increase in reverse current in zener breakdown is not as rapid and sudden as in avalanche breakdown, which involves impact ionisation about which also you have learnt in Sec. 1.4.2.

Avalanche breakdown

Zener diodes with less heavily doped *p*-type and *n*-type regions have wider depletion layers and, therefore, higher breakdown voltages (greater than 6 V). Notice in Fig. 2.11 that at voltages less than the reverse breakdown voltage, small reverse leakage current flows. As the reverse voltage approaches the reverse breakdown voltage, the electrons and holes ionised by the strong electric field **collide with other atoms creating more electron-hole pairs**, **which ionise more atoms**. This is the process of **impact ionisation** as you know from Sec. 1.4.2. A large number of electron-hole pairs created by impact ionisation result in a very rapid rise in the reverse current in the diode and breakdown. This process is called **avalanche breakdown**.

Note from Fig. 2.11 that avalanche breakdown occurs at much higher voltages than zener breakdown. Notice also that the avalanche breakdown region has a very sharp knee characterised by an almost vertical increase in reverse current. Therefore, the voltage is almost constant over most of the breakdown region for a large range of reverse current. So, large changes in diode current result in very small changes in diode voltage.

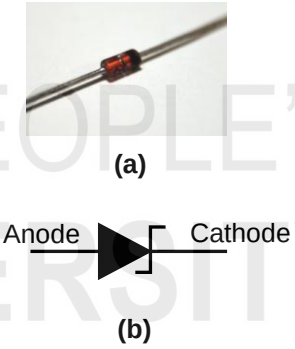


Fig. 2.12: a) An actual zener diode; b) circuit symbol of a zener diode.

Note further that, in zener breakdown, electrons and holes are created due to very strong electric field. It does not involve impact ionisation, which occurs in avalanche breakdown.

So, when a zener diode is reverse biased and the source voltage is greater than the breakdown voltage of the diode, its output voltage remains constant for a large variation in current. That is what makes the zener diode suitable for its application as a voltage regulator as you will learn in Unit 12. The symbol of zener diode is shown in Fig. 2.12. You may now like to solve an SAQ to fix these ideas about a zener diode.

SAQ 2 – Zener diode

Distinguish between a zener diode and a conventional *p*-*n* junction diode. What is the main difference between the processes that lead to zener breakdown and avalanche breakdown? Explain your answers.

Let us now discuss the light emitting diode or LED.

2.3.2 LED

An **LED** or **light emitting diode** is a *p-n* junction **diode that emits light when it is forward biased**. You may like to know: **How are LEDs constructed**? **How do LEDs work**?

LEDs are constructed using **heavily doped** *p-n* junction diodes made of **special compounds of semiconductors**. Due to heavy doping, the depletion region in LEDs is very thin as you have learnt in Sec. 2.3.1. Electrons and holes recombine in the junction region and emit light because these materials have band gaps suitable for light emission. We will explain this point shortly. Some of these materials are listed in Table 2.1, along with the colours produced by them in light from an LED.

Table 2.1: Some typical LED materials that give rise to different colours (wave lengths in nm)

LED material	Colour
Gallium Phosphide (610 – 760)	Red
Gallium Arsenide Phosphide (570-590)	Yellow
Gallium Indium Phosphide (500-570)	Green
Indium gallium Nitride (400-450)	Violet

Depending upon the application, the desired colour and brightness of light, efficiency, etc., manufacturers decide the size, material to be used, extent of doping of the LEDs, etc. Figs. 2.13a and b are schematic diagrams showing the construction of an LED. Metal contacts are provided through wires attached to the LED and then it is encased in a hard plastic shell, which also acts as a lens to focus the light emitted by the diodes (as in Fig. 2.13b).

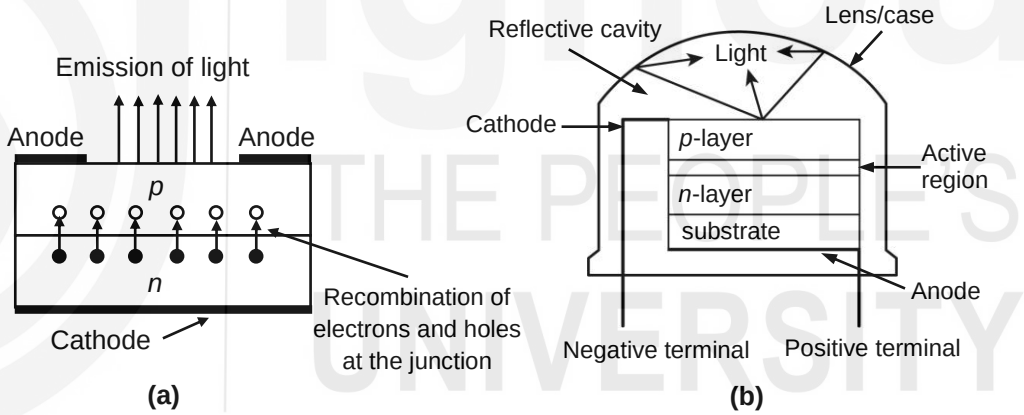


Fig. 2.13: a) Schematic diagram showing construction of LED; b) actual construction schematic of an encased LED.

You may be wondering: Why materials other than silicon or germanium are needed to construct light emitting diodes?

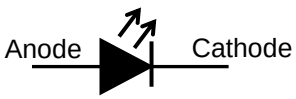
The answer lies in the working of LEDs. Let us understand its working based on its structure shown in Fig. 2.13a. You have learnt about generation and recombination of electron-hole pairs in semiconductors and *p-n* junction diodes. The energy of photons in the visible region corresponds to 2 to 3 eV. You can do the calculations of corresponding wavelengths yourself using the relation $E = hc/\lambda$ or work out SAQ 3e. Recall that you have done such calculations in BPHCT-137 and BPHET-141.

In *p-n* junction diodes made of materials like gallium arsenide phosphide (GaAsP) and gallium phosphide (GaP) that have *direct band gap* (read the margin remark), energy is emitted in the form of **light** (rather than heat) **when electron-hole pairs recombine in the thin junction region**. By suitably doping semiconducting materials like gallium arsenide, etc., LEDs that radiate

The most useful semiconducting materials for LEDs are the ones with **direct band gaps**. In such semiconductors, electrons and holes on either side of the band gap have the same value of crystal momentum and thus direct recombination of electrons and holes is possible. The detailed explanation of the physics of direct band gap semiconductors is beyond the scope of this course.

red, green, yellow, orange, violet light or infrared are produced. You may have seen these colours in strings of LEDs in festival lights. If the semiconductor is translucent (colourless), then **light is radiated through the junction**. Such a junction diode becomes a **source of light** and is called a **light emitting diode** (**LED**). Blue LEDs and the circuit symbol of an LED are shown in Figs. 2.14a and b.

(a)



(b)

Fig. 2.14: a) An actual LED; b) circuit symbol of LED.

Let us explain the process of emission of light in detail with the help of the energy band diagram. LED is basically a **forward biased** *p-n* junction diode in which the depletion region width and the resulting potential barrier across the junction are reduced. Electrons from the *n*-type region and holes from the *p*-type region flow more readily across the junction into the opposite type region. Thus, minority charge carriers are effectively injected across the junction by the application of the external voltage and a current is formed. The increased concentration of minority charge carriers in the opposite type region in the **forward biased LED** leads to recombination of charge carriers across the band gap.

Fig. 2.15 shows this process for a direct band gap semiconducting material. Note in Fig. 2.15 that the normally empty electron states in the conduction band of the *p*-type material and the normally empty hole-states in the valence band of the *n*-type material are populated with electrons and holes, respectively. These injected charge carriers recombine across the band gap and the energy released by this electron-hole recombination is equal to the band gap energy. It is released as light in the visible region for suitable semiconducting materials. An LED begins to emit light when voltage of more than 2 or 3 V is applied in the forward direction.

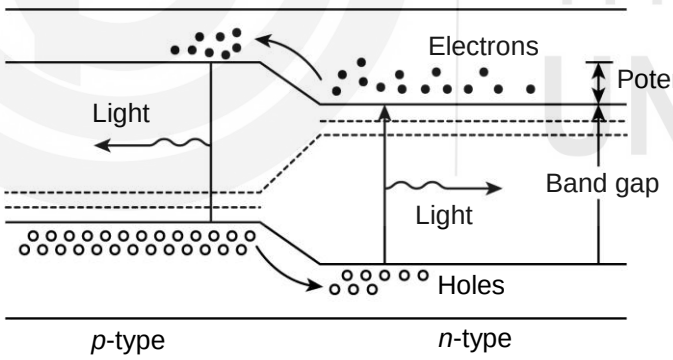


Fig. 2.15: Energy band diagram showing emission of light in an LED.

Thus, spontaneous light emission by charge carrier recombination is the working principle of an LED. The *I-V* characteristics of an LED are the same as those of an ordinary *p-n* junction diode (see Fig. 2.11). However, **LEDs are operated only in the forward bias, never in reverse bias**. **If operated in reverse bias, LEDs are damaged**.

You may like to know: How are the colours in LEDs produced? Recall from the physics course entitled Waves and Optics (BPHCT-137) that each colour of light has a unique wavelength. **The wavelength of light emitted by an LED depends on the band gap of the semiconducting materials used for constructing it**.

Fig. 2.16: Applications of LEDs in automobile lamps and for lighting homes and streets.
Source of pictures: commons.wikimedia.org

The band gaps in LEDs correspond to energies of electromagnetic radiation in the near infrared, visible or near-ultraviolet regions depending on their applications. In fact, the light emitted from an LED is **not monochromatic** but **its spectral width is so small that our eyes perceive it as a single colour**.

Initial LEDs were made from GaAs semiconductors, which emitted infrared and red wavelengths. Continuous advances in materials science have resulted in LEDs that emit light in a variety of colours used for lighting and display purposes. Now-a-days, there are quantum dot LEDs. Recall that you have learnt about quantum dots in Unit 9 of the course Elements of Modern Physics (BPHET 141). You know that quantum dots are semiconductor nanocrystals and can be tuned to emit any colour in the visible and infrared radiation.

LEDs have many applications all around us: For general lighting in homes and surroundings, in strings of festival lights, garden lights, automobile headlamps, TV displays, traffic signals, camera flashes, advertising, medical devices, biological detection, water purification, etc. (see Fig. 2.16). LEDs are ideal for high-speed data transmission (as in infrared remote controls and optical fibre communications) or information display in small battery-powered devices. LEDs have almost replaced the incandescent bulbs for lighting due to their low power consumption. Infrared LEDs have applications in burglar alarms, night vision devices, etc.

You should now revise the ideas presented in this section by attempting SAQ 3.

SAQ 3 – LED

Give short answers to the following questions:

a) Explain the process of emission of light by an LED.

b) List four semiconducting materials used for constructing LEDs.

c) Which region of the *I*-*V* characteristics do LEDs operate in? Why?

d) On what semiconductor parameter does the colour of the light emitted by an LED depend upon? Explain.

e) What band gaps will produce green and violet lights having wavelengths 550 nm and 380 nm, respectively, in an LED?

We now introduce another very useful junction diode, the solar cell.

2.3.3 SOLAR CELL

The solar cell finds applications in solar energy generation. Solar cells are arranged in large groups of thousands, called solar arrays, and used for electricity generation. The arrays convert sunlight into electrical energy, which is used everywhere in many cities to power homes, industries, markets and other public places, transportation, etc. (see Fig. 2.17). In fact, entire cities in the developed world run on solar energy produced by solar arrays.

So, **the solar cell is a very useful junction diode, which converts sunlight into electrical energy**. It is manufactured in such a way that the voltage developed across it can be used to provide electric supply. Or we can draw current from it. Let us explain its construction, working and *I-V* characteristics.

Fig. 2.17: **Solar arrays comprising thousands of solar cells.**

Solar cells are *p-n* junction diodes constructed so that the side of the diode exposed to the sunlight is very thin. So, if the *n*-side of the diode is exposed to the Sun, the *n*-type layer is very thin on top of a thick *p*-layer (you will understand why when you learn about the working of a solar cell). Suitable metallic contacts in the form of thin wires for current flow or voltage tapping are connected to the cell. The solar cell has an antireflective coating to reduce reflection and enable maximum absorption of sunlight (see Fig. 2.18). The solar cell is covered with a transparent film or toughened glass to allow light to fall on it and to protect it from dust, rain, snow, etc. You can see an animation of the process at:
https://en.wikipedia.org/wiki/Solar_cell#/media/File:Solartce3.gif

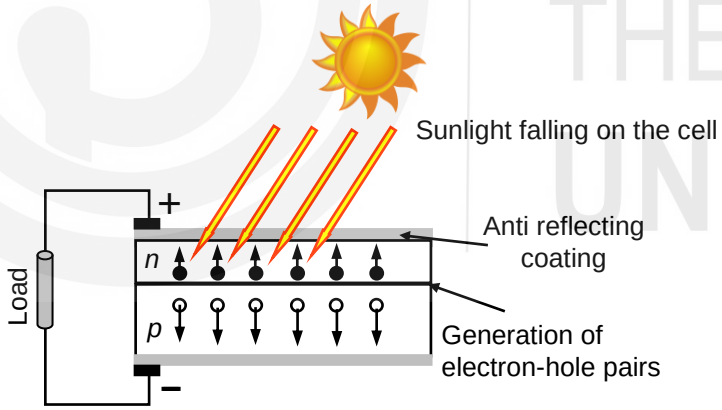
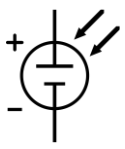


Fig. 2.18: Construction of a solar cell.

Initially, solar cells were made using silicon but these cells had a low efficiency of 15%. Since the average energy of photons in the sunlight is about 1.5 eV, GaAs is also used to manufacture solar cells because of its band gap, which is 1.42 eV. Later advances in materials science have led to solar cells made up of organometallics having increased efficiency. However, silicon is still used in large solar panels because it is inexpensive and rugged. Figs. 2.19a and b show a close-up of a solar cell in a solar array and its symbol.

(a)



(b)

Fig. 2.19: a) A single solar cell; b) circuit symbol of solar cell.

Let us now describe the **working** of a solar cell. In a solar cell, photons of energy equal to or more than the band gap of the semiconductor material are made to fall on the junction region of the diode to produce free electrons and electron-hole pairs. You may ask: **Why should light fall especially in the**

junction region? The answer is that if light were to fall in the *n*-region or *p*-region, it would simply create electron-hole pairs that would recombine immediately. But we would not like this to happen because then we would not be able to create a significant potential difference across the diode.

What we wish to do is increase the barrier potential to a significant level to use the diode as a voltage source. You will understand this point when you understand what happens when photons fall in the junction region. Refer to Fig. 2.18. When sunlight falls on the junction region through the thin *n*-layer, and the energy of the photons is greater than the band gap, electrons are freed from the covalent bonds; electron-hole pairs are created in the region. These electron-hole pairs do not recombine immediately. Why?

It happens due to the barrier potential or the electric field across the junction. Its polarity is such that electrons in the junction region move to the *n*-side and holes move to the *p*-side of the diode (see Fig. 2.20a). So, the number of electrons in the *n*-region increases making it more negative. Similarly, the number of holes in the *p*-region increases making it more positive. What is the result? A voltage is generated across the diode. This is known as **photovoltaic voltage** because we are using photons to generate voltage: [photo (from photons) + voltaic from voltage)]. **The process in which light is used to generate voltage is called the photovoltaic effect**.

But can this movement of electrons to the *n*-region and holes to the *p*-region continue forever? The answer is no, it cannot. This is because electrons that accumulate in large numbers in the *n*-region will repel other electrons moving to the *n*-side, and so will holes repel other holes entering the *p*-side. These will move back towards the depletion region and recombine. So, the photovoltaic voltage does not increase infinitely but is limited to a maximum value.

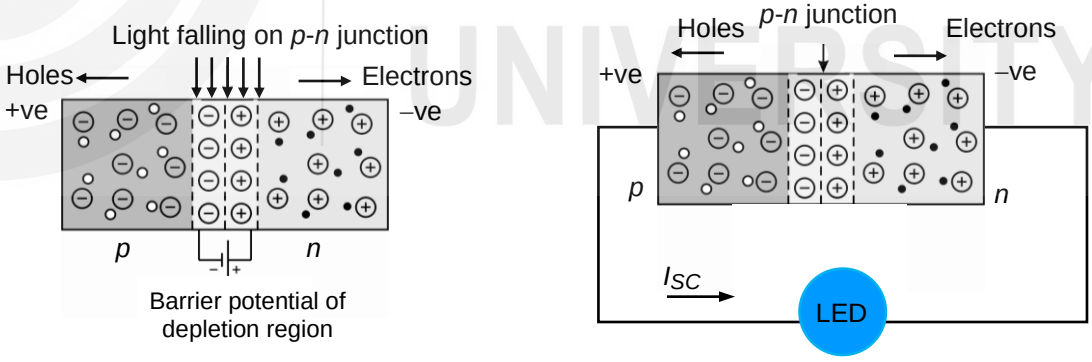


Fig. 2.20: **a) Working of a solar cell; b) flow of current.**

Now, what will happen if we attach two metallic contacts to the *n*- and *p*-ends of the diode and connect them with a wire to some device, say, an LED? Refer to Fig. 2.20b. Electrons will flow towards the positive side of the potential barrier, setting up a current in the circuit and recombine with holes in the *p*-region. But since sunlight is falling continuously on the junction, electron-hole pairs will be continuously created and electrons will keep flowing. This results in a continuous, constant current flow (I_{SC}) in the circuit. The LED will glow continuously! Notice that **there is no external battery** and yet current is flowing in the circuit. How? This is because **the solar cell is itself a battery**!

To sum up, **photons falling in the depletion region create electron-hole pairs**. Before these charge carriers recombine, electrons move to the *n*-region and holes to the *p*-region due to the strong electric field at the junction. Thus, **electrons accumulate in the *n*-region and holes in the *p*-region, resulting in the photovoltaic voltage across the diode**. When a device is connected in a circuit across the diode, current flows in it.

Let us now understand the *I*-*V* characteristics of a solar cell shown in Fig. 2.21. Initially the external voltage is zero, and so the circuit is an **open circuit**. If we do not connect any wires, the circuit is open, the current is zero. But you see that the *p*-end of the diode is positive. This is a **forward bias**. So, **when current is zero, we have a positive maximum photovoltaic voltage across the diode**. This is denoted by V_{oc}. **It is the maximum open circuit voltage when the current through the solar cell is zero**.

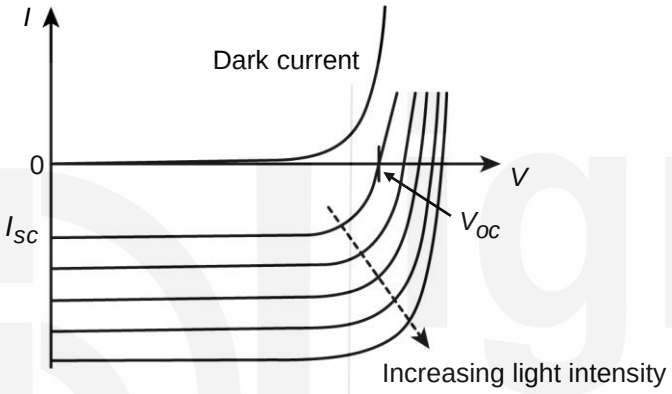


Fig. 2.21: *I*-*V* **characteristics of a solar cell.**

Now, we short-circuit the solar cell, that is, just connect a wire across its ends by removing the LED in Fig. 2.20b. We see that a constant current flows in the circuit. It is the **maximum** current in the circuit due to the solar cell's photovoltaic voltage.

This is called the **maximum short-circuit current** and we denote it by I_{sc}. It depends on the intensity of the incident sunlight. The more the numbers of photons, higher is the maximum short-circuit current in the solar cell. This is shown by other curves in Fig. 2.21.

Note that the maximum short-circuit current is flowing when **the externally applied voltage is zero**. Why is it negative in the *I*-*V* characteristic curve?

Since electrons flow towards the *n*-end inside the solar cell, in the external circuit, electrons flow from the *n*-end to the *p*-end to complete the circuit. This is opposite to the case of the conventional *p*-*n* junction diode shown in Fig. 2.6, where the electrons flow from the negative end of the battery towards the *n*-end of the diode, and emerge out of the diode from its *p*-end. Hence, the current flowing in the solar cell circuit is opposite (negative) as compared to the forward biased *p*-*n* junction diode. Therefore, I_{sc} is a negative current.

The curve marked **dark current** in Fig. 2.21 shows the current in the solar cell under *no-light* (dark) conditions. It reflects the behaviour of a solar cell when

no light falls on it, which is like a normal *p-n* junction diode. You may like to end this section with an SAQ to fix these ideas in your mind.

SAQ 4 – Solar cell

a) What is a solar cell?

b) Explain the working principle of a solar cell, in brief.

c) State one difference in the materials used for making solar cell and LED.

2.3.4 Photodiode

A photodiode is the ordinary *p-n* junction diode that generates electric current when light falls on it. Semiconducting materials such as Si, Ge, InGaAs, PbS are used for making different types of photodiodes. A photodiode is constructed to optimize its sensitivity to light incident on it for greater current flow. It is also known as **photodetector** or **photosensor**. **It operates in the reverse biased mode** and **converts light energy into electrical energy**. The underlying principle of a photodiode is the **photoelectric effect** about which you have studied in Unit 4 of the course BPHET-141. When light falls on a reverse biased photodiode, electric current due to minority charge carriers is generated in it. The photodiode is constructed as shown in Fig. 2.22.

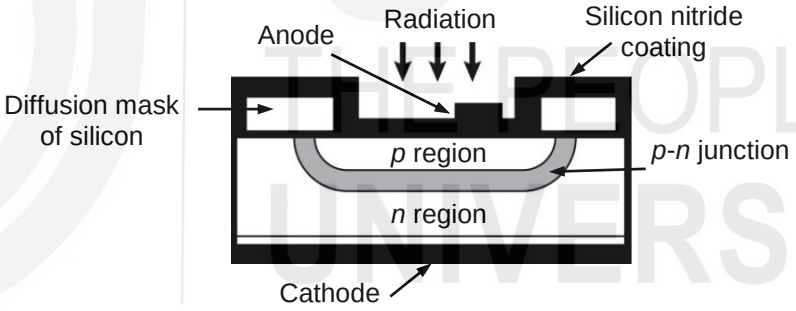


Fig. 2.22: **Construction of a photodiode.**

Actually, all *p-n* junction diodes can function as photodiodes when light falls on them. You may ask: How? Let us explain the mechanism.

Recall from the discussion in Sec. 2.2 that the reverse current in a *p-n* junction diode is largely made up of minority charge carriers. As you know these carriers are generated at room temperatures or higher temperatures. At room temperatures, valence electrons acquire enough thermal energy to break free from covalent bonds in the semiconductor and electron-hole pairs are produced. Although the lifetime of minority charge carriers is small, these contribute to the reverse current while they exist.

Light energy has the same effect on generation of minority charge carriers as thermal energy. When a beam of light of energy greater than the band gap of the semiconductor is incident on a *p-n* junction diode, it can break covalent bonds, and produce electron-hole pairs. The more is the light that falls on the diode, the larger is the reverse current. So, the reverse current in the diode

depends on the number of photons of appropriate energy falling on it. The reverse current is called the **photocurrent**.

In many *p*-*n* junction diodes, this is an unwanted effect as these will not function correctly if they are illuminated by unwanted light of appropriate energy that can produce reverse current. This is prevented by encasing them in opaque packaging so that no electromagnetic radiation reaches the junction. Exactly the opposite is done in a photodiode. Photodiodes are constructed with a window or optical fibre connection to allow light to reach the sensitive region in the device. In actual construction, photodiodes may contain optical filters, built-in lenses, and may have large or small surface areas.

Light falling on the photodiode produces free electrons and holes that **increase the reverse current through it**. The larger the number of photons, i.e., greater the intensity of light, greater is the number of minority charge carriers produced, and larger is the reverse current. Typically, the reverse current in photodiodes is of the order of tens of microamperes.

As you may have guessed by now, **photodiodes operate in the reverse biased region** of the *I*-*V* characteristic curve, which is the same as Fig. 2.11 with a larger reverse current (see Fig. 2.23). Recall that in a reverse biased diode, the *p*-side is connected to the negative terminal and the *n*-side to the positive terminal of the power source or the battery. Remember that **in most applications of a photodiode, the reverse voltage is nowhere near the breakdown voltage**.

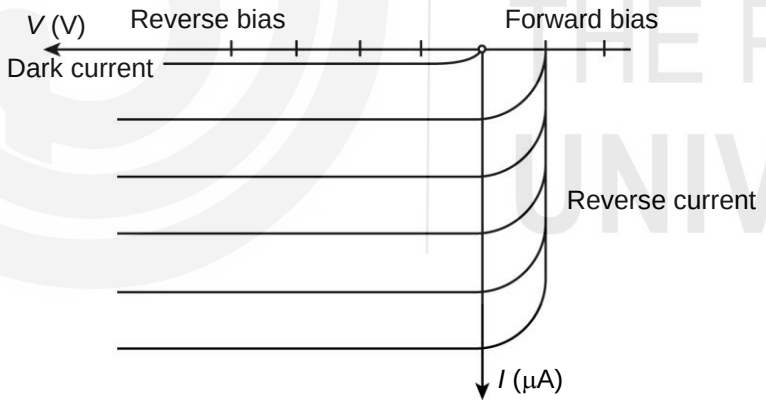
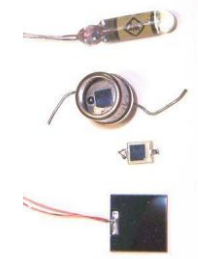


Fig. 2.23: *I*-*V* **characteristics of a photodiode.**

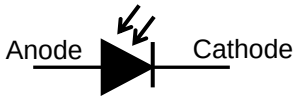
In Fig. 2.23, notice that the topmost curve is labelled as the dark current. This is the small reverse current that flows when no light is incident on the photodiode. Also, note that the reverse current increases as the intensity of the incident light is increased.

To sum up, a **photodiode** is a *p*-*n* junction diode that converts light into electrical current, which is generated when photons of appropriate energy are incident on the *p*-*n* junction producing minority charge carriers. A photodiode is also called a photo-detector, light detector and photo-sensor. Fig. 2.24 shows an actual photodiode and its symbol.

From this description, you can visualise many applications of photodiodes, wherein we need to convert light into an electrical current. Can you now see



(a)



(b)

Fig. 2.24: a) A few actual photodiodes; b) circuit symbol of a photodiode.

that **the solar cell is a large area photodiode**? You may now like to revise what you have learnt in this section. Try SAQ 5.

SAQ 5 – Photodiode

a) How is a photodiode different from an ordinary *p-n* junction diode?

b) Calculate the maximum wavelength of photons that can be detected by a photodiode made by a semiconductor of band gap 3.0 eV.

c) What is dark current in a photodiode?

d) Does the photodiode operate in the breakdown region?

We end this section by listing a few applications of photodiodes. Photodiodes are used in many electronic products around us such as compact disc players, smoke detectors, infrared remote control devices used to control televisions, air conditioners, camera light meters, medical devices, etc. One of the commonest uses of photodiodes is in switching on street lights after dark, or room lights when anyone enters it.

Photodiodes are also used as photodetectors in charge-coupled devices (used to collect light from the stars in huge telescopes), photoconductors, and photomultiplier tubes.

Let us now summarise the contents of this unit.

2.4 SUMMARY

Concept	Description
p-n ***junction and*** *p-n* ***junction diode, barrier potential and depletion region***	■ A *p-n* junction is formed from a single intrinsic semiconducting crystal when controlled amounts of donor and acceptor impurities are added to it. It has the following features: • The boundary or interface between the *n*-type and *p*-type parts of the crystal is called the *p-n* **junction**. A single *p-n* junction is a diode, which is **a two-terminal device formed by such doping**. • Due to the concentration gradient in the diode, **diffusion** of majority charge carriers takes place across the *p-n* junction, which leads to the accumulation of positive and negative ions near the junction. As a result, an electric field is created near the junction **preventing further movement of majority charge carriers**. • The electrostatic potential associated with this electric field is known as the **barrier potential**. The barrier potential makes the *p*-side negative with respect to the *n*-side in the diode. • Since movement of majority charge carriers across the *p-n* junction is prevented by the barrier potential, it results in the formation of a region near the junction where **no mobile charge carriers are present**. This is known as the **depletion region** or **space-charge region.**

I-V characteristics of a p-n junction diode

■ The *I-V* characteristics of a *p-n* junction diode have the following features:

- When a *p-n* junction diode is **forward biased**, that is, its *p*-side is positive with respect to the *n*-side by the externally applied voltage, a forward current typically of the order of a few milliamperes flows in the diode for applied voltages of less than 1 V beyond the **knee voltage.** The **knee voltage** is defined as the forward voltage at which the flow of the current through the *p-n* junction of the diode increases rapidly. The diode **offers low forward resistance to the flow of current in forward bias**.

- When the *p*-side of the *p-n* junction diode is negative with respect to its *n*-side, it is **reverse biased**. A very small **leakage current** or **reverse saturation current** due to drift of minority charge carriers, of the order of nanoamperes to microamperes for large values of the reverse voltage flows in a reverse biased *p-n* junction diode. In reverse bias, the diode **offers high resistance to the flow of current**. At **breakdown voltage**, which is the maximum reverse bias voltage that can be applied to a *p-n* junction diode, the current increases rapidly.

- Since a *p-n* junction diode allows current to flow only in one direction, it can function as a **rectifier**, which is a device that converts ac into dc.

Zener diode

■ **Zener diodes** are constructed like conventional *p-n* junction diodes made of silicon but are **very heavily doped**, so that their **depletion region is thin** giving rise to high electric fields across the junction. This results in the following features of the *I-V* characteristics of a zener diode:

- In the forward region, the zener diode conducts current like a conventional *p-n* junction diode. When the diode is reverse biased, high electric field in the depletion region results in breakdown. A **zener diode operates in the breakdown region** of its *I-V* characteristics.

- When a zener diode is reverse biased and the source voltage is greater than the breakdown voltage of the diode, its output voltage remains constant for a large variation in current making it useful for **voltage regulation**.

- Breakdown occurs due to two mechanisms in the zener diode: **Zener breakdown**, and **avalanche breakdown**.

- **Zener breakdown** occurs when the diode is reverse biased at typical voltages of 2 V to less than 6 V. This happens because the reverse bias increases further the high electric field in the thin depletion region, which creates a large number of electron-hole pairs resulting in rapid increase of current (breakdown).

- **Avalanche breakdown** occurs in less heavily doped *p-n* junction diodes having wider depletion layers and, therefore, higher breakdown voltages (greater than 6 V). It happens due to **impact ionisation** caused by collisions of electrons freed by the high electric field, with other atoms.

LED (light emitting diode)

- An **LED** or **light emitting diode** is a *p*-*n* junction **diode that emits light when it is forward biased**. It has the following features:
  - LEDs are constructed using **heavily doped** *p*-*n* junction diodes, due to which their depletion region is very thin. Energy is emitted in *p*-*n* junction diodes in the form of **light** (rather than heat) **when electron-hole pairs recombine in the thin junction region**.
  - **LEDs emit light** because these are made of **special type of compound semiconductors** with direct band gaps in the range 2eV to 3eV. The wavelength of radiation emitted by an LED depends on the band gap of the semiconducting materials used for constructing it.
  - LEDs can emit light of various colours in the visible region, and radiation in near infrared and near ultraviolet regions.
  - The *I*-*V* characteristics of an LED are the same as those of an ordinary *p*-*n* junction diode. **LEDs are always operated in the forward bias, never in reverse bias**. **LEDs are damaged if operated in reverse bias**.
  - LEDs are used in a variety of applications such as general lighting, displays in devices, medical devices, biological detection, water purification, high-speed data transmission, burglar alarms, night vision devices, etc.

Solar cell

- **The solar cell is a junction diode, which converts sunlight into electrical energy**. It is manufactured in such a way that the voltage developed across it can be used to provide electric supply. Or we can draw current from it. It has the following features:
  - Solar cell is constructed so that the side of the diode exposed to the sunlight is very thin. So, if the *n*-side of the diode is exposed to the Sun, the *n*-type layer is very thin on top of a thick *p*-layer. The solar cell has an antireflective coating to reduce reflection and enable maximum absorption of sunlight. It is covered with a transparent film or toughened glass to allow light to fall on it.
  - In a solar cell, photons of energy equal to or more than the band gap of the semiconductor material are made to fall specifically on the junction region of the diode. Photons of appropriate energy produce free electrons and electron-hole pairs in the depletion region. The electron-hole pairs do not recombine immediately because the barrier potential drives electrons in the junction region to the *n*-side and holes to the *p*-side of the diode making the *n*-region more negative and the *p*-region more positive. Thus, a voltage is generated across the diode, which is called the **photovoltaic voltage**. **The process in which light is used to generate voltage is called the photovoltaic effect**. A load connected to the solar cell allows flow of current in the circuit.
  - When the external voltage is zero, and no wires are connected across the solar cell terminals, the circuit is an **open circuit**. There is **a positive maximum open circuit photovoltaic voltage** (V_{OC}) **across the diode when** the current in it is zero. This is a **forward bias**

voltage. When the solar cell is short-circuited, and the externally applied voltage is zero, a constant **maximum short-circuit current** (I_{sc}) flows in the circuit, which increases with the intensity of the incident sunlight.

- The **dark current** in a solar cell is the current in it under *no-light* (dark) conditions. The *I-V* characteristic curve of the solar cell showing dark current reflects the behaviour of a solar cell when no light falls on it, which is like a normal *p-n* junction diode.

Photodiode

- A **photodiode** is the ordinary *p-n* **junction diode that generates electric current when light falls on it**. **It operates in the reverse biased mode** and **converts light energy into electrical energy**. It is also known as **photo-detector, light detector** and **photo-sensor**. It has the following features:
  - Photodiodes are constructed with a window or optical fibre connection to allow light to reach the sensitive region in the device. In actual construction, photodiodes may contain optical filters, built-in lenses, and may have large or small surface areas.
  - A beam of light of appropriate energy incident on a reverse biased photodiode produces electron-hole pairs and increases the reverse current. This reverse current is called the **photocurrent** and it increases with the intensity of light. Typically, the reverse current in photodiodes is of the order of tens of microamperes.
  - In most applications of a photodiode, **the reverse voltage is much less than the breakdown voltage**.
  - Photodiodes are used as light sensors in many electronic products, in charge-coupled devices, as photoconductors, and photomultiplier tubes.

2.5 TERMINAL QUESTIONS

1. Explain the rectifier action of a *p-n* junction diode on the basis of its *I-V* characteristics.
2. Fill in the blanks in the following Table:

Junction diode	Usually constructed of which materials	Bias useful for intended applications
p-n junction diode		
Zener diode		
LED		
Photodiode		

3. a) Using appropriate words or phrases, fill in the blanks in the following statements about zener breakdown and avalanche breakdown:

Zener breakdown occurs at .....V, whereas avalanche breakdown occurs at .....V. Zener breakdown occurs in .....doped *p-n* junction diodes having ..... depletion regions, whereas avalanche breakdown occurs in ..... doped *p-n* junction diodes having ..... depletion regions. Zener breakdown occurs due to.....Avalanche breakdown occurs due to.....

b) Explain briefly how the *I-V* characteristics of a zener diode suggest its use for voltage regulation.

4. a) Which materials are used for constructing LEDs, and why?

   b) Draw the *I-V* characteristics of an LED.

5. Give short answers to the following questions:

   a) Why is the top layer in a solar cell thin?

   b) How does a solar cell function like a battery?

   c) Explain the terms open circuit voltage and short circuit current for a solar cell.

   d) Why does the short circuit current increase in a solar cell with the intensity of light?

6. Explain the underlying principle of a photodiode. What features in its construction are responsible for its function as a photodiode?

7. Write the name of the junction diode against each of the applications listed below:

   a) TV display.....

   b) Converting ac into dc…….

   c) Providing energy for lighting, heating, etc…….

   d) Detecting fire with smoke……

   e) Protecting appliances by keeping voltages constant for fluctuations in power supply…….

   f) Collecting light in huge telescopes…….

   g) Water purification……..

   h) Providing energy for transportation…….

2.6 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. a) The depletion region width i) decreases when the forward biasing voltage across a *p-n* junction diode is increased and ii) increases when the reverse biasing voltage across it is increased.

   b) The knee voltage of a Ge *p-n* junction diode is 0.3 V. So, current will flow when the external bias is 0.5 V.

   c) A diode made of GaAs will start conducting when its forward biasing voltage exceeds the knee voltage 1.2 V.

2. The following features distinguish a zener diode from a conventional *p-n* junction diode:

   - Their doping. The zener diode has heavily doped *p* and *n* regions compared to a conventional *p-n* junction diode, which results in a thin depletion region compared to the normal diode.
   - The conventional *p-n* junction diode allows current to flow only in one direction while zener diode allows current to flow in both directions (in the reverse direction beyond the breakdown voltage).
   - The conventional *p-n* junction diode operates at far lower reverse biasing voltages than the breakdown voltage but the zener diode is operated at breakdown voltages.

   The main difference between zener breakdown and avalanche breakdown is as follows:

   - Zener breakdown is caused at lower voltages (2 V to less than 6 V) due to high electric fields caused by the thin depletion region in the diode. The electric field frees electrons from the covalent bonds and thus creates electron-hole pairs leading to large currents and breakdown.
   - Avalanche breakdown occurs at much higher voltages (> 6 V) due to impact ionization. Since the depletion region is wider due to less heavy doping, electrons freed by the high electric field collide continually with atoms in the semiconducting material and generate a large number of electron-hole pairs leading to breakdown.

3. a) When electron-hole pairs in the **very thin depletion region** of heavily doped forward biased *p-n* junction diode (LED) **recombine**, energy is emitted in the form of light. This happens because semiconducting materials like gallium arsenide phosphide, gallium phosphide, etc. used to construct LEDs have direct band gaps suitable for light emission. Thus, LEDs that radiate red, green, yellow, orange, violet light, near infrared and near ultraviolet radiations can be produced.

   b) Four semiconducting materials used for constructing LEDs are: Gallium Phosphide, Gallium Arsenide Phosphide, Gallium Indium Phosphide and Indium Gallium Nitride. Of course, there are many more materials that are used for producing LEDs.

   c) LEDs operate in the forward biased region of the *I-V* characteristics. This is because when the LED is forward biased, electrons and holes are injected into the region across the junction and recombine to emit light. In the energy band model, we say that spontaneous emission of light occurs due to recombination of minority charge carriers across the band gap.

   d) The colour of the light emitted by an LED depends upon the band gap of the semiconducting material used to manufacture it. If the band gap equals the energy corresponding to the specific wavelength of the colour of light, that colour is emitted.

   e) The wavelength of green and violet colours in light are 550 nm and 380 nm, respectively. Using the expression $E = hc/\lambda$ and substituting

the values of h, c (from the Table of Physical Constants given at the end of this block) and λ, we get the band gap for green light as:

$$E = \frac{6.63\times10^{-34}\times3.0\times10^{8}}{550\times10^{-9}}\,\text{J} = 3.6\times10^{-19}\,\text{J} = \frac{3.6\times10^{-19}}{1.6\times10^{-19}} = 2.25\,\text{eV}$$

Band gap for violet light is:

$$E = \frac{6.63\times10^{-34}\times3.0\times10^{8}}{380\times10^{-9}}\,\text{J} = 5.2\times10^{-19}\,\text{J} = \frac{5.2\times10^{-19}}{1.6\times10^{-19}} = 3.25\,\text{eV}$$

4. a) The solar cell is a *p*-*n* junction diode which converts sunlight into electrical energy. It is constructed so that the side of the diode exposed to the sunlight is very thin. So, if the *n*-side of the diode is exposed to the Sun, the *n*-type layer is very thin on top of a thick *p*-layer. The sunlight is made to fall on the junction region, which leads to the solar cell functioning as a battery.

   b) When sunlight falls on the junction region through the thin *n*-layer, and the energy of the photons is greater than the band gap, electron-hole pairs are generated in the region. Due to the polarity of the barrier potential across the junction, these electrons and holes do not recombine in the junction region. Rather electrons move to the *n*-side and holes move to the *p*-side of the diode thereby generating a potential difference across the diode.

   c) The main difference in the materials used for making a solar cell and an LED is that the LEDs are made of special semiconducting materials with direct band gaps of 2 V to 3 V corresponding to energies in the visible, near infrared and near ultraviolet regions. Solar cells are largely made up of silicon, which has a band gap of 1.12 eV.

5. a) A photodiode is different from an ordinary *p*-*n* junction diode in that it generates electric current when light falls on it. An ordinary *p*-*n* junction diode generates current at thermal energies.

   b) The maximum wavelength of photons that can be detected by a photodiode made by a semiconductor of band gap 3.0 eV is calculated from $E = hc/\lambda$ with $E = 3.0\,\text{eV}$. Substituting the values of h, c and E, we get:

$$\lambda = \frac{hc}{E} = \frac{6.63\times10^{-34}\times3.0\times10^{8}}{3.0\times1.6\times10^{-19}}\,\text{m} = 4.1\times10^{-7}\,\text{m} = 410\,\text{nm}$$

   Note that we have converted eV into joules for this calculation.

   c) Dark current in a photodiode is the small reverse current that flows when no light is incident on the photodiode.

   d) The photodiode does not operate in the breakdown region.

Terminal Questions

1. Refer to Fig. 2.11. It shows clearly that appreciable current flows in the *p*-*n* junction diode when it is forward biased. When it is reverse biased, the current through it is very small. So, a *p*-*n* junction diode allows current to

flow only in one direction and can be used as a rectifier that converts ac into dc.

2. Refer to the following Table:

Junction diode	Usually constructed of which materials	Bias useful for intended applications
p-n junction diode	Si, Ge, GaAs	Both forward and reverse bias but at reverse voltages less than breakdown voltages
Zener diode	Si	Reverse bias at breakdown voltages
LED	Gallium Phosphide, Gallium Arsenide Phosphide, Gallium Indium Phosphide, Indium Gallium Nitride and many more materials with appropriate direct band gaps	Only in the forward bias
Photodiode	Si, Ge, Indium Gallium Arsenide, Lead Sulphide	Reverse bias

3. a) Zener breakdown occurs at 2V to less than 6V, whereas avalanche breakdown occurs at more than 6V. Zener breakdown occurs in heavily doped *p-n* junction diodes having very thin depletion regions, whereas avalanche breakdown occurs in less heavily doped *p-n* junction diodes having wider depletion regions. Zener breakdown occurs due to electron-hole generation caused by high electric fields whereas Avalanche breakdown occurs due to due to electron-hole generation caused by impact ionization in high electric fields.

   b) At the breakdown voltage in the reverse biased region of the *I-V* characteristics of a zener diode, voltage remains constant for a large variation in current. So, even if the current varies by large amounts, the voltage across the load connected in parallel with the zener diode remains constant. Therefore, it can be used for voltage regulation.

4. a) Some examples of semiconducting materials used for constructing LEDs are: Gallium Phosphide, Gallium Arsenide Phosphide, Gallium Indium Phosphide and Indium Gallium Nitride because these materials have direct band gaps suitable for emission of radiation in the visible, near infrared and near ultraviolet regions which lend them to specific applications of LEDs.

   b) See Fig. 2.11 in the forward biased region.

5. a) The top layer in a solar cell is thin so that sunlight falling on it reaches the junction region without creating electron-hole pairs in the *n*- or *p*-region constituting the top layer, so that electron-hole pairs are created only in the junction region.

   b) A solar cell functions like a battery due to the following process:

   When sunlight falls on the junction region of a solar cell through its thin top layer, and the energy of the photons is greater than the band gap, electron-hole pairs are created in the region. These electron-hole pairs

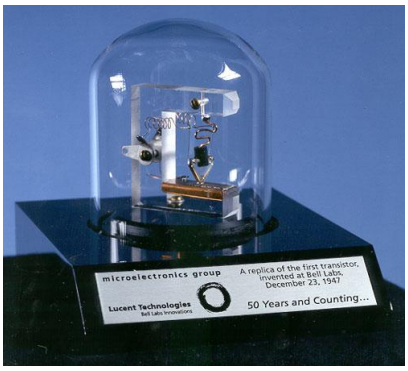
do not recombine immediately because of the polarity of the barrier potential, which makes electrons in the junction region to move to the *n*-side and holes to the *p*-side of the diode. So, the number of electrons in the *n*-region increases, making the *n*-end of the solar cell more negative and the number of holes in the *p*-region increases making it more positive at its *p*-end. Thus, a voltage is generated across the diode. The solar cell functions as a voltage source or a battery from which current can be drawn across a load.

c) Open circuit voltage in a solar cell is the voltage across the solar cell when no wires connect its two terminals, that is, when the current through the solar cell is zero. Short circuit current in a solar cell is the current due to its photovoltaic voltage when the cell is short circuited by joining wires to its terminals.

d) The short circuit current increases in a solar cell with the intensity of light because the number of photons falling on the junction increases with the intensity of light. Hence, the number of electron-hole pairs created increases resulting in an increase in the current.

6. The underlying principle of a photodiode is based on the photoelectric effect. The photoelectric effect is that when light of appropriate energy falls on a metal, it leads to emission of electrons from metals, which constitute a current. Similarly, when light of appropriate energy (greater than or equal to the band gaps) falls on a reverse biased photodiode, it generates electrons and holes, which are available for conduction. And so, current flows in the photodiode. Photodiodes are constructed like conventional *p*-*n* junction diodes. But whereas conventional *p*-*n* junction diodes are packaged in opaque casings to prevent light falling on them, photodiodes are constructed with a window or optical fibre connection to allow light to reach the sensitive region in the device. A photodiode may also contain optical filters, built-in lenses to facilitate its functioning.

7. a) TV display: LED
   b) Converting ac into dc: conventional *p*-*n* junction diode
   c) Providing energy for lighting, heating, etc.: Solar cells in solar panels
   d) Detecting fire with smoke: Photodiode
   e) Protecting appliances by keeping voltages constant for fluctuations in power supply: Zener diode
   f) Collecting light in huge telescopes: Photodiode
   g) Water purification: LEDs
   h) Providing energy for transportation: Solar cells in solar panels



The picture above shows the replica of the first transistor created by Lucent Technologies in 1997, to commemorate the 50th anniversary of the invention of the transistor. Transistors are used in a wide range of applications. You will learn about their basic physics in this unit.

Source of the picture: https://en.wikipedia.org/ https://clintonwhitehouse4.archives.gov/Initiatives/Millennium/capsule/mayo.html

UNIT 3

TRANSISTORS

Structure

STUDY GUIDE

In this unit, you will learn about transistors, which are essentially devices comprising two *p-n* junctions and are called **double junction devices**. You have studied about the bipolar junction transistor in your senior secondary school physics courses in detail. So, you are familiar with the construction, working and biasing conditions of the bipolar junction transistor. In this unit, you will also learn about the field effect transistor. You have to focus on how these devices are constructed, and the physics underlying their working. Then you will learn the contents of the next unit better, which describes various *I-V* characteristic curves of the BJT that lead to its applications as an amplifier and a switch. You should revise Sec. 2.2 of Unit 2 related to the biasing and working of the *p-n* junction diode to learn these concepts well. Note down your difficulties if you do not understand any part of the explanation, and ask your Counsellor to clarify those. You should try to answer all SAQs and Terminal Questions on your own for better learning of the concepts of this unit.

"Somewhere, something incredible is waiting to be known."	***Carl Sagan***

3.1 INTRODUCTION

In Unit 2, you have learnt about semiconductor devices called **junction diodes** such as the *p-n* junction diode, the zener diode, LED, solar cell and the photodiode. You have learnt that by adjusting the doping and the construction of the *p-n* junctions, these diodes could be used in a wide variety of applications. In this unit, you will learn how the use of two *p-n* junctions in a single device alters its working, and characteristics that makes it even more useful. Such devices are called **double junction devices**.

In this unit, you will learn about two such devices: **bipolar junction transistor** named so because both electrons and holes carry current in it (*bi* means two); and the **field effect transistor**, which is *unipolar* because current flow in it is due only to one type of charge, electrons or holes (*uni* means one).

Both types of transistors have numerous applications in modern electronic devices and systems as amplifiers and switches. Devices and systems, which involve amplifying signals, such as radio, television, audio systems, satellites and communication systems, space vehicles, power systems, signal generators, etc. use transistors as the basic building blocks. In the form of switching devices, these can be used for many applications such as in logic gates, microprocessors in automatic washing machines, calculators or computers, microcontrollers used for process control in industries, and even to regulate vehicular traffic on the roads.

Transistors are an integral part of the integrated chips (ICs) and mother boards used in various appliances and gadgets around us such as electronic lighters, toys, microwave ovens, modern refrigerators, smart watches, computers, mobile phones, etc. In fact, their use is consistently increasing. You will learn about these applications as you study this unit and the next blocks.

Transistors have brought about a revolution, which has helped upgrade technology and improve efficiency. It is, therefore, important that you understand the basic physics underlying the construction and working of these devices so that you understand how they work as amplifiers and switches. In Secs. 3.2 and 3.3, you will learn about the construction, biasing and working of bipolar junction transistors and junction field effect transistors, respectively. This knowledge will help you understand the *I-V* characteristics of BJTs under different biasing conditions, which we discuss in the next unit.

Expected Learning Outcomes

After studying this unit, you should be able to:

- describe the construction of bipolar junction transistors and junction field effect transistors;
- explain the biasing and working of bipolar junction transistors; and
- explain the biasing and working of junction field effect transistors.

3.2 BIPOLAR JUNCTION TRANSISTOR

The first bipolar junction transistor (BJT) was invented in 1948 by William Shockley, American physicist and inventor (read the margin remark). In this section, we discuss the construction, working and biasing conditions of a bipolar junction transistor.

From left: John Bardeen, William Shockley and Walter Brattain who worked in the Bell Laboratories in New Jersey, America, which is the research arm of the company American Telephone and Telegraph (AT&T). Shockley worked for ten years on the physics of the transistor but it was John Bardeen and William Brattain who created the “point contact transistor” under his supervision. However, the "bipolar" transistor, which was superior to the point contact transistor, was designed by Shockley. So, we can say that the bipolar junction transistor was largely Shockley's creation. The bipolar junction transistor was successfully demonstrated at the Bell Labs and its discovery was announced in a press release on July 4, 1951. Shockley, Bardeen and Brattain were jointly awarded the Nobel Prize in Physics in 1956.

3.2.1 Construction

Let us begin the discussion by asking: **What is a bipolar junction transistor (BJT)**? How is it constructed?

A bipolar junction transistor is a three-terminal device as shown in Fig. 3.1a. It is constructed by doping a single crystalline semiconducting material (usually silicon or germanium) in two ways:

i) by sandwiching a thin layer of *p*-type semiconductor between two outer layers of *n*-type semiconductor forming an *n*-*p*-*n* transistor; or

ii) by sandwiching a thin layer of *n*-type semiconductor between two outer layers of *p*-type semiconductor forming a *p*-*n*-*p* transistor.

This kind of doping makes it possible to use a BJT as an amplifier or a switch (see Fig. 3.1b).

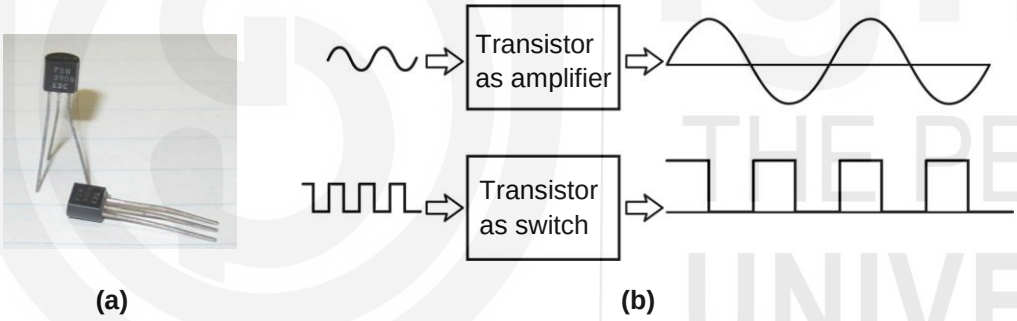


Fig. 3.1: **a) A bipolar junction transistor is a three-terminal device, which can be used b) as an amplifier and a switch.**

So, in a bipolar junction transistor, there are **three regions doped alternately** by *n*-type and *p*-type impurities to create an *n*-*p*-*n* or a *p*-*n*-*p* type of structure. Fig. 3.2 shows the structure of an *n*-*p*-*n* transistor.

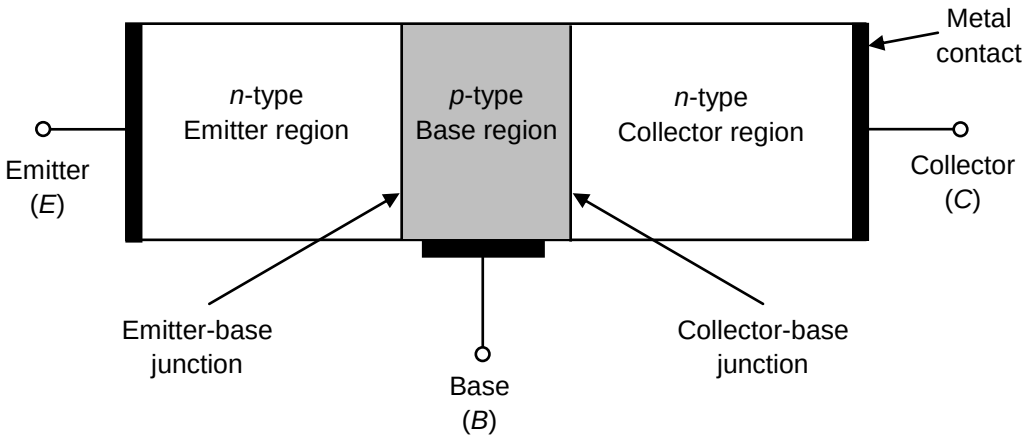


Fig. 3.2: **Structure of an *n*-*p*-*n* transistor.**

The three regions (*n*, *p* and *n*) shown in Fig. 3.2 are called **emitter** (*E*), **base** (*B*) and **collector** (*C*), respectively. So, the BJT is a **three-terminal device** with the electrical contact terminals attached to the emitter, base and collector. Sometimes, this is indicated by the letters E, B and C marked on the transistor.

The emitter, base and collector regions in a transistor are doped differently and have different widths. The emitter is doped heavily because its role is to 'emit' or inject charge carriers into the base region. The base region is lightly doped and is made very thin so that the heat loss due to recombination of charge carriers in it is reduced. The collector's doping levels lie between those of the emitter and the base. The collector is the largest of the three regions because it collects the charge carriers and has to dissipate more heat than the emitter or base. As you can see in Fig. 3.2, the transistor has two *p*-*n* junctions – one *p*-*n* junction is between the emitter and the base and the other *p*-*n* junction is between the base and the collector:

- **Emitter-base junction**, and
- **Collector-base junction**.

That is why a bipolar junction transistor can be thought of as having two *p*-*n* junctions back-to-back: one **emitter-base junction** and the other **collector-base junction**.

You may like to fix the structure of a BJT in your mind before you learn about its working. Try SAQ 1.

SAQ 1 – Bipolar junction transistor structure

a) Draw a labelled diagram showing the structure of a *p*-*n*-*p* transistor in the manner shown in Fig. 3.2 for an *n*-*p*-*n* transistor.

b) List the emitter, base, collector regions in a BJT in the order of (i) increasing levels of doping; (ii) increasing width.

3.2.2 Biasing and Working Mechanism

You can think of the *n*-*p*-*n* transistor as an *n*-*p* junction followed by a *p*-*n* junction. And the *p*-*n*-*p* transistor as a *p*-*n* junction followed by an *n*-*p* junction. So, we can understand its working in terms of the working of the two back-to-back *p*-*n* junctions.

Since there are two *p*-*n* junctions in a BJT, there will be **two depletion regions in the transistor** when no voltage is applied across the transistor due to diffusion of free electrons across the emitter-base and collector-base junctions. But **these will not be of the same width**. Can you say, why? This is because of the **different doping levels** of the emitter, base and collector.

Since the emitter is heavily doped, there will be a greater concentration of ions near the emitter-base junction preventing movement of charges across the junction. This will result in a thin depletion layer. Since the base region is

lightly doped, it will have a wider depletion region as the concentration of ions on that side of the junction is less. So, the depletion region will be of lesser width in the emitter region as compared to the base region. This means the depletion layer extends well into the base and only slightly into the emitter (see Fig. 3.3).

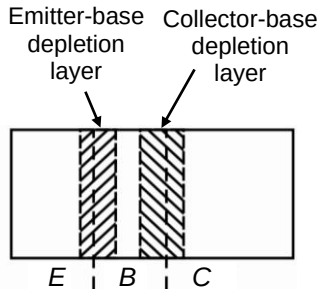


Fig. 3.3: **Depletion layers in a transistor when no voltage is applied.**

Similarly, at the collector-base junction, the depletion region in the base region is wider than that in the collector region. Since the collector is doped lesser as compared to the emitter, the collector-base depletion layer is wider than the emitter-base depletion layer (Fig. 3.3).

For each depletion layer, the barrier potential is about 0.7 V at 25°C for a silicon transistor and 0.3 V for a germanium transistor. Do you know that silicon transistors are more widely used than germanium transistors because of higher voltage rating, greater current ratings, and low temperature sensitivity? For our discussion, we will refer to silicon transistors, unless indicated otherwise.

In order that a transistor functions properly, we need to apply suitable voltages to its terminals. This is called the **biasing** of the transistor. So, the working of a bipolar junction transistor depends on the way the two *p*-*n* junctions in the transistor are biased. We take the case of an *n*-*p*-*n* transistor to explain transistor biasing and working mechanism. In our discussion, we consider an *n*-*p*-*n* transistor because it is more commonly used.

A typical biasing scheme of an *n*-*p*-*n* transistor is shown in Fig. 3.4. Note from Fig. 3.4 that in an *n*-*p*-*n* transistor, the **emitter-base** *p*-*n* **junction is forward biased while the collector-base** *p*-*n* **junction is reverse biased** for its normal operation as an amplifier.

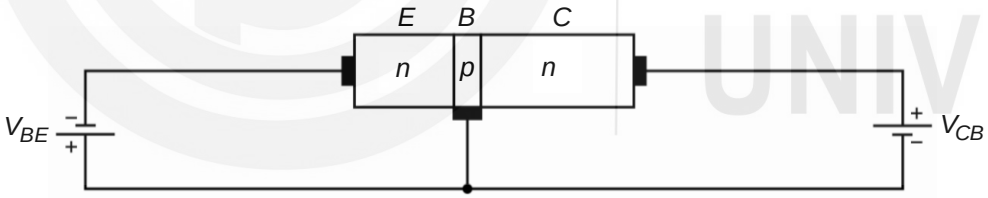


Fig. 3.4: **An** *n*-*p*-*n* **transistor when the emitter-base junction is forward biased and collector-base junction is reverse biased.**

We now ask: **What happens when we forward bias the emitter-base junction and reverse bias the collector-base junction** in a transistor? Refer to Figs. 3.5a, b and c. Note that the base (*p*-type) is biased positively with respect to the emitter (*n*-type). So, what will happen in this kind of biasing?

Note that there are free electrons in the emitter region (Fig. 3.5a). When the applied voltage V_{BE} is greater than the barrier potential (0.6 to 0.7 V for silicon transistor), these electrons enter the base region (see Fig. 3.5b). Once inside the base, these electrons can flow either through the thin base into the external base lead, or across the collector junction into the collector region. Since the base region is very thin and it receives a large number of electrons, therefore, for $V_{BE} > 0.7\,\text{V}$, most of these electrons diffuse into the collector depletion layer.

The free electrons in this layer are pushed into the collector region (by the depletion layer field and the positive voltage V_{CB} applied to the collector terminal) as shown in Fig. 3.5c, and flow into the external collector lead.

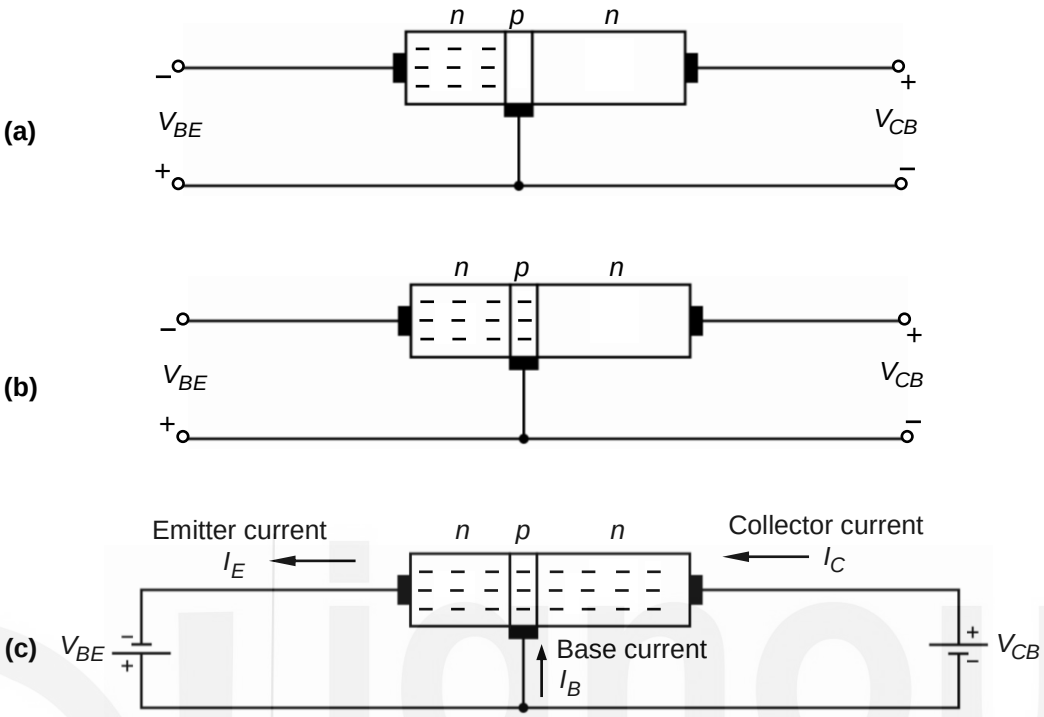


Fig. 3.5: **In an *n-p-n* transistor with forward-biased emitter and reverse-biased collector, a) emitter has many free electrons; b) free electrons are injected into base; c) free electrons diffuse to collector through base.**

So, due to the extremely thin base region and the positive voltage applied to the collector, almost all electrons diffuse through the base to the collector (Fig. 3.5c). This results in **current flow into the collector**. This current is called the **collector current** and denoted by I_C. It is of the order of milliamperes in a typical transistor.

Now let us see if any current flows through the base. Since the base region is very thin, most of the electrons cross to the collector region. Only a very small number of electrons flow out of the base terminal.

So, the **base current** that **flows into the base terminal** is very small. It is denoted by I_B and is of the order of microamperes. In most transistors, more than 95 percent emitter-injected electrons flow to the collector; less than 5 percent flow out in the external base lead.

The emitter current (denoted by I_E) is the current that flows **out of the emitter**. Typically, the emitter and collector currents are of the order of milliamperes and the base current is a few microamperes.

In sum, we can say that a steady stream of electrons leaves the negative source terminal and enters the emitter region. The forward bias forces these electrons to enter the base region. Almost all these electrons diffuse into the collector depletion layer, through the base due to the depletion layer field and the positively biased collector resulting in a steady stream of electrons into the collector region.

Fig. 3.6a shows the circuit symbol of an *n-p-n* transistor. The arrow shows the direction of current flow when the emitter-base *p-n* junction is biased in the forward direction, which means that the base is biased positively with respect to the emitter.

Similarly, the circuit symbol of a *p-n-p* transistor is shown in Fig. 3.6b.

You have learnt that the transistor has two *p-n* junctions, one of which is forward biased and the other reverse biased. Recall from Unit 2 that the forward biased *p-n* junction has low forward resistance and reverse biased *p-n* junction has high reverse resistance.

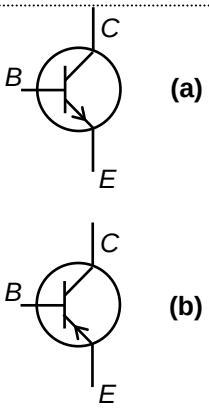


Fig. 3.6: **Circuit symbols of a)** *n-p-n* **transistor; b)** *p-n-p* **transistor.**

You should always remember that

> The emitter-base (*p-n*) junction in a transistor is forward biased and offers low forward resistance. The collector-base (*p-n*) junction in the transistor is reverse biased and, therefore, offers high reverse resistance.

This property of the transistor, of transferring a weak input signal from low resistance circuit to high resistance circuit, results in the input signal getting amplified in a particular configuration called the common emitter configuration (about which you will learn in Unit 4).

This happens because even if the collector current is in milliamperes, the **high reverse resistance** circuit results in an output signal having much larger amplitude than the input signal.

The transistor gets its name from what was historically referred to as *transresistance* : (*trans from transfer + istor from resistor*). Thus, transfer + resistor = transistor.

From this discussion, **you should not conclude that you can connect two discrete diodes back-to-back to get a transistor**.

This is because in such a circuit, each diode has two doped regions, and the overall circuit would have four doped regions. Also, the base region is not the same as in a transistor.

The key to transistor action, therefore, is the lightly doped thin base between the heavily doped emitter and the intermediately doped collector. Free electrons passing through the base stay in the base for a short time and reach the collector.

Effectively, the base current controls the BJT operation and it is a **current-controlled device**.

Let us now write down the relationship between the emitter, base and collector currents. Applying Kirchhoff's current law to the transistor as if it were a single node, we get:

$$I_E = I_B + I_C \qquad (3.1a)$$

Since $I_B << I_C$, we can write:

$$I_E \approx I_C \qquad (3.1b)$$

Always remember the sign convention for the currents in a BJT given ahead.

Recall that the convention for current direction is that the direction of current flow is taken to be opposite to that of the direction of electron flow. The emitter, base and collector **currents are taken to be positive** when these flow **into** the transistor. For the *n*-*p*-*n* transistor, the emitter current is negative and the collector and base currents are positive. For the *p*-*n*-*p* transistor, the emitter current is positive and the collector and base currents are negative.

Now, before we move on to the next section on the field effect transistor, you may like to revise the biasing and working of a bipolar junction transistor. Attempt SAQ 2.

SAQ 2 – Bipolar junction transistor biasing and working

a) Draw a labelled diagram similar to Fig. 3.4 showing how a *p*-*n*-*p* transistor is biased.

b) State the magnitudes of the emitter, base and collector currents in the *p*-*n*-*p* transistor. The emitter current in a BJT is 5.0 mA and the collector current is 4.998 mA. What is the value of the base current in it?

3.3 FIELD EFFECT TRANSISTOR

In this section, you will learn about the construction, biasing and working mechanism of the field effect transistor or FET. The FET is also a double junction device but it is different from the BJT: Whereas in the bipolar junction transistor both majority and minority charge carriers flow, in an FET **only majority charge carriers (either electrons or holes, but not both) flow**. That is why it is called a **unipolar device**. (Remember that **BJT** is a **bipolar device** because the currents in it are due to both electrons and holes).

The first FET was built and patented by the German physicist Heinrich Welker in 1945. It was called the junction field-effect transistor (JFET) about which you will learn in this section. A working JFET was built by George F. Dacey and Ian M. Ross in the Bell Laboratories in 1953.

Field effect transistors are of two types:

- **Junction field effect transistor** (JFET)
- **Metal oxide semiconductor field effect transistor** (MOSFET)

We will focus on the construction, biasing and working mechanism of JFET in our discussion. This device may be entirely new for you. So read the next two sections carefully.

3.3.1 Construction

Like the BJT, the JFET is also a three-terminal semiconductor device. It is made up of either an *n*-type or a *p*-type silicon substrate. Imagine a substrate

or bar of *n*-type semiconductor as shown in Fig. 3.7a. Note that its lower end is labelled as **source** and upper end as **drain**. If we apply a voltage so that the drain is positive with respect to the source, electrons will flow from the source to the drain in the substrate.

Now refer to Fig. 3.7b. To create a JFET, the *n*-type substrate/bar of semiconductor is doped heavily with *p*-type material at two sides so that there are two *p*-*n* junctions at the sides of the JFET. The *p*-regions are connected internally so that there is a **single gate** lead. This kind of JFET is called the *n*-channel JFET. The majority charge carriers in it are electrons.

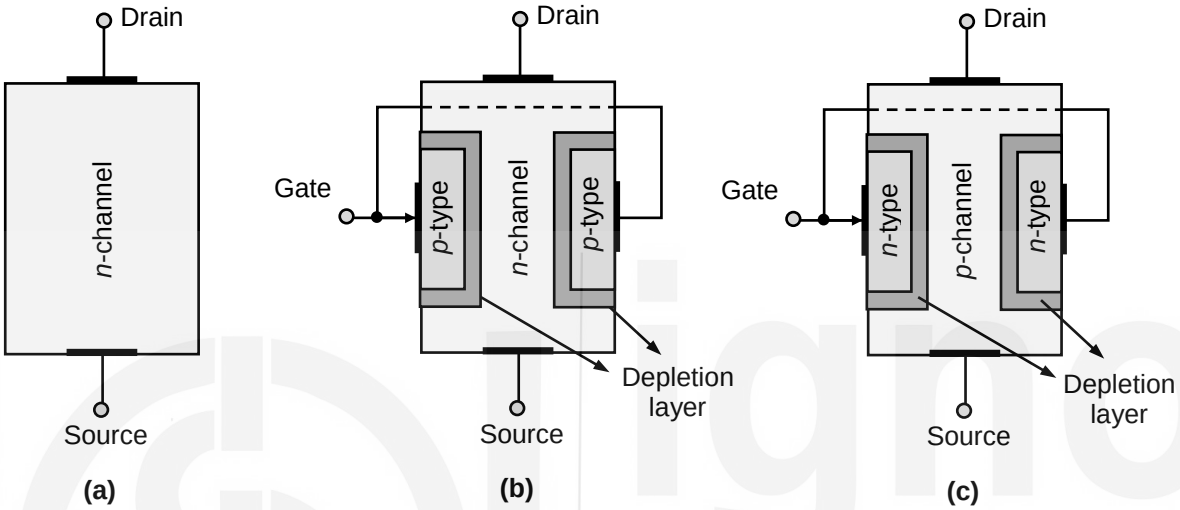


Fig. 3.7: **a) Drain and source in an *n*-channel JFET; b) its structure; c) a *p*-channel JFET.**

If the substrate or bar is of *p*-type material, then its sides are doped with *n*-type material and it is called a *p*-channel JFET (see Fig. 3.7c). Holes are majority charge carriers in it.

Just as we focused on the *n*-*p*-*n* transistor in Sec. 3.2, we will mainly discuss the *n*-channel JFET in this section. Then you can apply the ideas to *p*-channel JFET for practice and revision.

From Fig. 3.7b, note that the major part of the JFET is the *n*-channel. The top of the channel is connected through an ohmic contact to a terminal called the **drain** (*D*). The lower end of the channel is connected through another ohmic contact to a terminal called the **source** (*S*). The two *p*-type materials are connected to each other internally and externally to the third terminal called the **gate** (*G*).

So, always remember that

> In an *n*-channel JFET, the drain and source are connected to the two opposite ends of the *n*-channel and the gate to the two layers of the *p*-type material. The *n*-type material is heavily doped in the regions near the source and drain terminals.

You can see that like the BJT, the JFET is a three-terminal device (source, gate, drain) that has two *p*-*n* junctions. Thus, in an *n*-channel JFET,

Source *S* is the terminal of the JFET through which the majority carriers **enter** the channel giving rise to the source current I_S. Following the convention of the direction of current as opposite to the flow of electrons, we take the current flowing into the channel at *S* as positive.

Drain *D* is the terminal through which the majority carriers **leave** the channel giving rise to the drain current I_D. By convention, we take the current leaving the channel at *D* as positive.

Gate *G* is the term used for the terminal attached to the heavily doped (*p*+) regions on both sides of the *n*-type channel.

Channel is the region of *n*-type material between the two gate regions through which the majority carriers move from source to drain.

The drain-to-source voltage is denoted by V_{DS} and in an *n*-channel JFET, *D* is positively biased with respect to *S*.

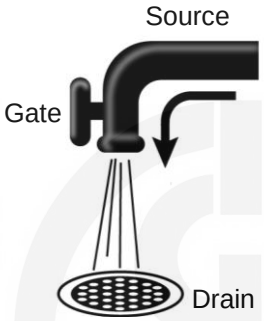


Fig. 3.8: **Analogy of water flow from water tap for the JFET.**

We can use an analogy to understand the role of these terminals in a simple way. Consider water flowing from a tap (Fig. 3.8). The source of water pressure is similar to the voltage applied between drain and source. Just as the water source establishes water flow from the tap, so does the voltage V_{DS} establish a flow of electrons from the source. A potential applied at the gate controls the flow of electrons to the drain (just as water flow is controlled by turning the tap's handle). Water flows out of the drain as do electrons in the JFET. As in a tap, the drain and source terminals are at the opposite ends of the *n*-channel.

You may now like to revise the construction of a JFET. Solve SAQ 3.

SAQ 3 – JFET construction

a) What is the main difference between the construction of an *n*-channel JFET and a *p*-channel JFET?

b) Name the majority charge carriers in *n*-channel and *p*-channel JFETs.

c) Fill in the blanks in the following sentence:

In a *p*-channel JFET, the source is the terminal through which ……..enter the substrate and leave the channel through the terminal called the …… The gate refers to the heavily doped….-type material on both sides of the JFET.

Let us now describe the biasing and working of an *n*-channel JFET.

3.3.2 Biasing and Working Mechanism

In the absence of an applied voltage, there exist depletion regions at the two *p*-*n* junctions on the sides of the *n*-channel JFET. Since the *p*-type material is heavily doped, the depletion layer is thin in the *p*-region and extends into the *n*-region (see Figs. 3.7a and b).

We now ask: **How is the JFET biased for its normal functioning**?

Refer to Fig. 3.9, which shows an *n*-channel JFET. We first consider the case when a positive voltage (V_{DS}) is applied to the drain with respect to the source and the gate is connected directly to the source so that the voltage V_{GS} between the gate and source is zero:

a) $V_{DS} > 0$ **and** $V_{GS} = 0$ V

In this case, the gate and the source are at the same potential and under no bias conditions when V_{DS} is zero, the depletion layers at the *p*-*n* junctions will be as shown in Fig. 3.7a.

Now what happens when the drain is biased positively with respect to the source by a battery or power supply (V_{DD})? Electrons in the *n*-channel move towards the drain since it is positively biased. This results in the flow of drain current in the direction shown in Fig. 3.9. The flow of electrons is unhindered in the channel and hence, the source current is equal to the drain current when V_{GS} is zero.

REMEMBER: The polarity of the drain-source voltage will be reversed for a *p*-channel JFET and so will the directions of the drain and source currents but the underlying physics remains the same.

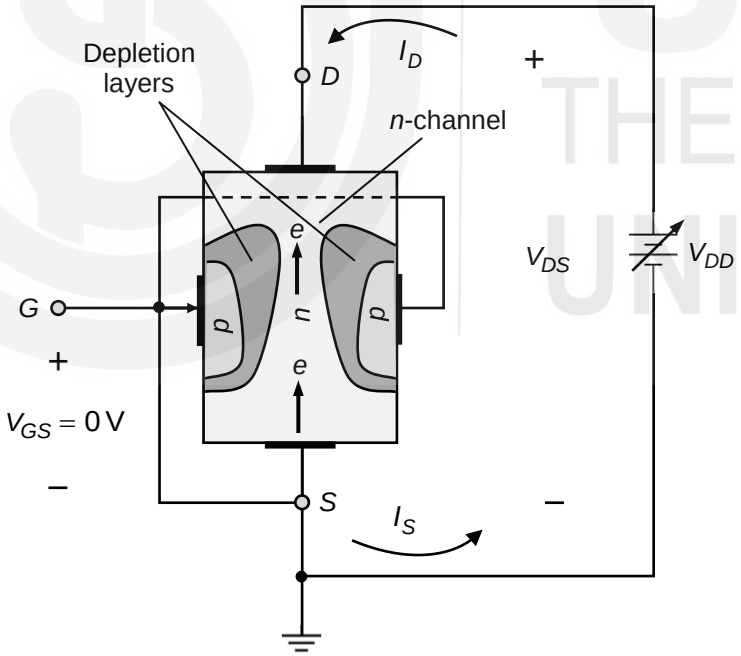


Fig. 3.9: **Working of JFET when** $V_{DS} > 0$ **and** $V_{GS} = 0$ V**.**

Also note from Fig. 3.9 that the *p*-*n* junctions are reverse biased in this case. You can see that the *p*-regions are negatively biased with respect to the *n*-region near the drain because the gate terminal is connected to the source terminal, which is negatively biased with respect to the drain. So, virtually no electrons move to the gate and the gate current is zero. Thus, we have:

$$I_D = I_S \qquad (3.2a)$$

and $I_G = 0$ (3.2b)

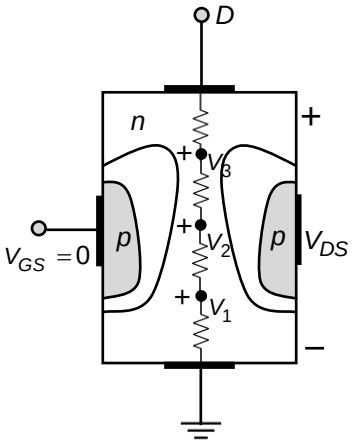


Fig. 3.10: **Variation in reverse biasing voltages across the *p-n* junctions; $V_3 > V_2 > V_1$ and $V_{GS} = 0$ V.**

Now did you note in Fig. 3.9 that the depletion layers of the *p-n* junctions are wider towards the drain end as compared to the source end? You may like to know: **Why is it so**?

This is because of the resistance of the channel. Let us assume that the resistance of the channel is uniform. Then we can break it down into segments of resistances as shown in Fig. 3.10. The voltage applied between the drain and the source drops across these resistances, and we have maximum voltage at the drain and minimum at the source.

As a result, the upper ends (towards the drain end of the channel) of both *p-n* junctions will be more reverse biased compared to their lower ends (towards the source end of the channel). In fact, the reverse bias between the *p*-type gate and *n*-type channel is **zero near the source end** and **maximum near the drain end**. Recall from Unit 2 that the greater the reverse bias in a *p-n* junction, the wider will be the depletion layer.

Therefore, **the depletion layers are much wider near the drain end and extend more into the channel in comparison with the source end**. This is how we get the **wedge shape of the *n*-channel** shown in Fig. 3.9.

Now what happens when we increase V_{DS} ? The drain current will increase in accordance with Ohm's law. Fig. 3.11 shows the drain characteristics of a JFET, the plot of the drain current (I_D) and drain-source voltage (V_{DS}) for $V_{GS} = 0$. Note that when V_{DS} is gradually increased, in the initial stages, the curve is linear. This is because the resistance is constant in this ohmic region.

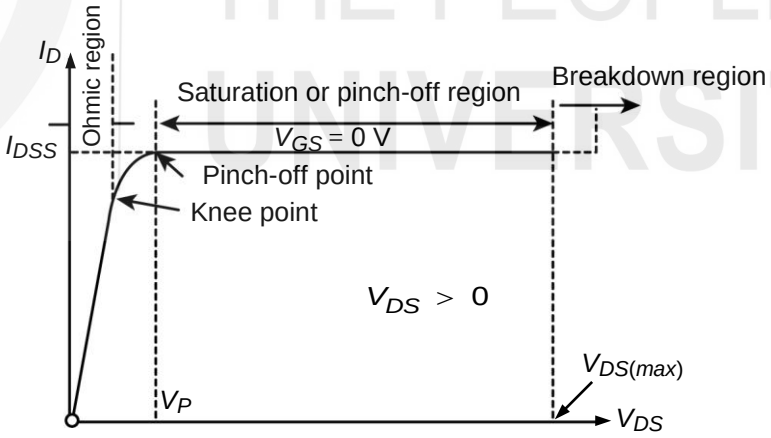


Fig. 3.11: **Drain characteristics of JFET when $V_{DS} > 0$ and $V_{GS} = 0$ V.**

As V_{DS} is increased further, the depletion layers become wider and the channel width reduces. The reduced width decreases the path of conduction for electrons, causing the resistance of the channel to increase. Then the curve begins to level off and eventually the drain current I_D saturates. After that it does not increase any further with an increase in V_{DS}. The drain-source voltage at which the drain current saturates is called the **pinch-off voltage** and denoted by V_P. So, I_D saturates for $V_{DS} \geq V_P$. This happens when V_{DS} is increased to a value such that the two depletion layers "touch" each other as shown in Fig. 3.12 resulting in *pinch-off*.

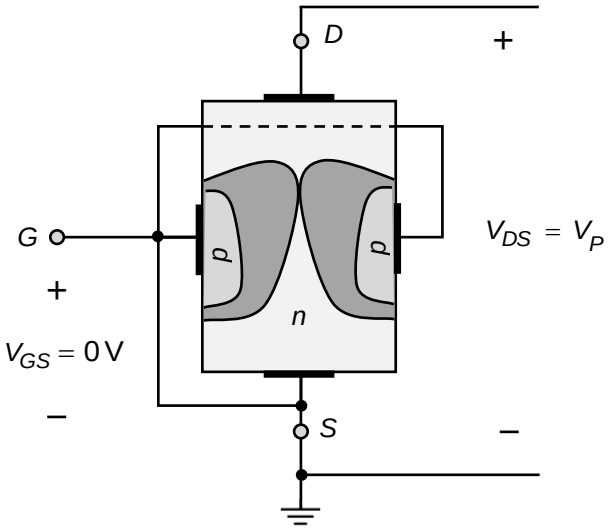


Fig. 3.12: **Pinch-off in JFET;** $V_{DS} = V_P$ **and** $V_{GS} = 0$ V**.**

Do not misunderstand the term pinch-off. It does NOT mean that the drain current is "pinched off". That is, the drain current does not become zero!

Actually, the drain current becomes constant (saturates) and is denoted by I_{DSS} as shown in Fig. 3.11. What happens is that a very small channel does exist and the saturation current flows through it. The notation I_{DSS} is used for **the saturation current when the gate and the source are short-circuited**. So, always remember that

> I_{DSS} **is defined by the conditions** $V_{DS} \geq |V_P|$ **and** $V_{GS} = 0$ **and is the maximum drain current in a JFET.**

You may like to know: **Why does the drain current not become zero at the pinch-off voltage**? It is because even if the depletion layers from both sides are touching, there is always some narrow channel allowing the drift current to pass through.

Thus, when V_{DS} becomes greater than V_P, the drain current becomes constant and remains at the same level. In effect, the JFET becomes a current source for $V_{DS} \geq V_P$ (Fig. 3.13).

Note from Fig. 3.13 that the drain saturation current I_{DSS} is constant but the voltage V_{DS} $(> V_P)$ is determined by the load in the circuit.

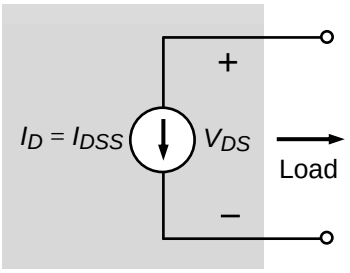


Fig. 3.13: Current source equivalent of JFET when $V_{DS} > V_P$ **and** $V_{GS} = 0$ V**.**

Let us now ask: **What kind of drain characteristics do we obtain when the gate source voltage** V_{GS} **is non-zero**?

So, we consider the case when a **negative voltage is applied to the gate** of an *n*-channel JFET.

b) $V_{DS} > 0$ **and** $V_{GS} < 0$ V

Refer to Fig. 3.14. Note from the figure that the voltage from gate to source is negative, that is, the gate is negatively biased with respect to the source. Thus, the gate-source *p*-*n* junctions are reverse biased by V_{GS}. So, if the

input signal is applied to the gate terminal, it will experience a high resistance. In other words, the JFET will offer high input impedance to the input signal. (It is always beneficial to have high input impedance so that the current drawn from the signal source is very small, and hence, there are no loading effects.) Also, the depletion layers will be wider due to the reverse biased gate junction as compared to the case when V_{GS} is zero (case a).

Now if we apply V_{DS}, pinch-off will occur at a lower value of voltage V_{DS} because depletion layer widths for pinch-off will now be reached at a **lower value** of V_{DS}. The gate current I_G will be zero in this case too because V_{GS} is negative and electrons will not flow through the gate.

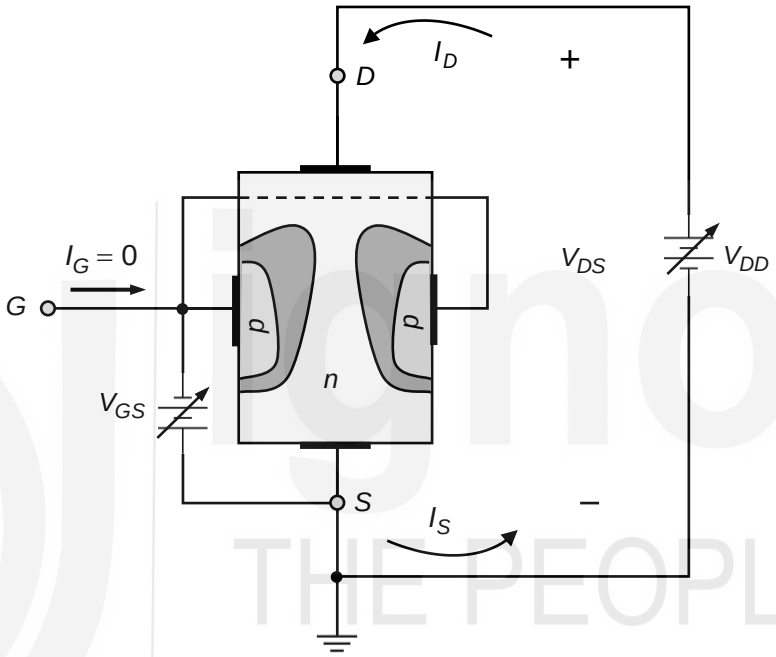


Fig. 3.14: **Working of JFET when $V_{DS} > 0$ and $V_{GS} < 0$.**

Therefore, the effect of applying a negative bias to the gate is that the drain current will reach its saturation level for a lower value of V_{DS}. Also, the saturation level value of the drain current (I_{DSS}) will decrease. Fig. 3.15 shows the drain (*I*-*V*) characteristics of an *n*-channel JFET.

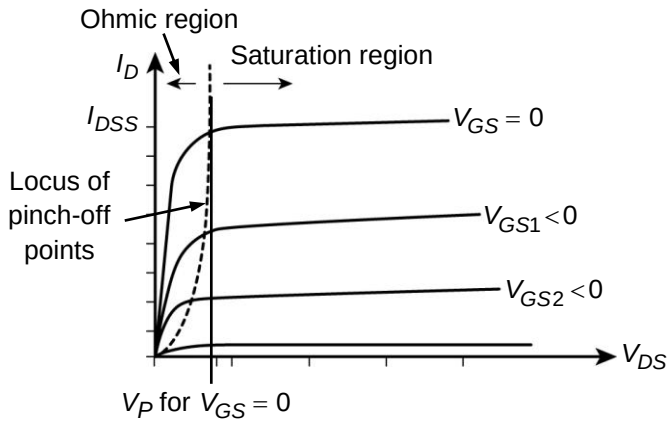


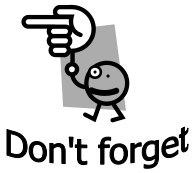
Fig. 3.15: **Drain characteristics of an *n*-channel JFET when $V_{DS} > 0$ for different values of $V_{GS} \leq 0$.**

Note from the figure that as the voltage V_{GS} is made more and more negative, saturation of the drain current occurs at lower values of V_{DS}. When V_{GS} reaches the pinch-off voltage, that is, $V_{GS} = -V_P$, the saturation current becomes zero for all practical purposes. At that point, the JFET is "turned off".

Always remember,

The gate-source voltage controls current flow in the JFET.

The voltage $V_{GS} = V_P$ is that value of V_{GS} for which the drain current is zero ($I_D = 0$). The voltage V_P is <u>negative</u> for an *n*-channel JFET and <u>positive</u> for a *p*-channel JFET.



For very high values of V_{DS}, these curves suddenly become almost vertical (see Fig. 3.11). The vertical rise in current indicates breakdown of the JFET. The operation of JFET at such high voltage has to be avoided to prevent damaging it.

The circuit symbol of an *n*-channel JFET is shown in Fig. 3.16. **Remember that the direction of the arrow represents the direction in which the gate current would flow if the *p*-*n* junction were forward biased**.

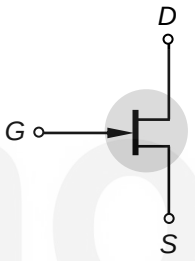


Fig. 3.16: **Circuit symbol of an *n*-channel JFET.**

With this, we end the discussion on JFET. You should try SAQ 4 to check whether you have understood the biasing and working of a JFET.

SAQ 4 – JFET biasing and working

a) Draw the circuit diagram analogous to Fig. 3.9 for a *p*-channel JFET for $V_{GS} = 0$ V.

b) Draw the circuit symbol of a *p*-channel JFET.

In this section, you have learnt about one type of a field effect transistor, namely, the JFET. We have not discussed the MOSFET about which you will learn in higher classes.

You have learnt that in the junction field effect transistor, the output current, i.e., the drain current between the drain and the source is controlled by the gate-to-source voltage V_{GS}, the input voltage applied at the gate with reference to the source.

So, a field-effect transistor **uses an electric field** to control the flow of current in a semiconductor. That is where the name field effect transistor comes from. And that is why a JFET is a **voltage-controlled device**.

You have learnt that the value of V_{GS} determines how much drain current flows in a JFET. When $V_{GS} = 0$, the saturation drain current I_{DSS} flowing through the JFET is maximum. This is why a JFET is referred to as **a normally on** device.

If $V_{GS} = |V_P|$, ($-V_P$ for *n*-channel JFET and $+V_P$ for *p*-channel JFET), the depletion layers touch and the drain current is cut off. Then the JFET is turned off.

We now summarise what you have learnt in this unit.

3.4 SUMMARY

Concept	Description
Bipolar junction transistor and its construction	■ A bipolar junction transistor is a double junction three-terminal device. It is constructed by doping a single crystalline semiconducting material (usually silicon or germanium) in two ways: • by sandwiching a thin layer of *p*-type semiconductor between two outer layers of *n*-type semiconductor forming an *n*-*p*-*n* transistor; or • by sandwiching a thin layer of *n*-type semiconductor between two outer layers of *p*-type semiconductor forming a *p*-*n*-*p* transistor. This kind of doping makes it possible to use a BJT as an amplifier or a switch. Both electrons and holes give rise to the current in a BJT. Hence, it is given the name bipolar. The three regions in an *n*-*p*-*n* transistor (*n*, *p* and *n*) or a *p*-*n*-*p* transistor (*p*, *n* and *p*) are called **emitter** (*E*), **base** (*B*) and **collector** (*C*), respectively.
Biasing and working of BJT	■ Transistor action in a BJT occurs because of the different doping levels of the emitter, base and collector regions, their different widths and the way it is biased. • The base is lightly doped and is very thin compared to the emitter and collector regions. The emitter region is more heavily doped compared to the collector region. • Typically, in a transistor used as an amplifier, the emitter-base junction is forward biased and the collector-base junction is reverse biased. • The emitter current is large (in milliamperes) and the base current is very small (in microamperes or nanoamperes) so that the collector current is almost equal to the emitter current: $I_E = I_B + I_C$ Since $I_B << I_C$, we can write: $I_E \approx I_C$ • The BJT is a current-controlled device as its operation is controlled by the currents in it.
Junction field effect transistor	■ The field effect transistor is also a double junction three-terminal device that has two *p*-*n* junctions. An *n*-channel JFET is constructed by heavily doping the sides of an *n*-type substrate with *p*-type material so that there are two

p-*n* junctions at the sides of the JFET. A *p*-channel JFET is constructed by heavily doping the sides of a *p*-type substrate with *n*-type material. JFETs have only one type of charge carriers for current flow and that is why these are unipolar devices. Electrons are the majority charge carriers in an *n*-channel JFET and holes in a *p*-channel JFET.

The three terminals are called the source (*S*), gate (*G*) and drain (*D*) analogous to the emitter, base and collector in a BJT. The gate regions are heavily doped as compared to the channel. So, the depletion layer extends more in the channel than in the gate regions.

Biasing and working of JFET

- For its normal functioning, an *n*-channel JFET is biased so that its drain is positive with respect to the source, and the gate-source voltage is zero or negative.

Thus, when $V_{DS} > 0$ and $V_{GS} = 0$ V, electrons in the *n*-channel move from the source towards the drain resulting in the drain current into the JFET. The source current is equal to the drain current when V_{GS} is zero:

$$I_D = I_S \quad \text{and} \quad I_G = 0$$

- When $V_{DS} > 0$, the depletion layers of the *p*-*n* junctions are much wider near the drain end and extend more into the channel in comparison with the source end, resulting in a **wedge-shaped** *n*-channel.
- When V_{DS} is relatively low, the drain current increases in accordance with Ohm's law and the drain current versus the drain-source voltage curve is linear.
- As V_{DS} is increased further, the depletion layer widths increase due to which the channel width decreases causing increased impedance to the flow of drain current. This results in saturation of the drain current, which does not increase with any further increase in V_{DS}.
- The drain-source voltage at which the drain current saturates is called the **pinch-off voltage** V_P.
- Pinch-off in a JFET occurs when V_{DS} is increased to a value such that the two depletion layers "touch" each other.
- I_{DSS} is defined by the conditions $V_{DS} \geq |V_P|$ and $V_{GS} = 0$ and is the maximum drain current in any JFET.
- When $V_{DS} > 0$ and $V_{GS} < 0$, the drain current reaches its saturation level for a lower value of V_{DS} and its value decreases.
- As the voltage V_{GS} is made more and more negative, saturation of the drain current occurs at lower values of V_{DS}.
- When $V_{GS} = -V_P$, the saturation current becomes zero and at that point, the JFET is "turned off".
- At very high values of V_{DS}, breakdown occurs in a JFET.

3.5 TERMINAL QUESTIONS

1. Fill in the blanks in the following sentences:

   a) The BJT and JFET have ……*p*-*n* junctions. That is why these are called……..junction devices.

   b) A BJT is a ……..controlled device while a JFET is a……controlled device.

   c) The …… and the ……. control the flow of current in the BJT and JFET, respectively.

   d) In the BJT, current flow is due to ……while in a JFET, current flow is due to ……

   e) In a BJT, the ……..junction is forward biased and the ……junction is reverse biased; in a JFET, the ……. junction is reverse biased.

2. a) Name the majority and minority charge carriers in *p*-*n*-*p* and *n*-*p*-*n* transistors and explain which currents these constitute in them.

   b) What is meant by bipolar and unipolar devices?

3. Draw a *p*-*n*-*p* transistor biased for amplifier operation and label the currents that flow in it. Explain its working.

4. Can collector current be larger than emitter current in a BJT? Explain.

5. Write the main differences in the working mechanisms of the *p*-*n*-*p* and *n*-*p*-*n* transistors.

6. Why is JFET called a field effect transistor?

7. Explain pinch-off in a *p*-channel JFET with the help of a diagram.

8. Draw the labelled drain characteristics of a *p*-channel JFET showing the ohmic, saturation and breakdown regions. What is the value of the drain current when $V_{GS} = V_P$ in a *p*-channel JFET? When is the drain current in it maximum?

9. Explain why the JFET is termed a voltage-controlled device.

10. When a *p*-channel JFET is “turned off”, what happens to the saturation current in it? When the gate voltage becomes more negative in an *n*-channel JFET, what happens to the channel between the depletion layers?

3.6 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. a) Fig. 3.17 shows the labelled diagram of the structure of the *p*-*n*-*p* transistor.

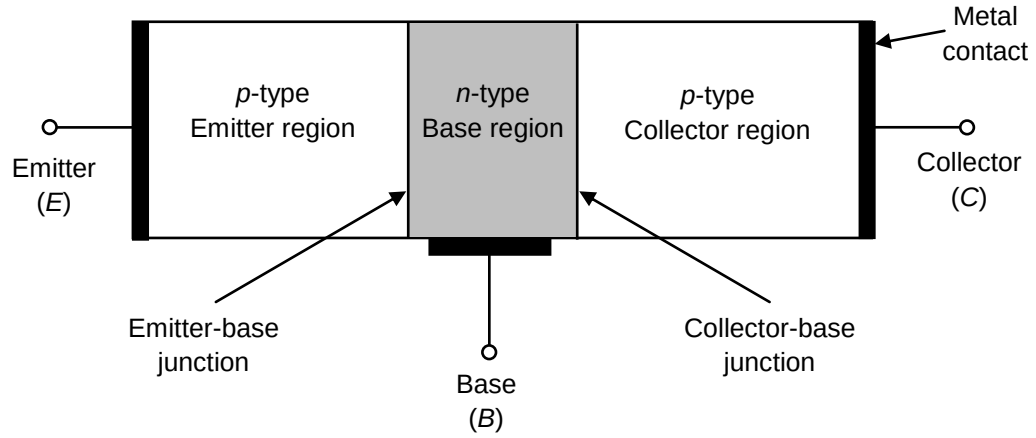


Fig. 3.17: **Structure of a *p-n-p* transistor.**

b) (i) increasing levels of doping: base, collector, emitter;

(ii) increasing width: base, emitter, collector.

2. a) Fig. 3.18 shows the labelled diagram of the biasing of the *p-n-p* transistor.

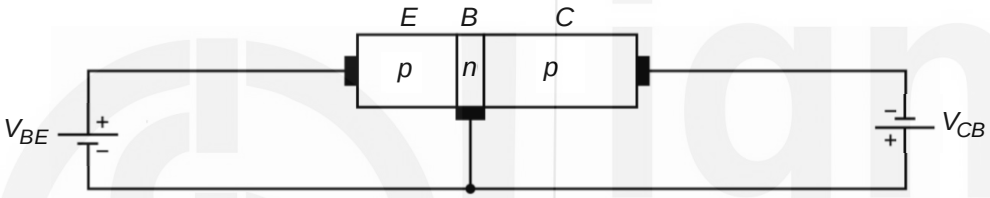


Fig. 3.18: **A *p-n-p* transistor when the emitter-base junction is forward biased and collector-base junction is reverse biased.**

b) The magnitudes of the emitter, base and collector currents in the *p-n-p* transistor are the same as those in an *n-p-n* transistor, that is, the emitter and collector currents are in milliamperes and the base current is in microamperes or nanoamperes. The base current is given from Eq. (3.1a) as:

$$I_B = I_E - I_C = 5.0\,\text{mA} - 4.998\,\text{mA} = 0.002\,\text{mA} = 2.0\,\mu A$$

3. a) In the *n*-channel JFET, the substrate/bar is made of *n*-type semiconductor and its sides are doped by *p*-type semiconductors. But in a *p*-channel JFET, the substrate/bar is made of *p*-type semiconductor and its sides are doped by *n*-type semiconductors.

b) The majority charge carriers in *n*-channel and *p*-channel JFETs are electrons and holes, respectively.

c) In a *p*-channel JFET, the source is the terminal through which holes enter the substrate and leave the channel through the terminal called the drain. The gate refers to the heavily doped *n*-type material on both sides of the JFET.

4. a) The circuit diagram analogous to Fig. 3.9 for the *p*-channel JFET when $V_{GS} = 0$ V is shown in Fig. 3.19.

b) The circuit symbol of a *p*-channel JFET is shown in Fig. 3.20.

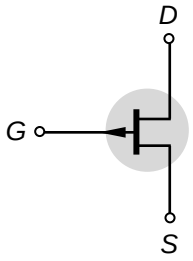


Fig. 3.20: **Circuit symbol of a *p*-channel JFET.**

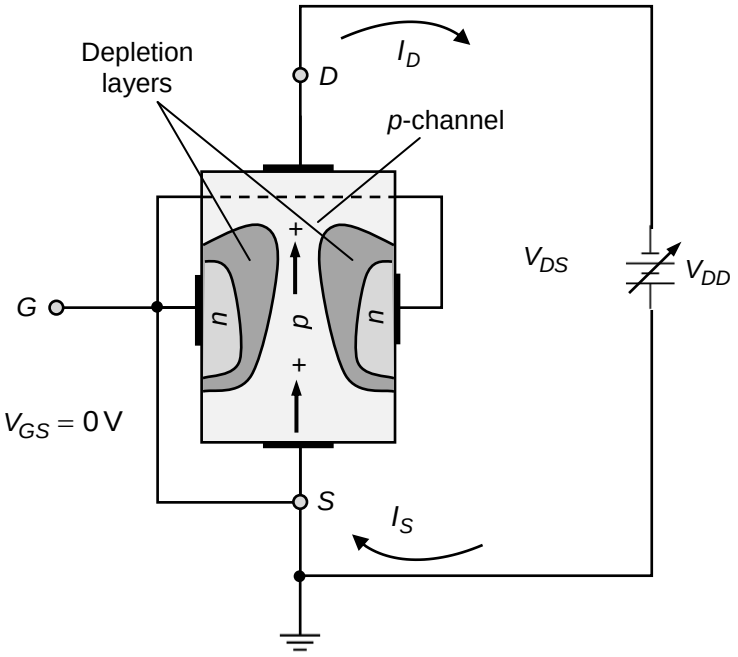


Fig. 3.19: **Circuit diagram for *p*-channel JFET when $V_{GS} = 0$ V.**

Terminal Questions

1. a) The BJT and JFET have two *p*-*n* junctions. That is why these are called double junction devices.
   b) A BJT is a current-controlled device while a JFET is a voltage-controlled device.
   c) The base and the gate control the flow of current in the BJT and JFET, respectively.
   d) In the BJT, current flow is due to both majority and minority charge carriers (electrons and holes) while in a JFET, current flow is due to only majority charge carriers (electrons or holes).
   e) In a BJT, the emitter-base junction is forward biased and the collector-base junction is reverse biased; in a JFET, the gate-source junction is reverse biased.

2. a) In a *p*-*n*-*p* transistor, the majority and minority charge carriers are holes and electrons, respectively. The emitter and collector currents are due to the motion of holes, and the base current is due to the flow of electrons. In an *n*-*p*-*n* transistor, the majority and minority charge carriers are electrons and holes, respectively. The emitter and collector currents are due to the flow of electrons, and the base current is due to the motion of holes.
   b) In bipolar devices, current flow is due to both majority and minority charge carriers. In unipolar devices, current flow is due to only majority charge carriers.

3. Fig. 3.21 shows a *p*-*n*-*p* transistor biased for amplifier operation along with the currents that flow in it. Its working is analogous to that of an *n*-*p*-*n* transistor and is described below. Note from Fig. 3.21 that in a *p*-*n*-*p* transistor too, the emitter-base *p*-*n* junction is forward biased while the collector-base *p*-*n* junction is reverse biased for its normal operation as an amplifier.

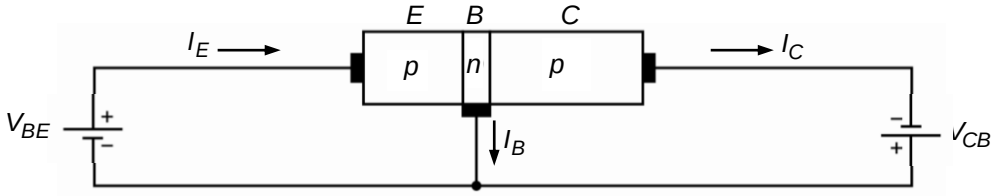


Fig. 3.21: **A *p*-*n*-*p* transistor biased for amplifier operation with its emitter-base junction forward biased and collector-base junction reverse biased.**

When the emitter (*p*-type) is biased positively with respect to the base (*n*-type), the holes in the emitter region enter the base region when the applied voltage V_{BE} becomes greater than the barrier potential. Once inside the base, the holes can flow either through the thin base into the external base lead, or across the collector junction into the collector region. Since the base region is very thin and it receives a large number of holes, therefore, most of these holes diffuse into the collector depletion layer. The free holes in this layer are pushed into the collector region (by the depletion layer field and the negative voltage V_{CB} applied to the collector terminal) and flow out of the collector. Due to the extremely thin base region and the negative voltage applied to the collector, almost all holes diffuse through the base to the collector. This results in current flow out of the collector. Since the base region is very thin, most of the holes cross to the collector region. Only a very small number of holes flow out of the base terminal. Therefore, the base current that flows out of the base terminal is very small. The emitter current flows into the emitter. Typically, the emitter and collector currents are of the order of milliamperes and the base current is a few microamperes.

4. No, the collector current cannot be larger than the emitter current in a BJT since a very small number of majority charge carriers from the emitter do flow out of (or into) the base of the BJT. Thus, the emitter current is the sum of the base current and the collector current.

5. The main differences in the working mechanisms of the *p*-*n*-*p* and *n*-*p*-*n* transistors are in their charge carriers and biasing, and the directions of the resulting currents. In an *n*-*p*-*n* transistor, the majority charge carriers are electrons, the *p*-type base is biased positively with respect to the *n*-type emitter, and the *n*-type collector is biased positively with respect to the *p*-type base. The emitter current flows out of the emitter, the base and collector currents flow into the base and collector regions, respectively. In a *p*-*n*-*p* transistor, the majority charge carriers are holes, the *n*-type base is biased negatively with respect to the *p*-type emitter, and the *p*-type collector is biased negatively with respect to the *n*-type base. The emitter current flows into the emitter, the base and collector currents flow out of the base and collector regions, respectively.

6. In a field effect transistor, the output current, i.e., the drain current between the drain and the source is controlled by the gate-to-source voltage V_{GS}, the input voltage applied at the gate with reference to the source. So, it uses an electric field to control the flow of current in a semiconductor. That is why the JFET is called a field effect transistor.

7. Fig. 3.22 shows the pinch-off in a *p*-channel JFET.

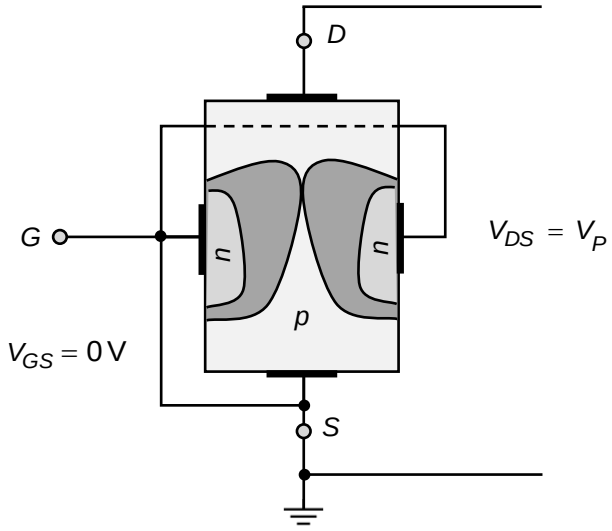


Fig. 3.22: **Pinch-off in *p*-channel JFET; $V_{DS} = V_P$ and $V_{GS} = 0$ V.**

Just as in an *n*-channel JFET, when V_{DS} is increased in a *p*-channel JFET, the reverse bias across the gate-source *p*-*n* junction increases and the widths of the depletion layers increase resulting in a reduced channel width. This decreases the path of conduction for holes, causing the resistance of the channel to increase. Eventually the drain current I_D saturates. After that it does not increase any further with an increase in V_{DS}. So when V_{DS} is increased to a value V_P such that the two depletion layers "touch" each other as shown in Fig. 3.22, pinch-off occurs in a *p*-channel JFET.

8. The drain characteristic curve for a *p*-channel JFET is shown in Fig. 3.23. The drain current in a *p*-channel JFET is zero when $V_{GS} = V_P$. It is maximum in a *p*-channel JFET when $V_{GS} = 0$ and $V_{DS} = V_P$.

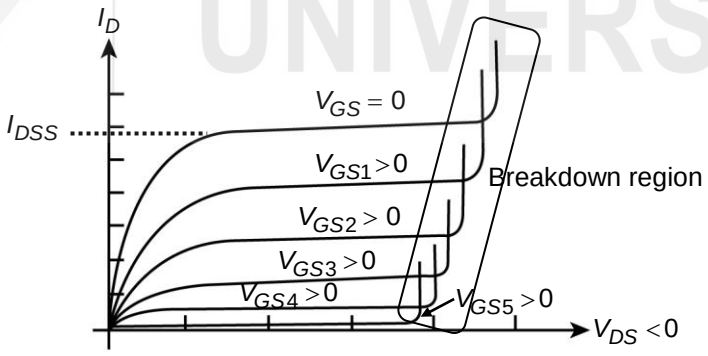


Fig. 3.23: **Drain characteristic curve for a *p*-channel JFET.**

9. The JFET is termed a voltage-controlled device because the gate-source voltage V_{GS} controls the flow of drain current in it.

10. When a *p*-channel JFET is "turned off", the saturation current in it becomes zero. When the gate voltage becomes more negative in an *n*-channel JFET, the channel width between the depletion layers decreases.

The picture above shows a stereo audio amplifier's power transistors mounted on heat sinks.
Source: https://en.wikipedia.org/
Picture credit: Daniel Christensen

UNIT 4

BIPOLAR JUNCTION TRANSISTOR BIASING

Structure

STUDY GUIDE

In this unit, you will learn about the biasing methods of a bipolar junction transistor, which you have learnt in your senior secondary (+2) physics. Essentially, it will be revisiting familiar concepts. But it will help if you revise Sec. 3.2 of Unit 3 before studying this unit. Once again, we advise you to study all sections of this unit thoroughly. Make a note of any concept that you do not understand so that you may ask us or the Counsellor. Solving the SAQs and Terminal Questions given in the unit will help you in reviewing your understanding of concepts better.

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

Marie Curie

4.1 INTRODUCTION

In Unit 3, you have learnt about the physics underlying the working of a bipolar junction transistor. You know about two of its major applications: as an amplifier and as a switch. In this unit, you will learn about how a BJT is biased for its operation so that it can be used for both these applications.

We begin the unit by explaining various configurations of the BJT in Sec. 4.2. You know that there are three terminals in a BJT and one of these has to be common in any given circuit having a BJT. So, you will learn about three configurations: **Common base**, **common emitter** and **common collector**.

We will then focus on the **common emitter configuration** in Sec. 4.3. We will discuss the input-output characteristics of the transistor, the parameters **dc alpha** and **dc beta** (current gain), the **load line** and **quiescent operating point**. Finally, in Sec. 4.4, we will discuss the **transistor biasing methods**.

In the next unit, you will learn about transistor circuit analysis.

Expected Learning Outcomes

After studying this unit, you should be able to:

- describe the common base (CB), common emitter (CE) and common collector (CC) configurations, and draw the corresponding circuits;
- draw and explain the features of input-output characteristics of a BJT in CE configuration;
- draw the load line and show the quiescent operating point on the output characteristics of a BJT in CE configuration and explain their function in its operation; and
- explain the transistor biasing methods with appropriate circuit diagrams.

4.2 CB, CE, CC CONFIGURATIONS OF A BJT

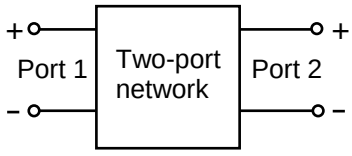


Fig. 4.1: **A two-port network has two ports, i.e., four terminals.**

In Unit 3, you have learnt that the bipolar junction transistor is a three-terminal device and its terminals are called emitter, base and collector. However, a network usually has two ports (input and output) as shown in Fig. 4.1. Hence, we require **four terminals** to represent any device. But since a transistor has only three terminals, we need to use one of these as a **common terminal** for the input and output sides. The common terminal is usually the terminal at ground potential or the closest to the ground potential. There are three possible configurations in a transistor in which one of the three terminals is common to both input and output ports. These are:

1. **Common base configuration**: When the **base** terminal of the transistor is common to input and output circuits.
2. **Common emitter configuration**: When the **emitter** terminal of the transistor is common to input and output circuits.

3. **Common collector configuration**: When the **collector** terminal of the transistor is common to input and output circuits.

In this section, you will learn about these three configurations and how a BJT is biased in all these.

Common Base Configuration

Recall Sec. 3.2 of Unit 3, in which you have learnt about one way of biasing a BJT. Fig. 3.4 actually shows the BJT in a common base configuration. That particular arrangement is called the **common base configuration** because the base terminal is common to both the input and output circuits of the BJT. We show the configuration again for both *n-p-n* and *p-n-p* transistors in Fig. 4.2. Note that this configuration has low input resistance and high output resistance as explained in Sec. 3.2 of Unit 3.

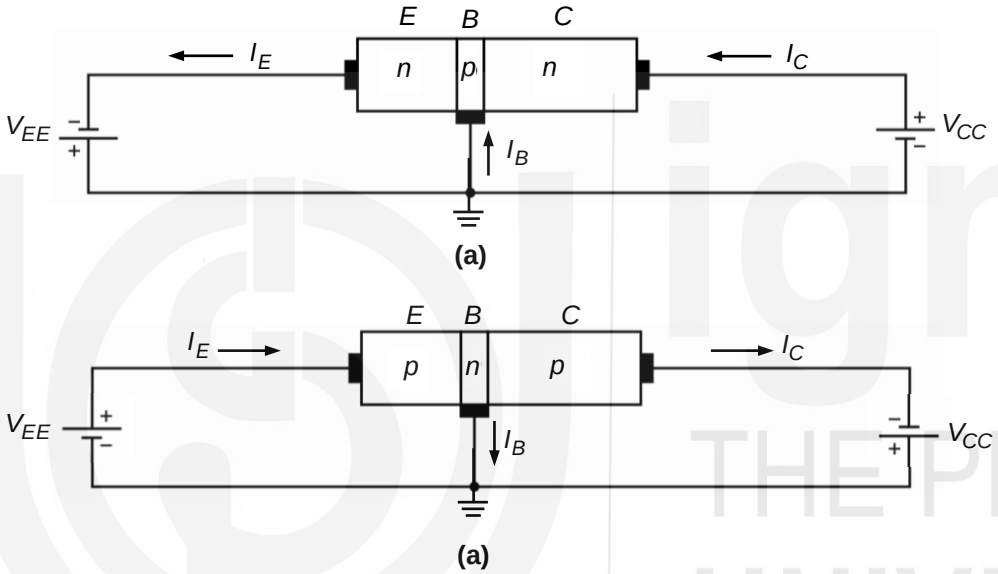


Fig. 4.2: **Common base configuration for a) an *n-p-n* transistor; b) a *p-n-p* transistor.**

Usually, we draw the configuration showing the transistor symbol. Note that in both Figs. 4.2a and b, V_{EE} and V_{CC} are the applied biasing voltages that establish currents in the directions shown in the common base configuration for the *n-p-n* and *p-n-p* transistors. By convention, these are defined to be opposite to the flow of electrons or in the direction of flow of holes. You must note the directions of the emitter and collector currents I_E and I_C, vis-à-vis the polarities of the biasing voltages V_{EE} and V_{CC}, respectively, in both types of transistors. So, in both cases, we have:

$$I_E = I_B + I_C \qquad (4.1a)$$

Since $I_B << I_C$, we can write:

$$I_E \approx I_C \qquad (4.1b)$$

The transistor symbols you have learnt in Unit 3 are arrived at for the common base configuration. You should practice drawing the circuits shown in Figs. 4.2a and b with the transistor symbols and the directions of emitter, base and collector currents for both *n-p-n* and *p-n-p* transistors. Try SAQ 1.

SAQ 1 – Common base configuration

Redraw Figs. 4.2a and b using the transistor symbols for *n-p-n* and *p-n-p* transistors.

We now illustrate the common emitter and common collector configurations using transistor symbols.

Common Emitter Configuration

The common emitter configuration is the most widely used configuration in transistor applications. It is shown in Fig. 4.3 for an *n-p-n* transistor. Note from Fig. 4.3 that in this configuration, the **emitter terminal is common** to both input and output circuits. Therefore, in this case, the input terminals are the base and emitter terminals, and the output terminals are the emitter and collector terminals. The base-emitter circuit is the input circuit and the collector-emitter circuit is the output circuit in this configuration.

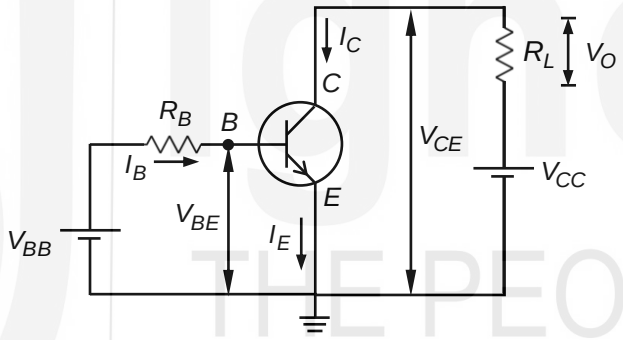


Fig. 4.3: **An *n-p-n* transistor in common emitter configuration.**

Note that the base, emitter and collector currents are shown in Fig. 4.3 with their directions as per the sign convention explained in Unit 3.

Note also that in the input circuit in the common emitter configuration, the applied voltage V_{BB} forward biases the emitter-base junction and the role of the resistance R_B is to limit the current in the input circuit. We can change the base current I_B by changing R_B or V_{BB}. When we change the base current, the collector current also changes as you have learnt in Sec. 3.2 of Unit 3. And a small base current controls a large collector current.

Now pay attention to the output circuit. Note that the applied voltage V_{CC} reverse biases the collector-base junction and again the current in the circuit is limited by the load resistance R_L. For proper operation of the transistor as explained in Unit 3, the applied voltage V_{CC} must reverse bias the collector-base junction.

The voltage V_{BE} represents the voltage between the base and emitter and the voltage V_{CE} represents the voltage between the collector and emitter. The voltage drop across the load resistance is the output voltage V_O. **Remember** that a double subscript notation has been used in the circuits of Figs. 4.2a, b and 4.3. These notations are used in transistor circuits.

We use the **same subscripts to represent source voltages**. For example, in Fig. 4.3, V_{BB} is the base voltage source and V_{CC}, the collector voltage source.

We use **different subscripts when the voltages are between two points**, for example, V_{BE} and V_{CE}. So, the voltage between the base and the emitter is denoted by V_{BE} and the voltage between the collector and the emitter by V_{CE}.

You should remember that the common emitter configuration also has low input resistance and high output resistance as explained in Sec. 3.2 of Unit 3.

You may now like to work out SAQ 2 to check if you have learnt the common emitter configuration well.

SAQ 2 – Common emitter configuration

Draw a diagram showing a *p-n-p* transistor in the common emitter configuration.

We now give a brief description of the common collector configuration.

Common Collector Configuration

When collector is common to the input and output circuits, we get the common collector configuration. The common collector configuration is used in applications that require impedance matching since it has high input impedance and low output impedance unlike the input and output resistances in the common base and common emitter configurations. Fig. 4.4 shows the common collector configuration for an *n-p-n* transistor.

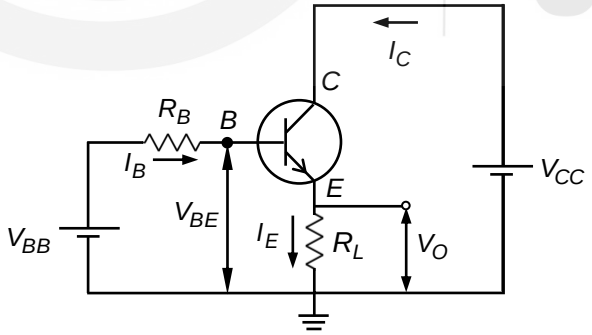


Fig. 4.4: **An *n-p-n* transistor in common collector configuration.**

Fig. 4.4 may look similar to Fig. 4.3 but actually it is not. You should note from Fig. 4.4 that the **collector is at ground potential** and the load resistor is connected between the emitter and ground.

For common collector configuration, the input voltage and current are V_{CB} and I_B, and the output voltage and current are V_{CE} and I_E.

You may like to pause for a while and revise what you have learnt about the common collector configuration. Attempt SAQ 3.

SAQ 3 – Common collector configuration

Draw the circuit showing common collector configuration for a *p-n-p* transistor.

Let us now discuss the common emitter configuration in some detail.

4.3 COMMON EMITTER CONFIGURATION

In this section, we will discuss the operation of a transistor in common emitter configuration. We will first draw the input and output characteristics for the transistor and explain them. Then we will define the parameters dc alpha and dc beta (current gain). Finally, we discuss the concepts of load line and quiescent operating point, which are important for the operation of a transistor.

4.3.1 Transistor Characteristics

We can plot three different *I-V* characteristics for each of the three configurations (CB, CE and CC) depending on the quantities involved: input, output and transfer characteristics.

However, we will be focusing on the CE configuration here. The currents and voltages involved in each of the three types of characteristics for the CE configuration are given in Table 4.1.

Table 4.1: Currents and voltages involved in different *I-V* characteristics of a BJT in common emitter (CE) configuration

Input characteristic	Output characteristic	Transfer characteristic
I_B versus V_{BE} for different values of V_{CE}	I_C versus V_{CE} for different values of I_B	I_C versus I_B for different values of V_{CE}

Of these, we will discuss the following two *I-V* characteristics in this section:

a) Input characteristics; and

b) Output characteristics.

Let us explain each of these characteristics for a bipolar junction transistor in the CE configuration.

Input characteristics

The input characteristics corresponding to the input circuit are a plot of the input current (the base current I_B on the *y*-axis) and the input voltage (base-

emitter voltage V_{BE} on the *x*-axis) for different values of the output voltage (collector-emitter voltage V_{CE}). The circuit to obtain this curve is the same as shown in Fig. 4.3. Fig. 4.5 shows the input characteristics for a few values of the collector-emitter voltage.

Note that the base current is in microamperes. Does this curve not look familiar to you? Recall the *I-V* characteristics of a forward biased *p-n* junction diode from Unit 2. The curve in Fig. 4.5 resembles those characteristics since the base-emitter junction is forward biased in a transistor. So, the base current flows only when the base-emitter voltage is equal to or greater than the knee voltage, which is 0.3 V for Ge transistor and 0.7 V for a silicon transistor.

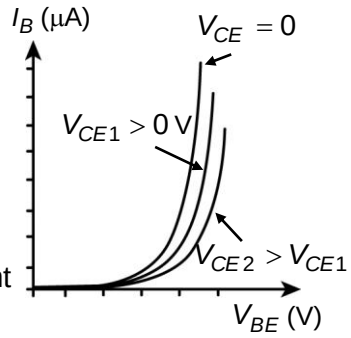


Fig. 4.5: **Input characteristics of a BJT in common emitter configuration.**

Output characteristics

The output characteristics of the BJT in CE configuration are shown in Fig. 4.6. Note that this is a plot of the output current (the collector current I_C on the *y*-axis) and the output voltage (collector-emitter voltage V_{CE} on the *x*-axis) for different values of the input current (base current I_B). For example, suppose we change V_{BB} to get a base current $I_B = 30\,\mu A$. For this fixed value of base current, we now vary V_{CC} in the circuit like the one shown in Fig. 4.3. Then we plot I_C for different values of V_{CE} to obtain a graph like the one shown in Fig. 4.6 for $I_B = 30\,\mu A$. You must remember that these curves are for a specific transistor. For other transistors, the values of I_C may be different for V_{CE} but the shape of the curves will be similar.

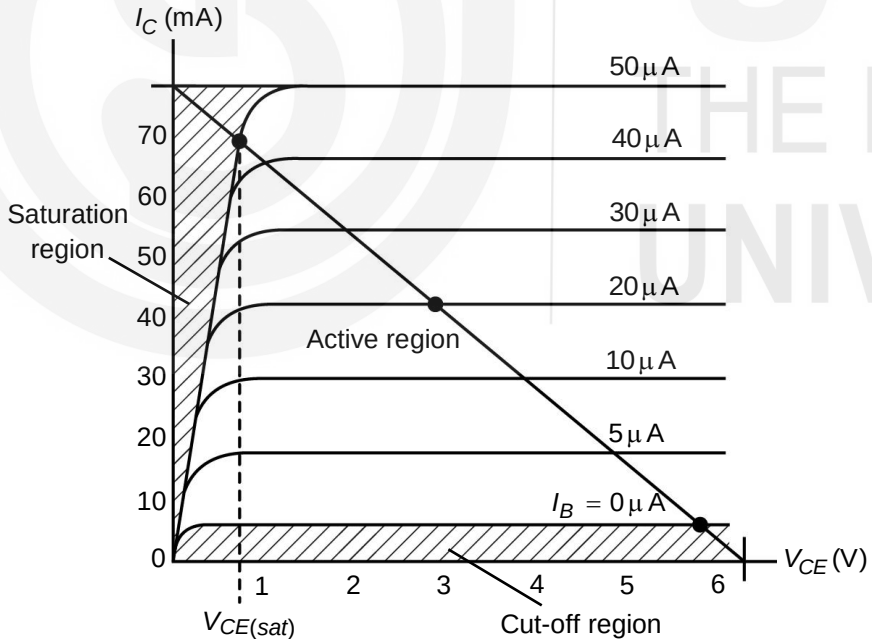


Fig. 4.6: **Output characteristics of a BJT in common emitter configuration.**

You should note from Fig. 4.6 that the base current is in microamperes, and the collector current is in milliamperes. Pay attention to the three regions marked on the output characteristics:

- Active Region
- Saturation Region
- Cut-off Region

All transistor output characteristics have an active region, a saturation region, and a cut-off region. We now discuss each of these three regions.

Active Region

Note from Fig. 4.6 that the active region of the output characteristic curve is the region to the right of $V_{CE} =$ a few tenths of a volt and above the curve for $I_B = 0$. Always remember that

The transistor operates in the active region when the base junction is forward biased and collector junction is reverse biased.

The transistor output current I_C responds most sensitively to the input signal in the active region. When V_{CE} increases from zero to a few-tenths of a volt, the collector current rises sharply in Fig. 4.6. When V_{CE} is above a few-tenths of a volt, the collector current becomes almost constant. Note that each collector current is many times greater than the corresponding base current. In different transistors, this ratio is different. The active region is the most important because amplification of signals is possible in the active region.

The active region is called the **linear** region because changes in the input signal produce proportional changes in the output signal. This means that the distortion in the output signal is low or negligible. Therefore, the **transistor has to operate in the active region when it is to be used as a voltage, current or power amplifier**. This means that the input, output voltages and input currents must be restricted to the values corresponding to the active region in the output characteristics.

Saturation Region

The **saturation region in the output characteristics** is the early sloping part of the characteristic curve, where V_{CE} is less than a few-tenths of a volt. In this region, the collector diode has insufficient positive voltage to collect all the free electrons injected into the base. Therefore, in this region, the base current is larger than normal. Remember that

The transistor operates in the **saturation region** when **both base-emitter and collector-emitter junction are forward biased by at least the knee voltage.**

So, in the saturation region, V_{CE} is a small negative voltage. Note from Fig. 4.6 that the saturation region is very close to the zero-voltage axis, where all the curves merge and fall rapidly toward the origin. When the collector-emitter diode is forward biased, there is an exponential increase in the collector current for small changes in collector-emitter voltage, as V_{CE} is increased to the value 0 V. You should note from Fig. 4.6 that the onset of

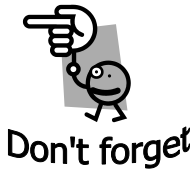
saturation takes place at the knee of the curves corresponding to different values of the base current.

Cut-off Region

The **cut-off region in the output characteristics** is the region below $I_B = 0$.

You must remember that

> The transistor operates in the **cut-off region when both the emitter and collector junctions are reverse biased.**



Note from Fig. 4.6 that even when the base current is zero, a small collector current still flows in the output circuit. This current is usually of the order of nanoamperes and cannot be seen in these curves. The region below $I_B = 0$ has been drawn larger than usual in Fig. 4.6 so that you may see it. This extremely small collector current is called **collector cut-off current** and denoted by I_{CEO}. Thus, remember that

> The cut-off current for the common emitter configuration is the collector current $I_C = I_{CEO}$ defined by the condition $I_B = 0$.

You may well ask: Why is there a finite, even if small, collector current in the cut-off region? This current is due to the minority charge carriers in the reverse biased collector junction. It is also due to surface leakage current which does not flow through the junction. Rather it flows around the junction and across the transistor surfaces. It is proportional to the voltage across the junction.

However, in a well-designed circuit, the collector cut-off current is small enough to ignore. So, while designing a transistor amplifier, we should take care that it does not operate in cut-off region (below $I_B = 0$) so that the output signal is undistorted.

When a transistor has to be operated as a switch, for example, in logic circuits in a computer, it has two points of operation: one in the saturation region and one in the cut-off region. This is because in the saturation region, the transistor allows current to flow and in the cut-off region, the current in the circuit is extremely small. Ideally, we must have $I_C = 0$ in the cut-off region for switching applications. But practically, since I_{CEO} is very small, cut-off does exist for switching purposes when $I_B = 0$ or $I_C = I_{CEO}$.

Next, we define current gains in a BJT. Before that you should solve SAQ 4 to fix the ideas of this section in your mind.

SAQ 4 – BJT in CE configuration

Write the relevant currents and voltages for the a) input and b) output characteristics of an ideal BJT in CE configuration. What are the values of V_{CE} and I_B in the saturation and cut-off regions of the output characteristics?

4.3.2 DC Alpha (α_{dc}) and DC beta (β_{dc})

In this section, we will define the parameters dc alpha (α_{dc}) and dc beta (β_{dc}) for the transistor in common emitter configuration.

By definition, **the dc alpha** is the **ratio of the dc collector current to the dc emitter current**:

$$\alpha_{dc} = \frac{I_C}{I_E} \tag{4.2a}$$

Recall from Eq. (3.1a) of Unit 3 that the collector current is almost equal to the emitter current. Therefore, the value α_{dc} is slightly less than 1. For example, the typical value of α_{dc} is more than 0.99 for a low power transistor. It is more than 0.95 even for high-power transistors.

By definition, **the dc beta** is the **ratio of the dc collector current to the dc base current**:

$$\beta_{dc} = \frac{I_C}{I_B} \tag{4.2b}$$

β_{dc} is also referred to as the **current gain**. This is because a small base current controls a much larger collector current.

Current gain β_{dc} typically ranges from 100 to 300 for low power transistors (operating at powers less than 1 W). For transistors operating at powers more than 1 W, that is, high power transistors, β_{dc} lies usually between 20 and 100. The current gain is a useful parameter for transistor applications as amplifiers.

An important point to note about the dc current gain of a transistor is that it depends on three parameters: the **transistor** itself, the **collector current**, and the **temperature**. For example, a different transistor will usually have a different dc current gain. Also, the dc current gain changes if the collector current or temperature changes. The range of variation in β_{dc} can be large and affects the design of a transistor amplifier.

From the definitions of α_{dc} and β_{dc}, we can establish a relation between them using Eq. (4.1a) as follows:

$$\alpha_{dc} = \frac{I_C}{I_E} = \frac{I_C}{I_C + I_B}$$

or $$I_C = \alpha_{dc}(I_C + I_B)$$

or $$I_C = \frac{\alpha_{dc}}{1-\alpha_{dc}} I_B$$

or $$\beta_{dc} = \frac{I_C}{I_B} = \frac{\alpha_{dc}}{1-\alpha_{dc}} \tag{4.3a}$$

From Eq. (4.3a), we get

$$\alpha_{dc} = \frac{\beta_{dc}}{1 + \beta_{dc}} \qquad (4.3b)$$

You may like to calculate the dc alpha and dc beta for a given transistor. Solve SAQ 5.

SAQ 5 – DC alpha and dc beta for a transistor

a) A transistor has a dc current gain of 200. If the base current is 10 μA, calculate the value of the collector current. What is the value of dc alpha?

b) Calculate the dc alpha and dc beta given that $I_C = 1.0$ mA and $I_B = 2.5\ \mu A$.

Lastly, in this section, we discuss two important concepts used in transistor amplifiers, namely, the load line and the quiescent operating point.

4.3.3 Load Line and Quiescent Operating Point

You have learnt in the previous section that the transistor should operate in the active region for its application as an amplifier. We now discuss the parameters that are important in this respect.

Refer to Fig. 4.7, which is Fig. 4.3 repeated here for ready reference.

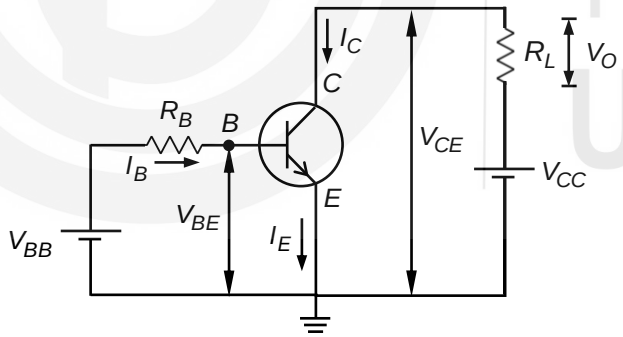


Fig. 4.7: **An** *n-p-n* **transistor in common emitter configuration.**

Let us apply Kirchhoff's voltage law to the output circuit in Fig. 4.7. We get:

$$V_{CE} = V_{CC} - I_C R_L \qquad (4.4a)$$

We can rewrite Eq. (4.4a) as:

$$I_C = \frac{V_{CC} - V_{CE}}{R_L} \qquad (4.4b)$$

Note that Eq. (4.4b) is a straight line between the variables I_C and V_{CE} with slope $(-\frac{1}{R_L})$ and intercept $\frac{V_{CC}}{R_L}$.

Also note that there is no signal applied in the input circuit and the voltages applied in both input and output circuits are dc voltages.

Fig. 4.8a shows the plot of the straight line given by Eq. (4.4b) on the output characteristic curve of the transistor in CE configuration. Note that from Eq. (4.4a), we can specify two points on the straight line as follows:

$$I_C = \frac{V_{CC}}{R_L} \quad \text{for} \quad V_{CE} = 0,$$

and $V_{CE} = V_{CC}$ for $I_C = 0$

When we join these two points on the output characteristics, we get the straight line given by Eq. (4.4a or b) superimposed on the curve. This line is called the **load line** (see Fig. 4.8a). It is called the load line because it is defined by the value of the load resistance in the output circuit. You have seen that the slope of the load line is determined by the value of the load resistance.

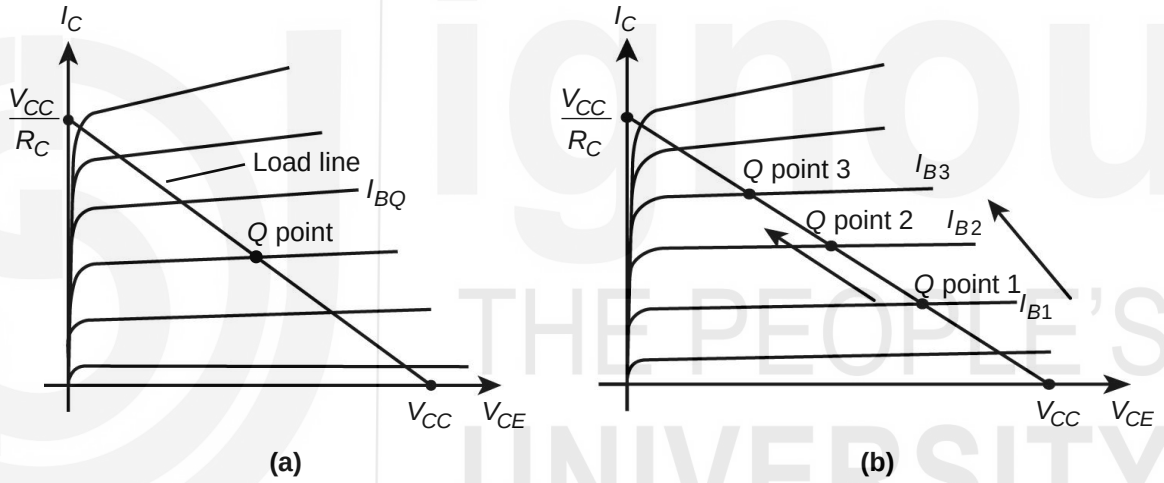


Fig. 4.8: **a) Load line and *Q* point; b) shifting of the *Q* point.**

You may like to know: **Why is the load line important in transistor amplifiers**? It is because the load line defines the operating conditions for the transistor in the active region for a given load resistance. You will understand this point better when we discuss the operating point. Let us do that now.

The **operating point** of a transistor in CE configuration is defined by the output voltage V_{CE}, output current I_C and input current I_B, for no input signal. It is also called the **quiescent point** or the *Q* **point.** Quiescent means quiet, resting, inactive or still.

For amplification of a small signal applied to the input circuit, the operating point is usually set up in the middle of the load line. It is shown in Figs. 4.8a and b.

Let us take an example showing a typical calculation for determining the operating point.

EXAMPLE 4.1 : DETERMINING *Q* POINT

Determine the *Q* point for a BJT in CE configuration given that $V_{CC} = 15\,\text{V}$, $\beta_{dc} = 100$, $I_B = 20\ \mu\text{A}$ and $R_L = 5.0\ \Omega$.

SOLUTION ■ We use Eqs. (4.2b and 4.4a) for the calculations.

From Eq. (4.2b), $I_C = \beta_{dc} \times I_B = 100 \times 20\mu\text{A} = 2\text{mA}$

The voltage drop across R_L is:

$$I_C R_L = 2\ \text{mA} \times 5.0\Omega = 10\,\text{V}$$

Thus, from Eq. (4.4a),

$$V_{CE} = 15\ \text{V} - 10\ \text{V} = 5\ \text{V}$$

The values $V_{CE} = 5$ V, $I_C = 2\text{mA}$ for $I_B = 20\ \mu\text{A}$ specify the *Q* **point** on the load line.

From the calculations given in Example 4.1, you would appreciate that the operating point depends on the dc current gain of the transistor, and is very sensitive to changes in it. For example, if the dc current gain were to change by a factor of even 2 or 3, the values of V_{CE} and I_C for a given I_B could be such that the *Q* point moved to the saturation region or cut-off. Then the transistor would not work as an amplifier. You will appreciate this point better when you study transistor amplifiers in the next block.

You may like to check if you have grasped these concepts well. Solve SAQ 6.

SAQ 6 – Load line and operating point

Determine the operating point for a transistor in CE configuration for

(i) $V_{CC} = 30$ V and $I_B = 30\mu\text{A}$

(ii) $V_{CC} = 10$ V and $I_B = 10\mu\text{A}$

given that $R_L = 5.0\text{k}\Omega$ and $\beta_{dc} = 100$.

So, now the question before us is: How is the transistor to be biased so that we get the desired load line and operating point for its applications? You will learn about various transistor biasing methods in Sec. 4.4.

4.4 TRANSISTOR BIASING METHODS

For optimum performance of the transistor, we need to apply appropriate external voltages of correct polarity in its input and output circuits. You have learnt in Sec. 4.3 that the operation of a transistor depends on its base current, collector current and input and output voltages. Thus, you must always remember that

Transistor biasing is the process of applying appropriate dc voltages to a transistor.

The voltages applied should be such that the **transistor functions as a linear device even when a small change occurs in the applied voltage**. You have learnt in Sec. 4.3.3 that the operating voltages and currents in the output and input circuits define the **operating point** or "**quiescent**" (*Q*) point of the transistor.

As you know, the *Q* point for a transistor in common emitter configuration is specified by the collector-emitter voltage V_{CE} and the collector current I_C. You should note that **for the transistor to operate linearly, the transistor biasing should be such that the *Q* point remains stable**. Also, the *Q* **point should not change with changes in current gain**.

We describe briefly three biasing methods for transistors in this section: **Fixed bias**, **self-bias** and **universal bias** for the common emitter configuration.

Fixed bias

Fixed bias is also called the base bias and in this kind of biasing, the **base current** (I_B) **in the transistor is kept constant** for given values of V_{CC}. This is done by connecting a constant resistor R_B having high resistance in the input circuit (see Fig. 4.9). By selecting a proper value of R_B, we can set the desired values of the base current and the collector current in the input and output circuits.

However, usually the fixed bias method using a single resistor is not used to bias a transistor for linear operation. This is because the biasing voltages and currents do not remain stable during transistor operation and the *Q* point is unstable.

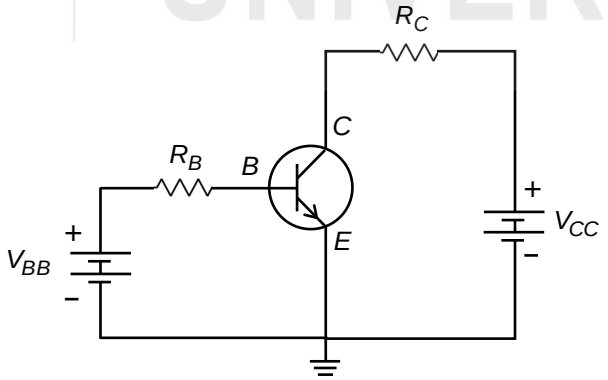


Fig. 4.9: **Fixed bias for an *n-p-n* transistor in common emitter configuration.**

Self-bias

In the **self-bias** method, also called the **collector feedback, the base resistor is connected to the collector rather than the supply voltage** as shown in Fig. 4.10. In this case, if β_{dc} increases due to some reason such as variation in temperature, say, then I_C will increase. This will result in larger voltage drop across R_C. Hence, the collector voltage will reduce. In effect,

I_C will reduce, correcting the increase caused by variation in β_{dc}. This type of biasing is somewhat more effective in improving the stability of the operating point. However, the circuit is still sensitive to changes in biasing voltages and the *Q* point is not that stable although it is better than a fixed bias circuit.

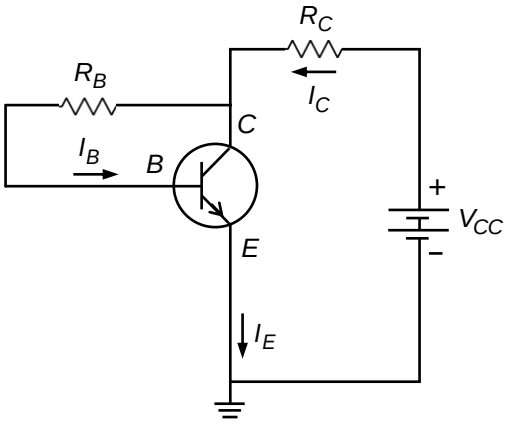


Fig. 4.10: **Self- bias for** an *n-p-n* **transistor in common emitter configuration.**

Universal bias

Universal bias is the biasing method most widely used for linear operation of the transistor (see Fig. 4.11). It is also known as the voltage divider circuit because the resistances R_1 and R_2 in Fig. 4.11 form the voltage divider. In this biasing method, only one battery or dc power source is required. It uses four resistors and provides a stable operating point.

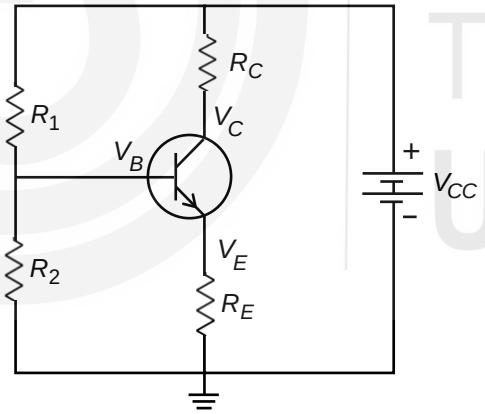


Fig. 4.11: **Universal bias for** an *n-p-n* **transistor in common emitter configuration.**

Note from Fig. 4.11 that the voltage across R_2 forward biases the emitter diode. In this biasing method, the *Q* point is independent of the dc beta gain. Let us explain how. If the current gain (β_{dc}) increases, the collector current (I_C) increases. This results in an increase in the emitter current I_E and, hence, the voltage drop across R_E increases. From the input circuit in Fig. 4.11, you can see that resistors R_1 and R_2 form a voltage divider across the dc power supply V_{CC}, and the current I_1 through these resistors is usually selected to be about one-tenth of the collector current. Since the base current is small, the current through R_2 can also be taken as $\sim I_1$. So, the voltage across R_2 is

$$V_{R_2} = I_1 R_2 = \frac{V_{CC} R_2}{R_1 + R_2} \qquad (4.5a)$$

Applying Kirchhoff's law to the input circuit, we get:

$$V_{R_2} = V_{BE} + V_{R_E} = V_{BE} + I_E R_E \quad (4.5b)$$

An increase in $I_E R_E$ results in a decrease in the base-emitter voltage V_{BE}, and, therefore, the base current is reduced. The reduced base current results in reduced collector current, which offsets the original increase in β_{dc}. We can depict this process by the following equations:

$$I_C \uparrow \Rightarrow I_E \uparrow \Rightarrow (I_E R_E) \uparrow \quad (4.6a)$$

Since $V_{R_2} = V_{BE} + I_E R_E = \text{constant}$, therefore,

$$V_{BE} \downarrow \Rightarrow I_B \downarrow \Rightarrow I_C \downarrow \quad (4.6b)$$

In Eqs. (4.6a and b), the arrow pointing upwards (indicated beside a physical quantity) shows that the quantity is increasing and the arrow pointing downwards shows that it is decreasing. Using Eq. (4.5a) in Eq. (4.5b) and the fact that $I_C \approx I_E$, we can write:

$$V_{R_2} = V_{BE} + I_C R_E = \text{constant} \quad (4.7)$$

Since the left-hand side of Eq. (4.7) is constant, a change in I_C causes V_{BE} to change in a direction so as to bring I_C back to its original value as depicted in Eq. (4.6b). Thus, the *Q* point is kept stable in the universal bias. Let us illustrate this with the help of an example.

EXAMPLE 4.2 : *Q* POINT IN UNIVERSAL BIAS

Calculate the dc voltages (base voltage, emitter voltage, collector voltage and collector to emitter voltage) and collector and emitter currents for the universal bias circuit of Fig. 4.11 given that $R_1 = 40\,\text{k}\Omega$, $R_2 = 5.0\,\text{k}\Omega$, $R_C = 5.0\,\text{k}\Omega$, $V_{CC} = 12\,\text{V}$ and $R_E = 1.0\,\text{k}\Omega$. Take $V_{BE} = 0.3\,\text{V}$ for the transistor used. Hence, locate the *Q* point.

SOLUTION ■ We use Eqs. (4.5a and 4.7), with Kirchhoff's law for these calculations.

The base voltage is given by: $V_B = V_{R_2} = \dfrac{V_{CC} R_2}{R_1 + R_2}$ (i)

Substituting the values $R_1 = 40\,\text{k}\Omega$, $R_2 = 5.0\,\text{k}\Omega$ and $V_{CC} = 12\,\text{V}$ in Eq. (i), we get

$$V_B = \frac{12\,\text{V} \times 5.0\,\text{k}\Omega}{(40 + 5.0)\,\text{k}\Omega} = 1.3\,\text{V}$$

$$\therefore \quad V_E = V_B - V_{BE} = (1.3 - 0.3)\,\text{V} = 1.0\,\text{V}$$

and $$I_E = \frac{V_E}{R_E} = \frac{1.0\,\text{V}}{1.0\,\text{k}\Omega} = 1.0\ \text{mA}$$

$$\Rightarrow \quad I_C \approx I_E = 1.0\ \text{mA}$$

Applying Kirchhoff's law to the output circuit, we get the collector voltage:

$$V_C = V_{CC} - I_C R_C = 12.0\,\text{V} - (1.0\,\text{mA} \times 5.0\,\text{k}\Omega) = 7.0\,\text{V}$$

Thus, the collector-emitter voltage is:

$$V_{CE} = V_C - V_E = 6.0\,\text{V}$$

Therefore, the *Q* point is located at (6.0 V, 1.0 mA) for this transistor.

We end this section with an SAQ for you.

SAQ 7 – Universal bias

For the circuit given in Fig. 4.11, state if the following quantities will increase or decrease if R_1 is increased by 50%?

a) V_{R_2} will…….b) V_{RE} will……. c) I_C will……d) V_C will…….

With this discussion on transistor biasing methods, we end this unit. Let us now summarise what you have learnt in this unit.

4.5 SUMMARY

Concept	Description
CB, CE, CC configurations of a BJT	■ There are three possible configurations in a transistor in which one of the three terminals is common to both input and output ports: • **Common base configuration**: When the **base** terminal of the transistor is common to input and output circuits. • **Common emitter configuration**: When the **emitter** terminal of the transistor is common to input and output circuits. • **Common collector configuration**: When the **collector** terminal of the transistor is common to input and output circuits.
Input and output characteristics in common emitter configuration	■ In common emitter configuration of the BJT, the **input characteristics** are a plot of the input current (the base current I_B on the *y*-axis) and the input voltage (base-emitter voltage V_{BE} on the *x*-axis) for different values of the output voltage (collector-emitter voltage V_{CE}). The **output characteristics** are a plot of the output current (the collector current I_C on the *y*-axis) and the output voltage (collector-emitter voltage V_{CE} on the *x*-axis) for different values of the input current (base current I_B).
Active, saturation and cut-off regions in the output characteristics	■ All transistors have an **active region**, a **saturation region**, and **a cut-off region** in the output characteristics: • The **active region** of the output characteristic curve is the region to the right of $V_{CE} =$ a few tenths of a volt and above the curve for $I_B = 0$. **The transistor operates in the active region when the base junction is forward biased and collector junction is reverse biased.**

- The transistor operates in the **saturation region** when **both base-emitter and collector-emitter junction are forward biased by at least the knee voltage.**
- The transistor operates in the **cut-off region when both the emitter and collector junctions are reverse biased.** The **cut-off current** for the common emitter configuration is the collector current $I_C = I_{CEO}$ defined by the condition $I_B = 0$.

DC current gains for the transistor in common emitter configuration

■ The **dc alpha is the ratio of the dc collector current to the dc emitter current:**

$$\alpha_{dc} = \frac{I_C}{I_E}$$

The **dc beta is the ratio of the dc collector current to the dc base current**:

$$\beta_{dc} = \frac{I_C}{I_B}$$

It is also referred to as the **current gain** because a small base current controls a much larger collector current. The relations between dc alpha and dc beta are given by:

$$\beta_{dc} = \frac{I_C}{I_B} = \frac{\alpha_{dc}}{1-\alpha_{dc}} \quad \text{and} \quad \alpha_{dc} = \frac{\beta_{dc}}{1+\beta_{dc}}$$

Load line and Q point

■ The load line defines the operating conditions for the transistor in the active region for a given load resistance. **The straight line on the output characteristic curve of the transistor in CE configuration joining the points**

$$I_C = \frac{V_{CC}}{R_L} \quad \text{for} \quad V_{CE} = 0,$$

and $V_{CE} = V_{CC}$ for $I_C = 0$

is the **load line.** Its slope is determined by the value of the load resistance in the output circuit. The **operating point** or **quiescent point** (*Q* point) of a transistor in CE configuration is defined by the output voltage V_{CE}, output current I_C and input current I_B, for no input signal. For amplification of a small signal applied to the input circuit, the operating point is usually set up in the middle of the load line and has to be stable.

Transistor biasing methods

■ Appropriate external voltages of correct polarity need to be applied in the input and output circuits of the transistor for optimum performance. Transistor biasing is the process of applying appropriate dc voltages to a transistor. Three biasing methods can be used:

- **Fixed bias**: In fixed bias, also called the base bias, the **base current** (I_B) **in the transistor is kept constant** for given values of V_{CC}. In the fixed bias method, the biasing voltages and currents do not remain stable during transistor operation and the *Q* point is unstable.

- **Self-bias**: In the **self-bias** method, also called the **collector feedback bias**, the base resistor is connected to the collector rather than the supply voltage. This type of biasing improves the stability of the operating point. However, the circuit is still sensitive to changes in biasing voltages and the *Q* point is not that stable although it is better than a fixed bias circuit.
- **Universal bias**: Universal bias is the most widely used biasing method for linear operation of the transistor. It is also known as the **voltage divider circuit** because two resistances form the voltage divider in the base biasing circuit. In this biasing method, only one battery or dc power source is required. It uses four resistors and provides a stable operating point.

4.6 TERMINAL QUESTIONS

1. Using the characteristics of Fig. 4.6, determine the
   a) collector current when the base current is 20 μA and collector-emitter voltage is 5 V;
   b) the approximate magnitude of the ratio of the collector current to base current of 20 μA when collector-emitter voltage is 3 V.
2. Determine the dc beta value for part a) of Terminal Question 1.
3. It is given that the current gain is 200, V_{CC} is 20 V and the base current is 40 μA for the CE configuration of a BJT. What is the value of the load resistor?
4. Calculate the base voltage, emitter voltage, collector voltage, collector to emitter voltage, and collector and emitter currents for the universal bias circuit of Fig. 4.11 given that $R_1 = 45\,\text{k}\Omega$, $R_2 = 5.0\,\text{k}\Omega$, $R_C = 5.0\,\text{k}\Omega$, $V_{CC} = 15\,\text{V}$ and $R_E = 1.0\,\text{k}\Omega$. Take $V_{BE} = 0.3\,\text{V}$ and hence, locate the *Q* point.

4.7 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. See Figs. 4.12a and b.

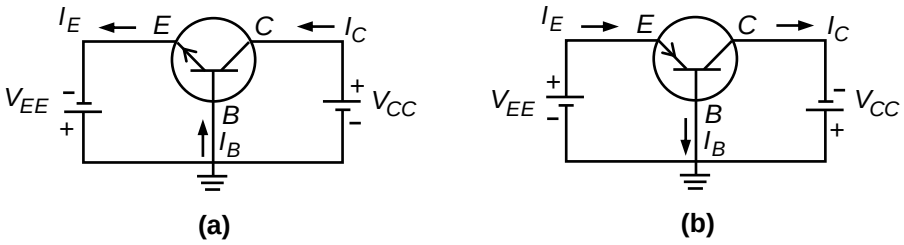


Fig. 4.12: **a) An** *n-p-n* **transistor; b) a** *p-n-p* **transistor in common base configuration.**

2. See Fig. 4.13.

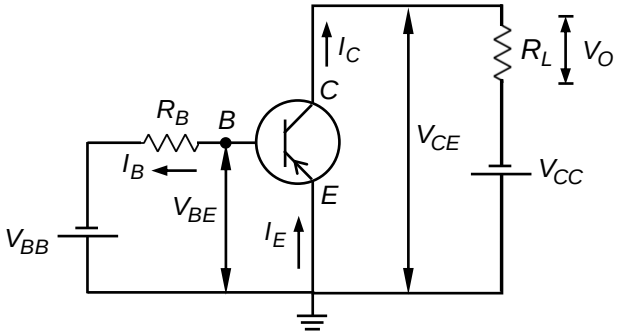


Fig. 4.13: **An *p-n-p* transistor in common emitter configuration.**

3. See Fig. 4.14.

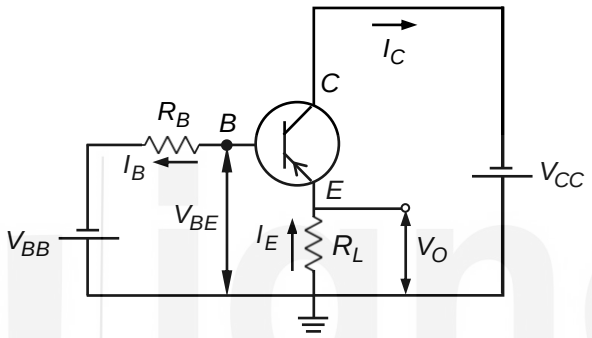


Fig. 4.14: **A *p-n-p* transistor in common collector configuration.**

4. For input characteristics: I_B, V_{BE}, V_{CE}.

   For output characteristics: I_C, V_{CE}, I_B.

   For saturation region, $V_{CE} <$ few tenths of 1 V, $I_B = 0$ and above.

   For cut-off region, $V_{CE} = V_{CC}$, below $I_B = 0$.

5. a) From Eq. (4.2b), $I_C = \beta_{dc} I_B$

$$= 200 \times 10\,\mu\text{A} = 2\ \text{mA}$$

From Eq. (4.3b), $\alpha_{dc} = \dfrac{\beta_{dc}}{1+\beta_{dc}}$

$$= \frac{200}{201} = 0.995$$

b) From Eq. (4.2b), $\beta_{dc} = \dfrac{I_C}{I_B}$

$$= \frac{1.0\,\text{mA}}{2.5\,\mu\text{A}} = 400$$

From Eq. (4.3b), $\alpha_{dc} = \dfrac{\beta_{dc}}{1+\beta_{dc}}$

$$= \frac{400}{401} = 0.998$$

6. i) From Eq. (4.2b), $I_C = \beta_{dc} I_B$

$= 100 \times 30\,\mu A = 3$ mA

whence $I_C R_L = 15\,V$

Therefore, $V_{CE} = V_{CC} - I_C R_L$

$= 30\,V - 15\,V = 15\,V$

ii) From Eq. (4.2b), $I_C = \beta_{dc} I_B$

$= 100 \times 10\,\mu A = 1\,mA$

whence $I_C R_L = 5\,V$

Therefore, $V_{CE} = V_{CC} - I_C R_L$

$= 10\,V - 5\,V = 5\,V$

7. a) decrease;

b) decrease;

c) decrease;

d) increase.

Terminal Questions

1. From the characteristics shown in Fig. 4.6,

a) collector current is approximately 40 mA or 41.2 mA;

b) at collector-emitter voltage of 3 V, the approximate magnitude of the ratio of collector current (~ 40 mA) to base current (20 μA) is 2000.

2. dc beta is the ratio of the collector current to base current and equals 2060.

3. $I_C = \beta_{dc} I_B = 200 \times 40\,\mu A = 8.0$ mA

$$R_L = \frac{V_{CC}}{I_C} = \frac{20\,V}{8.0\,mA} = 2.5\,k\Omega$$

4. Substituting the values $R_1 = 45\,k\Omega$, $R_2 = 5.0\,k\Omega$ and $V_{CC} = 15\,V$, the base voltage is given by:

$$V_B = V_{R_2} = \frac{V_{CC} R_2}{R_1 + R_2}$$

$$= \frac{15\,V \times 5.0\,k\Omega}{(45 + 5.0)\,k\Omega} = 1.5\,V$$

$\therefore \quad V_E = V_B - V_{BE}$

$= (1.5 - 0.3)\,V = 1.2\,V$

and $I_E = \frac{V_E}{R_E} = \frac{1.2\,\text{V}}{1.0\,\text{k}\Omega} = 1.2\,\text{mA}$

$\therefore \quad I_C \approx I_E = 1.2\,\text{mA}$

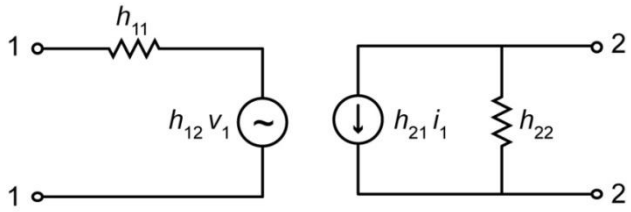
Applying Kirchhoff's law to the output circuit, we get the collector voltage as:

$$V_C = V_{CC} - I_C R_C = 15\,\text{V} - (1.2\,\text{mA} \times 5.0\,\text{k}\Omega) = 9.0\,\text{V}$$

Thus, the collector-emitter voltage is:

$$V_{CE} = V_C - V_E = 7.8\,\text{V}$$

Therefore, the *Q* point is located at (7.8 V, 1.2 mA) for this transistor.



UNIT 5

Analysis of transistor circuit becomes easier when it is represented in terms of hybrid parameters of the device. In this unit you will learn about how to analyse transistor circuits using these parameters.

TRANSISTOR CIRCUIT ANALYSIS

Structure

STUDY GUIDE

So far you have learnt about different semiconductor devices. When we want to use them in various applications, it is important to know their behaviour in advance under various current and voltage conditions. An equivalent circuit of a device in the form of basic components, like resistors, capacitors, voltage/current sources is very handy for this purpose.

In this unit, you will learn about the equivalent circuit of the bipolar junction transistor and express its circuit in the form of hybrid-parameters (*h*-parameters in short).

You will require the basic knowledge of calculus and first order differential equations for studying this unit. You should attempt all the SAQs and Terminal Questions given in this unit on your own, before turning to the answers and hints provided at the end of the unit.

"My first toy was a box of transistors."

Ann Makosinski

5.1 INTRODUCTION

In the earlier units you have learnt about different electronic devices. When you use them for some applications and you want to analyse the circuit, it is possible to do so by replacing these devices by some equivalent circuit. Therefore, in this unit, we will discuss how this can be done for a bipolar junction transistor.

A transistor equivalent circuit is basically a circuit consisting of ideal voltage and/or current "source" or "generators" and passive components (*R*, *L* and *C*), which acts electrically exactly like the transistor. In other words, the transistor can be replaced by an appropriate collection of generators and passive components of the equivalent circuit. The advantage of the equivalent circuit is that we can predict the transistor's exact behaviour (gain, etc.) by applying the basic circuit laws (Kirchhoff voltage and current laws) for the various loops and junctions and solving for the desired quantities by using only algebra and Ohm's law.

In an equivalent circuit, we consider the input and output currents and voltages as variables. We can write linear equations relating these variables. The coefficients in these equations are called the hybrid or *h*-parameters. We will discuss these parameters in detail in Sec. 5.2.

You have learnt in Unit 4 that the transistor is a three-terminal device, and we can apply supply voltages to these terminals by treating one of them as the common terminal. So, we have three basic configurations of transistor biasing, namely, CB, CC and CE configurations. In each of these configurations, one side is treated as input part and another as output part.

In Sec. 5.3, you will study about the representation of a transistor in its equivalent circuit form. You will also learn how to express various transistor parameters (like gains) in terms of *h*-parameters. In Sec. 5.4, you will learn how to analyse the common emitter (CE) configuration of the transistor and obtain expressions for voltage gain, current gain, input and output impedances in terms of *h*-parameters.

You will also apply a similar process for analysing common base (CB) and common collector (CC) configurations in Secs. 5.5 and 5.6, respectively.

Based on the relations obtained for different configurations, you will be able to compare the transistor configurations on the basis of their properties like gains and impedances and choose a suitable configuration for the intended application.

Expected Learning Outcomes

After studying this unit, you should be able to:

- obtain *h*-parameters for a 4-terminal network;
- represent the bipolar junction transistor in its equivalent circuit form;
- define all the *h*-parameters for a transistor;

- derive the expressions for voltage gain, current gain, input impedance and output impedance in terms of *h*-parameters for common-emitter configuration of the transistor; and
- obtain the expressions for the gains and impedances for common base and common collector configurations of the transistor.

5.2 *h*-PARAMETERS

To obtain the circuit parameters, let us begin by representing the circuit in a simple form. Circuit analysis becomes straightforward if we treat the circuit as a black box with two ends or ports, i.e., 4 terminals corresponding to input and output sides, as shown in Fig. 5.1.

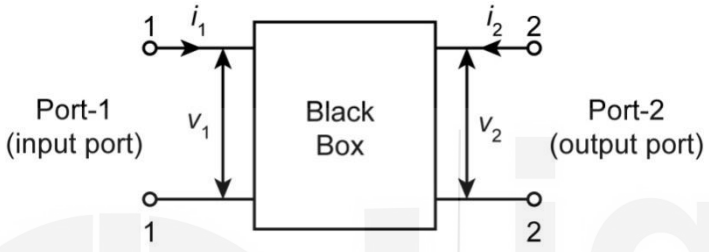


Fig. 5.1: **Circuit as a black box with 2 ports or 4-terminal.**

Here the voltage at the input terminal is v_1, and current flowing into the circuit is i_1 (positive). The voltage across output terminals is v_2 and the current flowing into the circuit is i_2 (positive). The relations between i_1, v_1, i_2 and v_2 can be represented by linear equations.

The current flowing into the circuit is considered to be positive, while the one coming out is taken to be negative.

If we consider the currents to be independent variables and voltages to be dependent variables, then we get a circuit as shown in Fig. 5.2a.

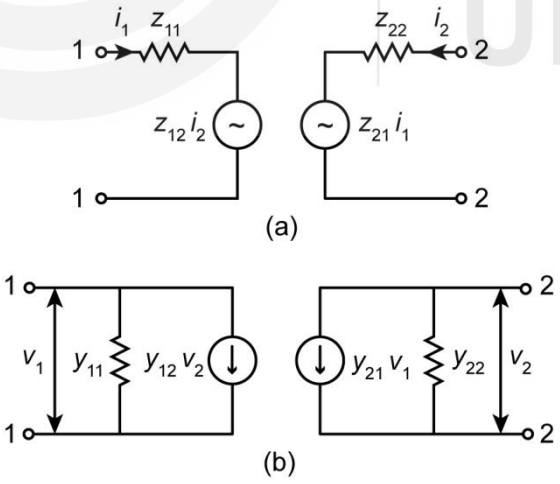


Fig. 5.2: **a) *z*- and b) *y*-parameter circuits.**

We can write the input and output voltages in terms of input and output currents as given below:

$$v_1 = z_{11} i_1 + z_{12} i_2 \qquad (5.1a)$$

$$v_2 = z_{21} i_1 + z_{22} i_2 \qquad (5.1b)$$

In these equations, the coefficients are impedances and are indicated by *z-parameters*.

You are aware that the impedances are indicated by *z* while the admittances are indicated by symbol *y*.

Another way to treat the equivalent circuit could be by considering the voltages to be independent variables, while currents to be dependent variables. In this case we can depict the circuit as shown in Fig. 5.2b. Here you will observe that the current sources are connected in parallel to admittances. The equations in this case are:

$$i_1 = y_{11}v_1 + y_{12}v_2 \quad (5.2a)$$

$$i_2 = y_{21}v_1 + y_{22}v_2 \quad (5.2b)$$

The coefficients in the equations are admittances represented by *y*-parameters. But a more appropriate combination of variables for analysing the transistor circuit is to consider the input current (i_1) and output voltage (v_2) as independent variables and input voltage (v_1) and output current (i_2) as dependent variables. Then we can write following relations:

$$v_1 = h_{11}i_1 + h_{12}v_2 \quad (5.3)$$

and

$$i_2 = h_{21}i_1 + h_{22}v_2 \quad (5.4)$$

The coefficients h_{11}, h_{12}, h_{21} and h_{22} are called the **hybrid parameters** or ***h*-parameters**. It is possible to obtain different characteristics of transistor circuits like current or voltage gains and input or output impedances using these parameters.

Fig. 5.3 shows the circuit with *h*-parameters.

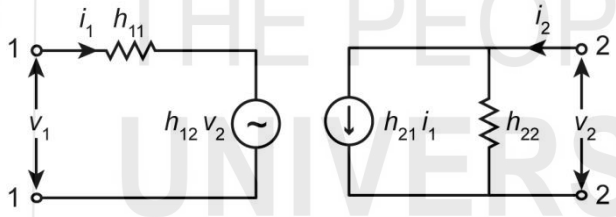


Fig. 5.3: 4-terminal network with *h*-parameters.

Let us now learn what the different *h*-parameters represent.

5.2.1 Interpretation of *h*-Parameters

Let us understand the physical significance of the *h*-parameters.

Refer to Fig. 5.3. To understand the meaning of h_{11} and h_{21}, let us assume that the output terminals (2-2) are short circuited as shown in Fig. 5.4, then the output voltage v_2 will be zero.

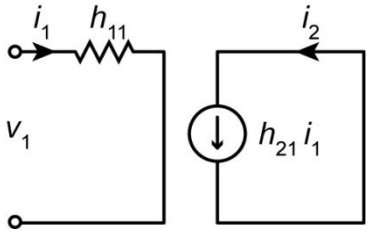


Fig. 5.4: ***h*-parameter circuit with short circuited output.**

Hence, from Eq. (5.3) and (5.4), we get

$$v_1 = h_{11} i_1 \qquad (5.5)$$

and $$i_2 = h_{21} i_1 \qquad (5.6)$$

a) Input impedance (h_{11})

From Eq. (5.5), we can write

$$h_{11} = \frac{v_1}{I_1} \quad \text{(with output short circuited)} \qquad (5.7)$$

It is clear that h_{11} is a quantity obtained by dividing a voltage term by a current term; that is, it represents an impedance and since both voltage and current are corresponding to the input part, we call it **input impedance** with short-circuited output.

b) Current gain (h_{21})

From Eq. (5.6), we can write

$$h_{21} = \frac{i_2}{I_1} \quad \text{(with output short-circuited)} \qquad (5.8)$$

Since this is a ratio of output and input currents, h_{21} represents the current gain of the circuit, when output is short-circuited.

After learning about the input impedance and current gain, let us find out the significance of remaining two *h*-parameters. For this purpose, we take the input terminals (1-1) in Fig. 5.3 to be open as shown in Fig. 5.5. With this, we get zero current entering the circuit from input side and so $i_1 = 0$. Hence, Eqs. (5.3) and (5.4) take the following form:

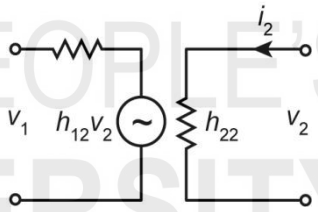


Fig. 5.5: ***h*-parameters circuit with open input terminals ($i_1 = 0$).**

$$v_1 = h_{12} v_2 \qquad (5.9)$$

and $$i_2 = h_{22} v_2 \qquad (5.10)$$

c) Reverse voltage gain (h_{12})

From Eq. (5.9), we can write

$$h_{12} = \frac{v_1}{v_2} \quad \text{(with input terminals open)} \qquad (5.11)$$

Because we have a ratio of input and output voltages, it is termed as **reverse voltage gain.** The word "reverse" is used to denote transfer from the output back to input.

Normally, we consider the gain of a circuit as a ratio of output (voltage or current) or input (voltage or current). It is referred to as *forward gain* or simply *gain* of the circuit.

d) Output admittance (h_{22})

From Eq. (5.10), we get the expression for h_{22} as

$$h_{22} = \frac{i_2}{v_2} \quad \text{(with input terminal open)} \qquad (5.12)$$

Here the parameter is the ratio of current to voltage; hence, it represents admittance (reciprocal of impedance). The unit of admittance is siemens, indicated by S (also referred to as mho, as it is reciprocal of ohm.)

Now, you have learnt the significance of all the *h*-parameters in terms of electrical quantities. Table 5.1 summarizes the meaning of each *h*-parameter and the required condition.

Table 5.1: *h*-parameters

Parameter	Meaning	Equation	Condition
h_{11}	Input impedance	$\frac{v_1}{i_1}$	Output shorted
h_{12}	Reverse voltage gain	$\frac{v_1}{v_2}$	Input open
h_{21}	Current gain	$\frac{i_2}{i_1}$	Output shorted
h_{22}	Output admittance	$\frac{i_2}{v_2}$	Input open

You may now like to attempt an SAQ.

SAQ 1 – *h*-parameters

a) In the circuit shown in Fig. 5.4, $v_1 = 50\,\text{mV}$, $i_1 = 20\,\mu\text{A}$ and $i_2 = 4\,\text{mA}$. Calculate the input impedance and current gain of the circuit.

b) In the circuit shown in Fig. 5.5, $v_1 = 50\,\text{mV}$, $i_2 = 4\,\text{mA}$ and $v_2 = 1\,\text{V}$. Calculate the reverse voltage gain and output admittance of the circuit.

Now that you are familiar with the *h*-parameters for any general circuit represented by a 2-port (4-terminal) network, we will apply them to bipolar junction transistor circuits. For this purpose, we first represent the transistor in its equivalent circuit form.

5.3 EQUIVALENT CIRCUIT OF A TRANSISTOR

Any device in practice behaves differently than its expected ideal characteristics. For example, a voltage source is expected to supply a constant voltage irrespective of the current we draw from it.

However, if you look at the output voltage of a voltage source, you will find that it decreases as we increase the current drawn from it. This is because an ideal voltage source is expected to have zero resistance. But in practice, there is always a finite (however small, but non-zero) resistance of a practical voltage source. Hence, we represent a practical voltage source as an ideal voltage source (V) in series with a finite source resistance (R_S) as shown in Fig. 5.6.

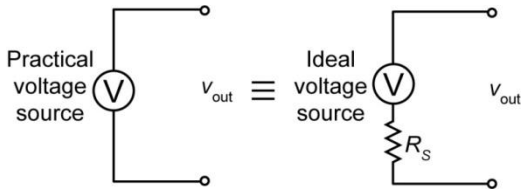


Fig. 5.6: **Practical voltage source.**

The same is true for a transistor circuit. We can replace the transistor by equivalent circuit components and then apply the circuit laws to obtain various transistor parameters.

Suppose some transistor's equivalent circuit is simply an ideal voltage generator of magnitude Av_{in} in series with a resistance R as shown in Fig. 5.7.

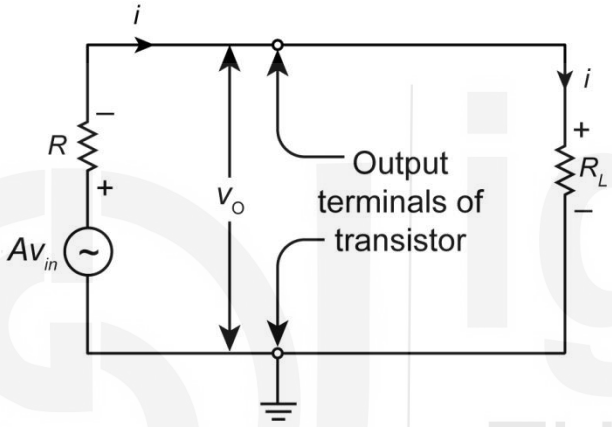


Fig. 5.7: **Simple equivalent circuit of a transistor.**

Algebraically, A is a positive number, and v_{in} is the amplitude of the input voltage to the transistor. The resistor R_L is the load resistance connected to the output terminals and current i is flowing through the circuit. The voltage generator produces a voltage A times as large as the input, so we may intuitively think of A as the voltage gain at this stage of the calculation. However, the voltage gain is

$$A_v = \frac{v_{out}}{v_{in}} = \frac{iR_L}{v_{in}} \qquad (5.13)$$

and the current i can be obtained from the Kirchhoff voltage law for the loop containing the generator, R and R_L. Hence,

$$Av_{in} - iR - iR_L = 0$$

or
$$i = \frac{Av_{in}}{R + R_L} \qquad (5.14)$$

Thus, the voltage gain becomes

$$A_v = \frac{v_{out}}{v_{in}} = \frac{iR_L}{v_{in}} = \frac{Av_{in}R_L}{(R + R_L)v_{in}}$$

or
$$A_v = A\left(\frac{R_L}{R + R_L}\right) \qquad (5.15)$$

We see that the voltage gain depends on A, R and R_L, and only as R_L becomes very large compared to R we get $A_V \cong A$.

This was a very simplistic approach just for understanding the treatment involved in working with equivalent circuits. But, in general, we consider the device to be a black box with four terminals. Two of these terminals act as input while the other two as output. You know that the transistor has three terminals, whereas our black box from which the equivalent circuit is developed has four. Hence, for the equivalent circuit to be applied to a transistor, one transistor terminal must be common between the input and output. This can either be the emitter, the collector or the base called respectively the "common emitter" (CE), "common collector" (CC) or the "common base" (CB) configurations. You have learnt about these configurations in Unit 4.

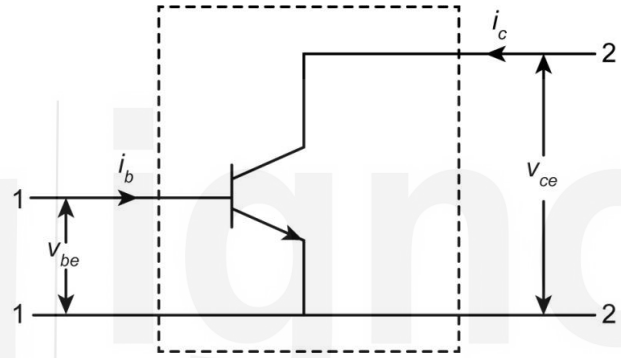


Fig. 5.8: **CE configuration of transistor as a 4-terminal network.**

Consider the transistor shown in Fig. 5.8, where we have taken the CE configuration. Here 1, 1 are the input terminals and 2,2 are the output terminals, i_b and v_{be} are the input current and input voltage while i_c and v_{ce} are the corresponding values of current and voltage in the output circuit, respectively. Thus, we have four quantities, two of which are currents and two are voltages. Of these four quantities, we can take two as dependent quantities and the other two as independent quantities. Then we express the dependent quantities in terms of the independent ones. For a transistor, an equivalent circuit by the selection of i_b and v_{ce} as independent and i_c, v_{be} as dependent is widely used due to the simplicity involved. Thus, we write

$$i_c = f(i_b, v_{ce}) \qquad (5.16)$$

and $$v_{be} = f(i_b, v_{ce}) \qquad (5.17)$$

So, in terms of the general 4-terminal circuit for CE configuration, we have

$$i_1 = i_b$$
$$i_2 = i_c$$
$$v_1 = v_{be}$$
$$v_2 = v_{ce}$$

For the general network, we have expressed v_1 and i_2 in terms of i_1 and v_2 in Eqs. (5.3) and (5.4). You must have noticed that Eq. (5.3) is merely the Kirchhoff voltage equation for the input and Eq. (5.4) is the Kirchhoff current equation for the output. From those equations, we can arrive at the equivalent circuit in terms of *h*-parameters.

The term $h_{21}i_1$ means that there is a current generator of magnitude h_{21} multiplied by i_1. The term $h_{22}v_2$ means the voltage v_2 appears across an admittance h_{22} (or equivalently across a resistance of $\frac{1}{h_{22}}$ ohms). The term $h_{11}i_1$ means the current i_1 flows through an effective resistance of h_{11} ohms. The term $h_{12}v_2$ means there is a voltage generator of magnitude $h_{12}v_2$. The input signal is represented by source e_s with internal resistance r_s and at the output, load r_L is connected. Therefore, we can draw the equivalent circuit as shown in Fig. 5.9. The physical significance of the *h*-parameters can be understood from the equivalent circuit of Fig. 5.9.

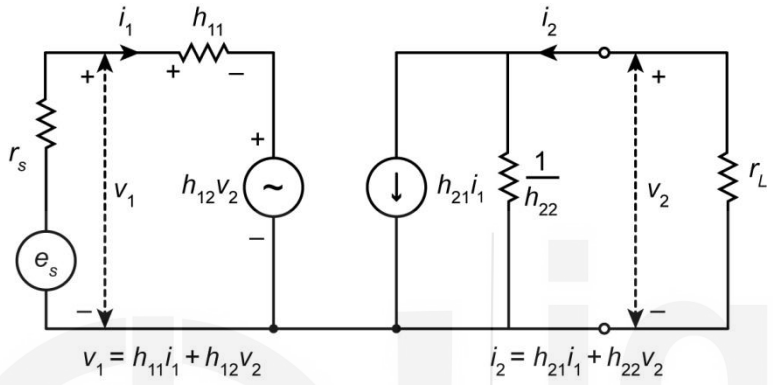


Fig. 5.9: **Transistor *h*-parameter equivalent circuit.**

The *input resistance* h_{11} is a resistance in the input circuit. The term $h_{12}v_2$ is the amplitude of a voltage generator in the input; it represents how much of the output voltage v_2 is transferred or fed back to the input and h_{12} is also called *reverse voltage transfer ratio*. The parameter h_{21} represents how much of the input current i_1 is transferred to the output, i.e., the *current gain*. The higher the value of h_{21}, the larger is the change in output current for a given input current change. We call h_{22} the *output admittance* because it is an admittance or conductance directly across the output terminals parallel to the load resistor r_L.

The *h*-parameters of transistor can be listed as follows:

$$h_i = h_{11}$$
$$h_r = h_{12}$$
$$h_f = h_{21}$$
$$h_o = h_{22}$$

where h_i = input impedance with output shorted
h_r = reverse voltage gain with input open
h_f = forward current gain with output shorted
h_o = output admittance with input open

To remember this, notice that the subscript is the first letter of the description.

i = input
r = reverse
f = forward
o = output

The *h*-parameters of a transistor depend on the configuration: CE, CC or CB. Because of this, the letter *e* is included for CE configuration, *c* for CC configuration and *b* for CB configuration. Table 5.2 summarises the notation for commonly used transistor *h*-parameters. As you can see, the parameters corresponding to CE are h_{ie},h_{re},h_{fe} and h_{oe}.

Table 5.2: *h*-parameters for different transistor configurations

Parameter	CE	CC	CB
h_{11}	h_{ie}	h_{ic}	h_{ib}
h_{12}	h_{re}	h_{rc}	h_{rb}
h_{21}	h_{fe}	h_{fc}	h_{fb}
h_{22}	h_{oe}	h_{oc}	h_{ob}

The general *h*-parameter equivalent circuit of Fig. 5.9 and Eqs. (5.3) and (5.4) are widely used to calculate the voltage gain, current gain, input impedance and output impedance of the transistor amplifier in its three configurations.

You may like to attempt an SAQ now.

SAQ 2 – Transistor *h*-parameters

Write down the electrical units of *h*-parameters h_{11}, h_{12}, h_{21} and h_{22}.

You will now learn how to obtain the expression for these transistor parameters.

5.4 ANALYSIS OF COMMON EMITTER AMPLIFIER

Fig. 5.10 shows a common emitter (CE) amplifier. A small sine wave is applied at the input. This produces variations in the base current. Because of current gain β, the collector current is an amplified sine wave of the same frequency. This sinusoidal collector current flows through the collector resistance and produces an amplified output voltage.

Notice that the ac output voltage is inverted with respect to the ac input voltage, meaning that it is 180° out of phase with the input. During the positive half cycle of input voltage, the base current increases, causing the collector current to increase. This produces a larger voltage drop across the collector resistor. Therefore, the collector voltage decreases, and we get the first negative half cycle of output voltage.

Conversely, on the negative cycle of input voltage, less collector current flows and the voltage drop across the collector resistor decreases. For this reason, the collector-to-ground voltage rises and we get the positive half cycle of the output voltage.

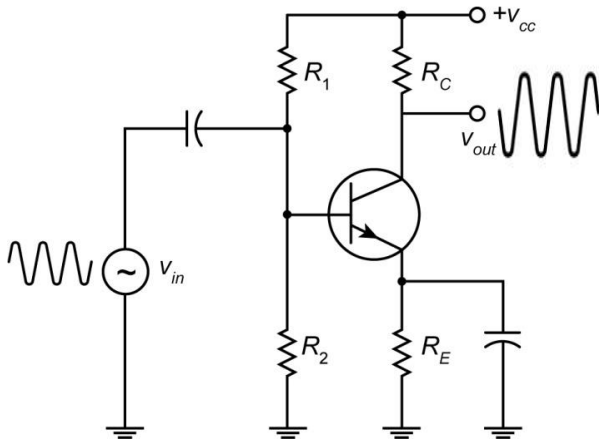


Fig. 5.10: **Common emitter amplifier.**

Now that you have learnt the working of CE amplifier, let us obtain the expressions for the circuit characteristics in terms of *h*-parameters. For this we represent the CE circuit in its *h*-parameter model shown in Fig. 5.11.

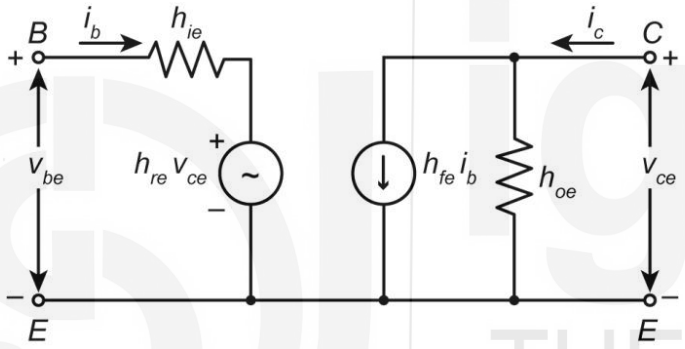


Fig. 5.11: **CE hybrid circuit.**

5.4.1 Current Gain

The current gain of an amplifier is defined as the ratio of the ac output current to ac input current. In symbols,

$$A_i = \frac{i_2}{i_1} \qquad (5.18)$$

Using Eq. (5.4), we can rewrite Eq. (5.18) as

$$A_i = \frac{h_{21}i_1 + h_{22}v_2}{i_1} = h_{21} + h_{22}\frac{v_2}{i_1}$$

From Fig. 5.9 you can see that, output voltage $v_2 = -i_2 r_L$. (The negative sign of current i_2 is because it is considered to be coming out of the output port and flowing into the load resistor.) Substituting this in Eq. (5.18), we get

$$A_i = h_{21} - h_{22}\frac{i_2 r_L}{i_1} = h_{21} - A_i h_{22} r_L$$

Solving for A_i, we get

$$A_i = \frac{h_{21}}{1 + h_{22} r_L} \qquad (5.19)$$

5.4.2 Voltage Gain

The voltage gain of an amplifier is defined as the ratio of the ac output voltage to ac input voltage:

$$A_v = \frac{v_2}{v_1} \tag{5.20}$$

From Eq. (5.3), we can write

$$A_v = \frac{v_2}{h_{11} i_1 + h_{12} v_2} = \frac{-i_2 r_L}{h_{11} i_1 - h_{12} i_2 r_L}$$

Dividing the numerator and denominator by i_2 gives

$$A_v = \frac{-r_L}{\frac{h_{11}}{A_i} - h_{12} r_L}$$

Substituting for A_i from Eq. (5.19) and simplifying we get:

$$A_v = \frac{-h_{21} r_L}{h_{11} + (h_{11} h_{22} - h_{12} h_{21}) r_L} \tag{5.21}$$

5.4.3 Input Impedance

The input impedance of a loaded two port network is

$$Z_{in} = \frac{v_1}{i_1} \tag{5.22}$$

On substituting for v_1 from Eq. (5.3), we get

$$Z_{in} = \frac{h_{11} i_1 + h_{12} v_2}{i_1} = h_{11} + \frac{h_{12} v_2}{i_1}$$

Using $v_2 = -i_2 r_L$, and

$$Z_{in} = h_{11} - \frac{h_{12} i_2 r_L}{i_1}$$

However, $\frac{i_2}{i_1} = A_i$, hence

$$Z_{in} = h_{11} - A_i \, h_{12} \, r_L$$

Using Eq. (5.19), we get

$$Z_{in} = h_{11} - \frac{h_{12} \, h_{21} r_L}{1 + h_{22} \, r_L} \tag{5.23}$$

5.4.4 Output Impedance

To get the output impedance, the source voltage shown in Fig. 5.9 is reduced to zero. When we drive the output terminals with a signal v_2, the circuit takes the form as shown in the Fig. 5.12.

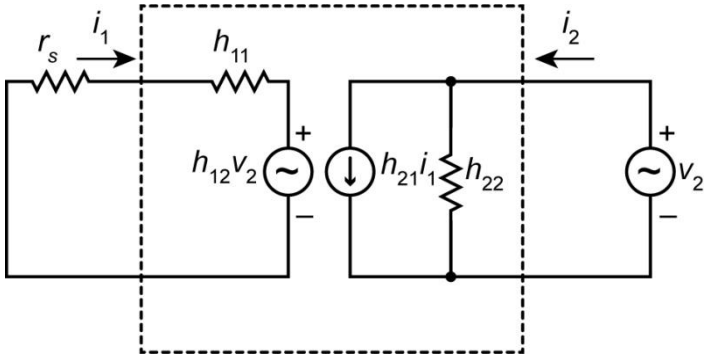


Fig. 5.12: **Determination of output impedance by removing input voltage.**

The ratio of v_2 to i_2 is the output impedance (reciprocal of admittance) of the two-port network:

$$Z_{out} = \frac{v_2}{i_2} \qquad (5.24)$$

Substituting for i_2 from Eq. (5.4) in Eq. (5.24), we have:

$$Z_{out} = \frac{v_2}{h_{21}i_1 + h_{22}v_2} \qquad (5.25)$$

Applying Ohm's law on the input side in Fig. 5.12 we can write:

$$i_1 = \frac{-h_{12}v_2}{r_s + h_{11}}$$

When this is substituted in Eq. (5.25), we get:

$$Z_{out} = \frac{r_s + h_{11}}{(r_s + h_{11})h_{22} - h_{12}h_{21}} \qquad (5.26)$$

Thus, for a CE amplifier, the *h* formulas can be written using the symbols from Table 5.2 as follows:

$$A_i = \frac{h_{fe}}{1 + h_{oe}r_L} \qquad (5.27a)$$

$$A_v = \frac{-h_{fe}\,r_L}{h_{ie} + (h_{ie}h_{oe} - h_{re}h_{fe})\,r_L} \qquad (5.27b)$$

$$Z_{in} = h_{ie} - \frac{h_{re}h_{fe}r_L}{1 + h_{oe}r_L} \qquad (5.27c)$$

$$Z_{out} = \frac{r_s + h_{ie}}{(r_s + h_{ie})h_{oe} - h_{re}h_{fe}} \qquad (5.27d)$$

Now, we will consider the analysis of other configurations of transistor amplifiers.

But before studying further, you may like to solve SAQ 3.

SAQ 3 – Analysis of CE configuration of transistor

For a transistor in CE configuration the values of *h*-parameters are:

$h_{11} = 5\,\text{k}\Omega$, $h_{12} = 1.2 \times 10^{-4}$

$h_{21} = 200$, $h_{22} = 10\,\mu\text{S}$.

Calculate the current gain, voltage gain and input impedance of the circuit if $r_L = 2\,\text{k}\Omega$.

5.5 ANALYSIS OF COMMON BASE AMPLIFIER

Fig. 5.13 shows a common base amplifier.

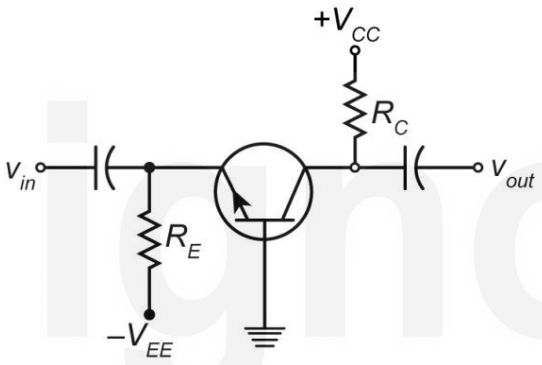


Fig. 5.13: **Common base amplifier.**

The base terminal is common to both the input and the output. The input is at the emitter, and the output is taken off the collector, i.e., across the collector resistor R_C.

In the common base configuration, the output is in phase with the input. A positive input makes the emitter more positive, that is v_{be} reduces, and hence, the collector current decreases. This decrease in the collector current causes the output voltage at the collector to rise, thus giving a higher positive output.

For calculating the current gain, voltage gain, input impedance and output impedance, we need to use the CB parameters h_{ib}, h_{rb}, h_{jb} and h_{ob} as shown in Fig. 5.14.

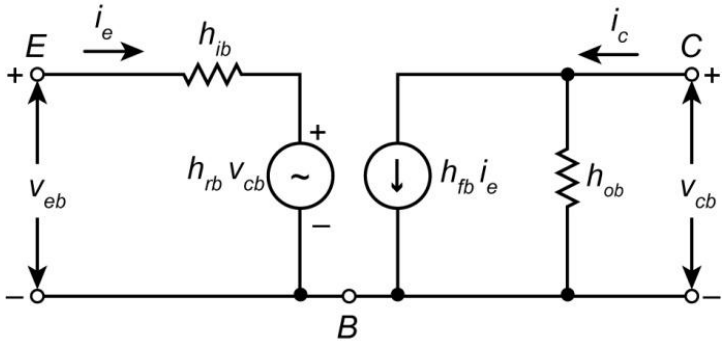


Fig. 5.14: **Hybrid circuit of CB configuration.**

Then the basic formulas become:

$$A_i = \frac{h_{fb}}{1 + h_{ob} r_L} \quad (5.28a)$$

$$A_v = \frac{-h_{fb} r_L}{h_{ib} + (h_{ib} h_{ob} - h_{rb} h_{fb}) r_L} \quad (5.28b)$$

$$Z_{in} = h_{ib} - \frac{h_{rb} h_{fb} r_L}{1 + h_{ob} r_L} \quad (5.28c)$$

$$Z_{out} = \frac{r_S + h_{ib}}{(r_S + h_{ib}) h_{ob} + h_{rb} h_{fb}} \quad (5.28d)$$

You may now like to attempt an SAQ.

SAQ 4 – Common base amplifier

For a transistor, the *h*-parameters for CB configuration are $h_{ib} = 20\Omega, h_{rb} = 3 \times 10^{-4}, h_{fb} = -0.98, h_{ob} = 0.5\,\mu s$. Calculate the A_i, A_v and Z_{in} . The load resistance is 1 kΩ.

Next, we consider the common collector amplifier.

5.6 ANALYSIS OF COMMON COLLECTOR AMPLIFIER

In the common collector configuration shown in Fig. 5.15, the collector terminal is common to both the input and the output. The input is at the base and the output is taken off the emitter, that is, across the emitter resistor R_E.

As the input goes more positive, the current flows in the transistor and I_E increases, which means that the output also goes more positive. In other words, the output voltage is in phase with the input voltage. The common collector amplifier is thus often called the "*emitter follower*" because the output voltage on the emitter "follows" the input voltage at the base.

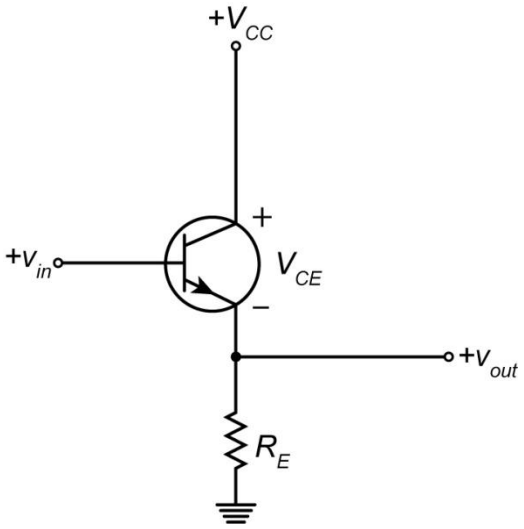


Fig. 5.15: **Common collector amplifier.**

To calculate the current gain, voltage gain, input impedance and output impedance we need to use the *h*-parameters in CC configuration: h_{ic}, h_{rc}, h_{fc} and h_{oc} as depicted in the hybrid circuit of CC configuration shown in Fig. 5.16.

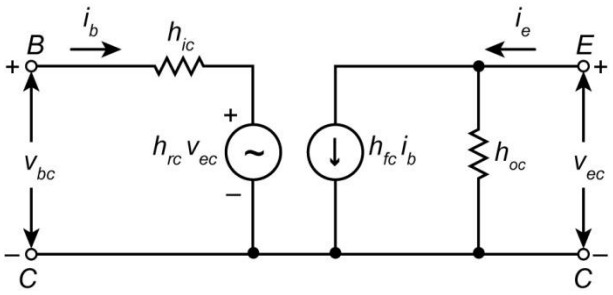


Fig. 5.16: **Hybrid circuit of CC configuration.**

Thee relations associated with CC configuration are as follows:

$$A_i = \frac{h_{fc}}{1 + h_{oc} r_L} \quad (5.29a)$$

$$A_v = \frac{-h_{fc} r_L}{h_{ic} + (h_{ic} h_{oc} - h_{rc} h_{fc}) r_L} \quad (5.29b)$$

$$Z_{in} = h_{ic} - \frac{h_{rc} h_{fc} r_L}{1 + h_{oc} r_L} \quad (5.29c)$$

$$Z_{out} = \frac{r_S + h_{ic}}{(r_S + h_{ic}) h_{oc} - h_{rc} h_{fc}} \quad (5.29d)$$

To sum up, in Table 5.3 we give comparison among three types of amplifier circuits.

Table 5.3: Comparative study of three types of amplifier circuits

Property	**Common emitter**	**Common base**	**Common collector**
Transistor impedance	Medium input impedance and medium output impedance	A relatively low input impedance and high output impedance	High input impedance and low output impedance
Current gain	Large current gain	Approximately no current gain	Current gain is high
Voltage gain	Large voltage gain	Voltage gain is fair	Voltage gain is less than unity
Phase of input and output signals	Input and output signals are 180° apart	No phase reversal	No phase reversal
Principal use	In transistor amplifiers	In very high frequencies	In driving low impedance loads like loudspeakers

5.7 SUMMARY

Concept	Description
4-terminal device	■ **A four terminal (two port) network** can be analysed by considering it to be a black box with **two specified currents (input-output)** and **two specified voltages (input-output).**
z-parameters	■ When the **currents are taken to be independent variables** and **voltages as dependent variables,** then we get ***z*-parameters.**
y-parameters	■ With **voltages as independent and currents as dependent variables,** we get ***y*-parameters.**
h-parameters	■ If **input current and output voltage** are taken to be **independent,** we get **hybrid (or *h*-) parameters.** The basic equations are: $v_1 = h_{11} i_1 + h_{12} v_2$ and $i_2 = h_{21} i_1 + h_{22} v_2$ where h_{11} is **input impedance,** h_{12} is **reverse voltage gain,** h_{21} is **current gain** and h_{22} is **output admittance.**
Transistor parameters	■ The **transistor parameters** in terms of *h*-parameters are: • **Current gain,** $A_i = \frac{i_2}{i_1} = \frac{h_{21}}{1 + h_{22} r_L}$ • **Voltage gain,** $A_V = \frac{v_2}{v_1} = \frac{-h_{21} r_L}{h_{11} + (h_{11} h_{22} - h_{12} h_{21}) r_L}$ • **Input impedance,** $Z_{in} = \frac{v_1}{i_1} = h_{11} - \frac{h_{12} h_{21} r_L}{1 + h_{22} r_L}$ • **Output impedance,** $Z_{out} = \frac{v_2}{i_2} = \frac{r_S + h_{11}}{(r_S + h_{11}) h_{22} - h_{12} h_{21}}$
h-parameters for different transistor configurations	■

Parameter	CE	CC	CB
h_{11}	h_{ie}	h_{ic}	h_{ib}
h_{12}	h_{re}	h_{rc}	h_{rb}
h_{21}	h_{fe}	h_{fc}	h_{fb}
h_{22}	h_{oe}	h_{oc}	h_{ob}

5.8 TERMINAL QUESTIONS

1. Using the hybrid circuit for the CB configuration shown in Fig. 5.14, prove that $h_{fb} = \frac{h_{fe}}{1 + h_{fe}}$.

2. For a transistor with the following *h*-parameters for CC configuration, calculate the values of current gain and output impedance:

$h_{ic} = 10\,\text{k}\Omega$

$h_{rc} = 1$

$h_{fc} = -100$

$h_{oc} = 20\,\mu\text{S}$

It is given that source resistance $r_S = 500\,\Omega$ and load resistance $r_L = 50\,\Omega$.

5.9 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. a) $h_{11} = \dfrac{v_1}{i_1} = \dfrac{50\,\text{mV}}{20\,\mu\text{A}} = 2.5\,\text{k}\Omega$

$h_{21} = \dfrac{i_2}{i_1} = \dfrac{4\,\text{mV}}{20\,\mu\text{A}} = 200$

b) $h_{12} = \dfrac{v_1}{v_2} = \dfrac{50\,\text{mV}}{1\,\text{V}} = 0.05$

$h_{22} = \dfrac{i_2}{v_2} = \dfrac{4\,\text{mA}}{1\,\text{V}} = 0.004\,\text{S} = 4\,\text{mS}$

2. h_{11} = ohm (Ω)

h_{12} = no units (ratio of voltages)

h_{21} = no units (ratio of currents)

h_{22} = siemens or mho (S or Ω^{-1})

3. Given that for CE configuration

$h_{11} = h_{ie} = 5\,\text{k}\Omega, \quad h_{12} = h_{re} = 1.2 \times 10^{-4},$

$h_{21} = h_{fe} = 200, \quad h_{22} = h_{oe} = 10\,\mu\text{S}$ and $r_L = 2\,\text{k}\Omega$

For CE configuration:

i) Current gain

$$A_i = \frac{h_{fe}}{1 + h_{oe} r_L} = \frac{200}{1 + 10 \times 10^{-6} \times 2 \times 10^3}$$

$$= \frac{200}{1 + 2 \times 10^{-2}} = \frac{200}{1.02\Omega} = 196.0$$

ii) Voltage gain

$$A_v = \frac{-h_{fe} r_L}{h_{ie} + (h_{ie} h_{oe} - h_{re} h_{fe}) r_L}$$

$$= -\frac{200 \times 2 \times 10^3}{(5 \times 10^3) + \left[(5 \times 10^3 \times 10 \times 10^{-6}) - (1.2 \times 10^{-4} \times 200)\right] \times (2 \times 10^3)}$$

$$= -\frac{400\times10^3}{5\times10^3 + [0.05-0.024]\times(2\times10^3)}$$

$$= \frac{-400\times10^3}{(5\times10^3)+(0.052\times10^3)} = \frac{-400}{5.052} = 79.18$$

iii) Input impedance

$$Z_{in} = h_{ie} - \frac{h_{re}h_{fe}r_L}{1+h_{oe}r_L}$$

$$= (5\times10^3) - \frac{1.2\times10^{-4}\times200\times2\times10^3}{1+10\times10^{-6}\times2\times10^3}$$

$$= 5000 - \frac{48}{1+0.02} = 5000 - 47.06 = 4952.9\,\Omega$$

4. For CB configuration,

$$A_i = \frac{h_{fb}}{1+h_{ob}r_L}$$

$$= \frac{-0.98}{1+(0.5\times10^{-6})(1\times10^3)} = \frac{-0.98}{1+0.0005} \approx -0.98$$

$$A_v = \frac{-h_{fb}r_L}{h_{ib}+(h_{ib}h_{ob}-h_{rb}h_{fb})r_L}$$

$$= -\frac{-0.98\times1\times10^3}{20+[(20\times0.5\times10^{-6})-(3\times10^{-4}\times(-0.98)]\times(1\times10^3)}$$

$$= \frac{0.98\times10^3}{20+[(10\times10^{-6})+(2.94\times10^{-4})]\times10^3}$$

$$= \frac{0.98\times10^3}{20+(0.1+2.94)\times10^{-4}\times10^3} = \frac{980}{20+0.304}$$

$$= 48.27$$

$$Z_{in} = h_{ib} - \frac{h_{rb}h_{fb}r_L}{1+h_{ob}r_L}$$

$$= 20 - \frac{(3\times10^{-4})\times(-0.98)(1\times10^3)}{1+(0.5\times10^{-6}\times1\times10^3)} = 20 + \frac{0.294}{1.0005}$$

$$= 20.29\,\Omega$$

Terminal Questions

1. For the common base circuit, the input current is i_e, whereas output current is i_c. Hence $h_{fb} = \frac{i_c}{i_e}$. The relation between the collector and emitter current is given by Eq. (3.1a):

$$i_e = i_c + i_b$$

Now, from Eq. (4.2b), you know that the current gain β of the transistor is given by

$$\beta = \frac{i_c}{i_b},$$

But in common emitter configuration,

$$\text{Current gain} = h_{fe} = \frac{i_c}{i_b} (= \beta)$$

From Eq. (3.1a), we can write

$$i_e = i_b + i_c = i_c\left(\frac{1}{h_{fe}} + 1\right)$$

Current gain in CB configuration:

$$h_{fb} = \frac{i_c}{i_e} = \frac{i_c}{i_c\left(\frac{1}{h_{fe}} + 1\right)} = \frac{h_{fe}}{1 + h_{fe}}$$

2. Current gain in CC configuration is given by:

$$A_i = \frac{h_{fc}}{1 + h_{oc} r_L}$$

$$= \frac{-100}{1 + (20 \times 10^{-6} \times 50)} = \frac{-100}{1.001} \approx -100$$

Output impedance, $Z_{out} = \dfrac{r_s + h_{ic}}{(r_s + h_{ic}) h_{oc} - h_{rc} h_{fc}}$

$$= \frac{500 + (10 \times 10^3)}{[(500 + (10 \times 10^3)) \times 20 \times 10^{-6}] - (1 \times -100)}$$

$$= \frac{10.5 \times 10^3}{(10.5 \times 10^3 \times 20 \times 10^{-6}) + 100}$$

$$= \frac{10.5 \times 10^3}{0.210 + 100} \approx \frac{10.5 \times 10^3}{100} = 105\,\Omega$$