ANALOG & DIGITAL ELECTRONICS

M.Sc. PHYSICS

SEMESTER-I, PAPER-IV

LESSON WRITERS

Prof. Sandhya Cole
Professor
Department of Physics
University College of Sciences
Acharya Nagarjuna University

Prof. R.V.S.S.N. Ravi Kumar
Professor
Department of Physics
University College of Sciences
Acharya Nagarjuna University

Prof. G. Naga Raju
Professor
Department of Physics
University College of Sciences
Acharya Nagarjuna University

Prof. Ch. Linga Raju
Professor
Department of Physics
University College of Sciences
Acharya Nagarjuna University

EDITOR

Prof. Sandhya Cole
Professor
Department of Physics
University College of Sciences
Acharya Nagarjuna University

DIRECTOR, I/c

Prof. V. Venkateswarlu
M.A., M.P.S., M.S.W., M.Phil., Ph.D.
Professor
Centre for Distance Education
Acharya Nagarjuna University
Nagarjuna Nagar 522 510

Ph: 0863-2346222, 2346208
0863- 2346259 (Study Material)
Website www.anucde.info
E-mail: anucdedirector@gmail.com

M.Sc. PHYSICS: ANALOG & DIGITAL ELECTRONICS

First Edition : 2025

No. of Copies :

This book is exclusively prepared for the use of students of M.Sc. Physics, Centre for Distance Education, Acharya Nagarjuna University and this book is meant for limited circulation only.

Published by:

Prof. V. VENKATESWARLU
Director, I/c
Centre for Distance Education,
Acharya Nagarjuna University

Printed at:

FOREWORD

Since its establishment in 1976, Acharya Nagarjuna University has been forging ahead in the path of progress and dynamism, offering a variety of courses and research contributions. I am extremely happy that by gaining 'A+' grade from the NAAC in the year 2024, Acharya Nagarjuna University is offering educational opportunities at the UG, PG levels apart from research degrees to students from over 221 affiliated colleges spread over the two districts of Guntur and Prakasam.

The University has also started the Centre for Distance Education in 2003-04 with the aim of taking higher education to the door step of all the sectors of the society. The centre will be a great help to those who cannot join in colleges, those who cannot afford the exorbitant fees as regular students, and even to housewives desirous of pursuing higher studies. Acharya Nagarjuna University has started offering B.Sc., B.A., B.B.A., and B.Com courses at the Degree level and M.A., M.Com., M.Sc., M.B.A., and L.L.M., courses at the PG level from the academic year 2003-2004 onwards.

To facilitate easier understanding by students studying through the distance mode, these self-instruction materials have been prepared by eminent and experienced teachers. The lessons have been drafted with great care and expertise in the stipulated time by these teachers. Constructive ideas and scholarly suggestions are welcome from students and teachers involved respectively. Such ideas will be incorporated for the greater efficacy of this distance mode of education. For clarification of doubts and feedback, weekly classes and contact classes will be arranged at the UG and PG levels respectively.

It is my aim that students getting higher education through the Centre for Distance Education should improve their qualification, have better employment opportunities and in turn be part of country's progress. It is my fond desire that in the years to come, the Centre for Distance Education will go from strength to strength in the form of new courses and by catering to larger number of people. My congratulations to all the Directors, Academic Coordinators, Editors and Lesson-writers of the Centre who have helped in these endeavors.

Prof. K. Gangadhara Rao
M.Tech., Ph.D.,
Vice-Chancellor I/c
Acharya Nagarjuna University.

M.SC. PHYSICS

SYLLABUS

SEMESTER - I, PAPER - IV

104PH24 - ANALOG & DIGITAL ELECTRONICS

Course Objectives:

- Introduction of Semiconductor Devices and the Opto-electronic devices and their analysis.
- Operational Amplifiers, Construction and working DC and AC analysis, Effect of Feedback.
- Acquiring the Knowledge in Communication Electronics and then the Digital electronics.
- Architecture of 8085 Microprocessor, Instruction set, Addressing modes and some illustrative programmes.
- Introduction to 8051 Microcontroller.

UNIT-I

Semiconductor Devices

Diodes, Junction diode, Tunnel diode, Photo diode, transistors, Silicon controlled rectifier, Uni junction transistor, Field effect transistor, JFET & MOSFET, CMOS, Opto-electronic devices: Solar cells, Photo detectors, LEDs.

Learning Outcomes:

- Construction and working of Tunnel diode, photodiode, Silicon Controlled Rectifier, Uni-junction Transistor.
- Know the Characteristics of FET, MOSFET and CMOS.
- Construction, working and applications of Solar Cells and LED's.

UNIT II

Operational Amplifiers

Differential Amplifier - Circuit Configurations - dual input, balanced output differential amplifier - DC analysis - Ac analysis, inverting and non-inverting inputs CMRR. Block diagram of a typical Op - Amp - analysis. Open loop configuration inverting and non - inverting amplifiers. Op-amp with negative feedback – voltage series feedback- effect of feedback on closed loop gain input resistance output resistance bandwidth and output offset voltage- voltage follower. Practical Op-amps

Input Offset Voltage- input bias current-input offset current, total output offset voltage, CMRR frequency response. Summing amplifier, Scaling and Averaging amplifiers, integrator and differentiator.

Oscillators Principles - oscillator types - frequency stability – response The phase shift oscillator, Weinbridge oscillator – Multi vibrators- Monostable and a stable-comparators- Square wave and triangular wave generators- voltage regulators.

Learning Outcomes:

- To learn about the Differential amplifier and then the Operational amplifier, AC and DC analysis, Characteristics, Effect of Feedback.
- Oscillators Principles, Construction and working of different types of Oscillators. Clear picture of Multivibrators and then the Comparators using Operational amplifiers.

UNIT III

Communication Electronics

Amplitude modulation - Generation of AM waves - Demodulation of AM waves DSBSC modulation. Generation of DSBSC waves. Coherent detection of DSBSC waves, SSB modulation, Generation and detection of SSB waves. Vestigial side band modulation, Frequency Division Multiplexing (FDM).

Learning Outcomes:

Acquiring knowledge in Communication electronics AM & FM, modulation and Demodulation.

UNIT IV

Digital Electronics

Simplification of Boolean Expressions: Algebraic Method, Karnaugh Method, EX - OR, EX - NOR Gates, Combinational Logic Gates - Decoder - encoders- Multiplexer (data selectors) - Application of Multiplexer - De multiplexer(data distributors), Sequential Logic - Flip-Flops: A 1- Bit Memory - The R-S Flip Flop, JK Flip-Flop - JK Master Slave Flip-Flop - T- Flip-Flop - D Flip-Flop - Shift Registers - Synchronous and Asynchronous Counters - Cascade Counters - A/D to D/A Converters.

Learning Outcomes:

- Expressing the Boolean expressions in a simple way, Karnaugh method.
- Learning the Combinational Logic Circuits, Multiplexer and Demultiplexer.
- Learning about the Sequential Logic circuits- Flip-Flops, Registers and Counters

UNIT V

Microprocessors

Introduction to microcomputers - memory - input/output - interfacing devices 8085 CPU - Architecture - BUS timings - Demultiplexing the address bus generating control signals - instruction set - addressing modes - illustrative programmes - writing assembly language programmes - looping, counting and indexing - counters and timing delays - stack and subroutine. Introduction to micro controllers - 8051 micro controllers - architecture & pin description.

Data interpretation and analysis - Precision and accuracy - Error analysis, propagation of errors. Least squares fitting.

Learning Outcomes:

- Knowledge about the Microprocessor and its Architecture
- Learning the Instruction Set, Addressing modes
- Writing the programmes using 8085 instructions
- Little knowledge about 8051 Microcontroller

Course Outcomes:

At the end of the course the student is expected to assimilate the following and possesses basic knowledge of the following.

- The design and functional performance of various semiconductor and optoelectronic devices such as Diodes, transistors, Solar Cells, Photo detectors and LEDs.
- To learn about the Differential amplifier and then the Operational amplifier, AC and DC Analysis, Characteristics, Effect of Feedback.
- Acquiring knowledge in Communication electronics AM &FM, modulation and Demodulation
- Learning the Combinational Logic Circuits, Multiplexer and Demultiplexer
- Learning about the Sequential Logic circuits- Flip-Flops, Registers and Counters
- Knowledge about the Microprocessor and its Architecture

Text and Reference Books:

1) Electronic Devices and Circuits - G.K.Mithal (Khanna)
2) Integrated Electronics- Jacob Millman & C.C. Halkies (TMH)
3) Op-Amps & Linear Integrated Circuits - Ramakanth A. Gayakwad
4) Op-Amps & Linear Integrated Circuits - D. Mahesh Kumar (MacMillan)
5) Digital Principles and Applications by A.P.Malvino and Donald P.Leech TMH 1993.
6) Microprocessor Architecture, Programming and Applications with 8085/8086 by Ramesh S.Gaonkar, Wiely-Eastern 1987.
7) Digital Electronics: An Introduction to theory and Practical - William. H.Gotnman.

(104PH24)

M.Sc. DEGREE EXAMINATION, MODEL QUESTION PAPER

M.Sc. PHYSICS-FIRST SEMESTER

ANALOG & DIGITAL ELECTRONICS

Time: Three hours **Maximum: 70 marks**

Answer ALL Questions

All Questions Carry Equal Marks

1 a) Explain the working principle and applications of a silicon-controlled rectifier (SCR).

b) Compare the construction and characteristics of JFET and MOSFET.

OR

c) Describe the instruction set classification of the 8085 microprocessors.

d) Write an 8085-assembly language program to perform multiplication of two numbers.

2 a) Explain the DC and AC analysis of a differential amplifier.

b) What is the voltage follower configuration of an operational amplifier? Discuss its applications.

OR

c) Describe the working and frequency stability of a phase shift oscillator.

d) Compare monostable and Astable multivibrators, with circuit diagrams.

3 a) What is amplitude modulation (AM)? Explain its advantages and disadvantages.

b) Describe the working of coherent detection of DSBSC waves.

OR

c) Compare frequency division multiplexing (FDM) and time division multiplexing (TDM).

d) Explain the process of generation and detection of SSB waves.

4 a) What is the Karnaugh map method for simplifying Boolean expressions?

b) Explain the working of a JK master-slave flip-flop with a neat diagram.

OR

c) What is the function of a decoder and encoder in digital circuits?

d) Compare synchronous and asynchronous counters with examples.

5 a) Explain the bus architecture and timing diagrams of the 8085 microprocessors.

b) Describe the stack and subroutine operations in the 8085 microprocessors.

OR

c) What are the key differences between the 8085 microprocessor and 8051 microcontrollers

d) Explain the importance of error analysis and least squares fitting in data interpretation.

CONTENTS

LESSON-1

ADVANCEMENTS IN SEMICONDUCTOR TECHNOLOGY: DIODES, TRANSISTORSAND RECTIFIERS

1.0 AIM AND OBJECTIVES:

The aim of this study is to explore the construction, characteristics, and applications of semiconductor devices such as diodes, transistors, and silicon-controlled rectifiers. It focuses on understanding the working principles of p-n junction diodes, tunnel diodes, and photodiodes, analyzing their current-voltage characteristics, and examining their role in rectification, amplification, and switching applications. The study also investigates the impact of doping on the depletion region, the function of transistors in electronic circuits, and the significance of semiconductor materials like silicon and germanium. Additionally, it covers the working of silicon-controlled rectifiers in power control applications and the tunneling effect in tunnel diodes. This knowledge is essential for understanding modern electronic components and their practical applications in various fields.

STRUCTURE:

1.1 DIODES:

A diode is a dispositive made of a semiconductor material, which has two terminals or electrodes (di-ode), that act like an on-off switch. When the diode is “on”, it acts as a short circuit and passes all current. When it is “off”, it behaves like an open circuit and passes no current. The two terminals are different and are marked as plus and minus in figure 1.1. If the polarity of the applied voltage matches that of the diode (forward bias), then the diode turns “on”. When the applied voltage polarity is opposite (reverse bias), it turns “off”. Of course this is the theoretical behaviour of an ideal diode, but it can be seen as a good approximation for a real diode.

A diode is simply a p-n junction (see 'Introduction into Semiconductor Physics') with the following characteristics:

- Under forward bias, it needs a small voltage to conduct. This voltage drop is maintained during conduction.
- The maximum forward current is limited by heat-dissipation ability of the diode. Usually it is around 1000 mA.
- There is a small reverse current

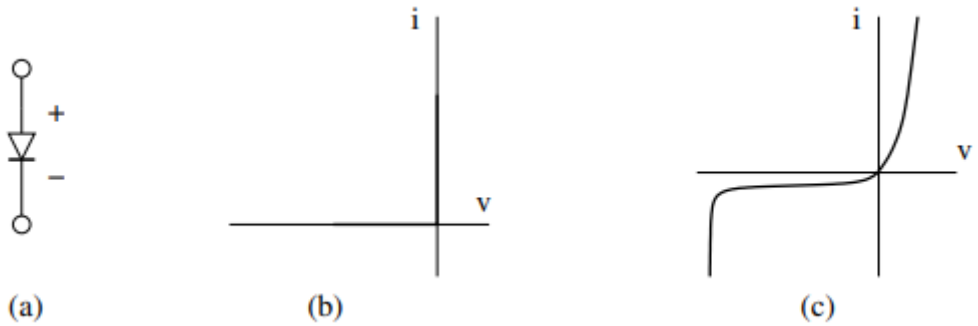


Figure 1.1: (a) Diode symbol. (b) Current-voltage characteristics of an ideal diode. (c) IV curves for a real diode.

Every diode has a maximum reverse voltage (breakdown voltage) that cannot be exceeded without diode damage.

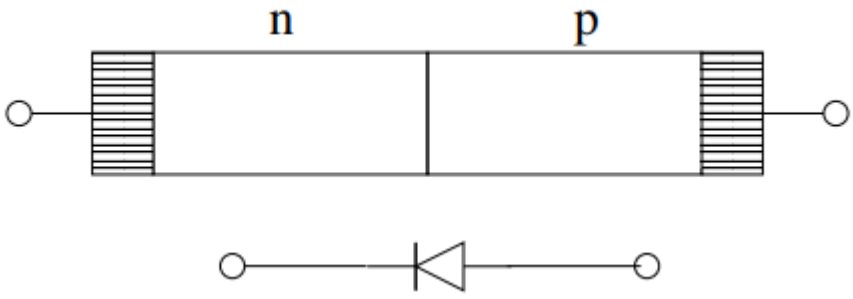


Figure 1.2: A p − n diode junction structure and the equivalent device schematics.

1.2 JUNCTION DIODE:

1.2.1 Basic Operation:

The diode is fabricated of a semiconductor material, usually silicon, which is doped with two impurities. One side is doped with a donor or n-type impurity which releases electrons into the semiconductor lattice. These electrons are not bound and are free to move about. Because there is no net charge in the donor impurity, the n-type semiconductor is electrically neutral.

The other side is doped with an acceptor or p-type impurity which imparts free holes into the lattice. A hole is the absence of an electron which acts as a positive charge. The p-type semiconductor is also electrically neutral because the acceptor material adds no net charge.

Figure 1.3 (a) illustrates the cross section of the diode. The junction is the dividing line between the n-type and p-type sides. Thermal energy causes the electrons and holes to move randomly. Electrons diffuse across the junction into the p-type side and holes diffuse across the junction into the n-type side. This causes a net positive charge to develop in the n-type side and a net negative charge to develop in the p-type side. These charges set up an electric field across the junction which is directed from the n-type side to the p-type side. The electric field opposes further diffusion of the electrons and holes. The region in which the electric field exists is called the depletion region. There are no free electrons or holes in this region because the electric field sweeps them out.

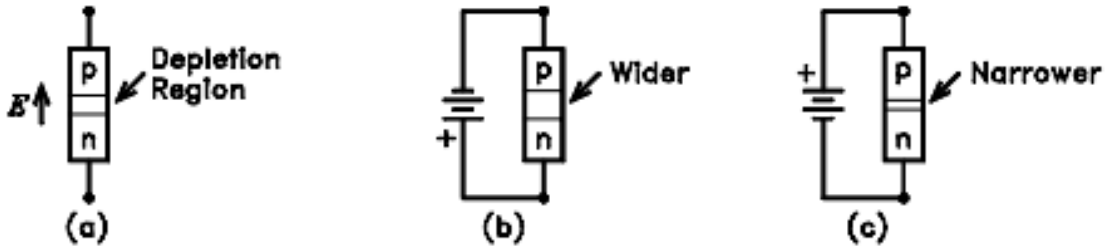


Figure 1.3: (a) Diode cross section. (b) Reverse biased diode. (c) Forward biased diode

Figure 1.3(b) shows the diode with a battery connected across it. The polarity of the battery is such that it reinforces the electric field across the junction causing the region to widen. The positive terminal pulls electrons in the n-type side away from the junction. The negative terminal pulls holes in the p-type side away from the junction. No current can flow. The diode is said to be reverse biased. Figure 1.3(c) shows the diode with the battery polarity reversed. The battery now tends to cancel out the electric field in the depletion region, causing its width to decrease. The positive terminal forces holes toward the junction. The negative terminal forces electrons toward the junction. A current flows which increases rapidly if the applied voltage is increased. The diode is said to be forward biased.

1.3 TUNNEL DIODE:

A Tunnel diode is a heavily doped p-n junction diode in which the electric current decreases as the voltage increases. In tunnel diode, electric current is caused by “Tunneling”. The tunnel diode is used as a very fast switching device in computers. It is also used in high-frequency oscillators and amplifiers.

1.3.1 Symbol of Tunnel Diode:

The circuit symbol of tunnel diode is shown in the below figure1.4. In tunnel diode, the ptype semiconductor act as an anode and the n-type semiconductor act as a cathode.

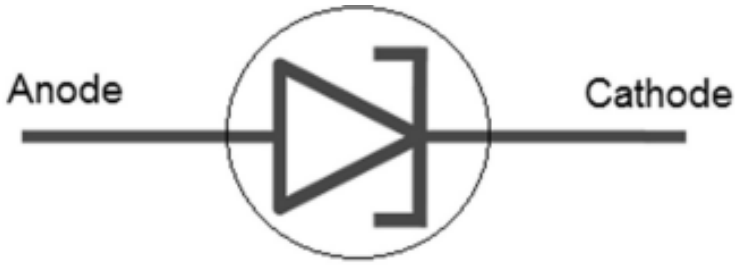


Figure 1.4 Symbol of Tunnel Diode

The anode is a positively charged electrode which attracts electrons whereas cathode is a negatively charged electrode which emits electrons. In tunnel diode, n-type semiconductor emits or produces electrons, so it is referred to as the cathode. On the other hand, p-type semiconductors attract electrons emitted from the n-type semiconductor, so p-type semiconductor is referred to as the anode.

Leo Esaki observed that if a semiconductor diode is heavily doped with impurities, it will exhibit negative resistance. Negative resistance means the current across the tunnel diode decreases when the voltage increases. In 1973 Leo Esaki received the Nobel Prize in physics for discovering the electron tunneling effect used in these diodes.

A tunnel diode is also known as Esaki diode which is named after Leo Esaki for his work on the tunneling effect. The operation of tunnel diode depends on the quantum mechanics principle known as “Tunneling”. In electronics, tunneling means a direct flow of electrons across the small depletion region from n-side conduction band into the p-side valence band.

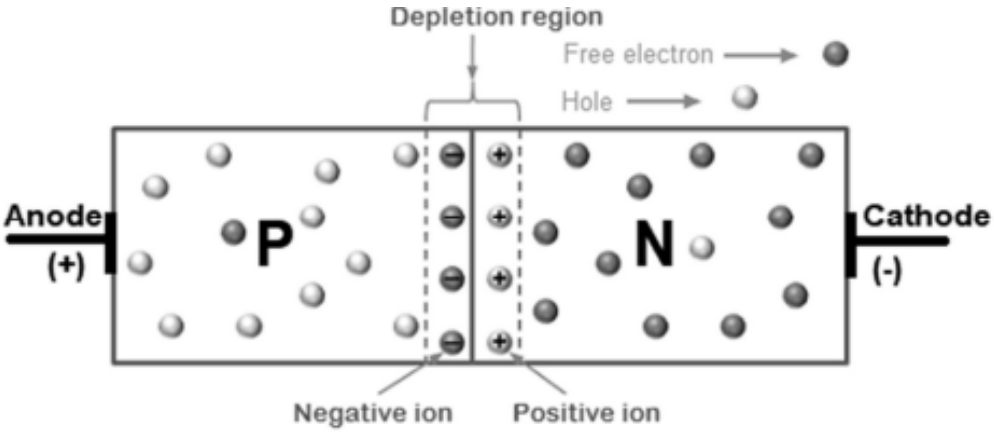


Figure 1.5 Construction of Tunnel Diode

Germanium material is commonly used to make tunnel diodes. They are also made from other types of materials such as gallium arsenide, gallium antimonide, and silicon.

1.3.2 Width of the Depletion Region in Tunnel Diode:

The depletion region is a region in a p-n junction diode where mobile charge carriers (free electrons and holes) are absent. Depletion region acts like a barrier that opposes the flow of electrons from the n-type semiconductor and holes from the p-type semiconductor.

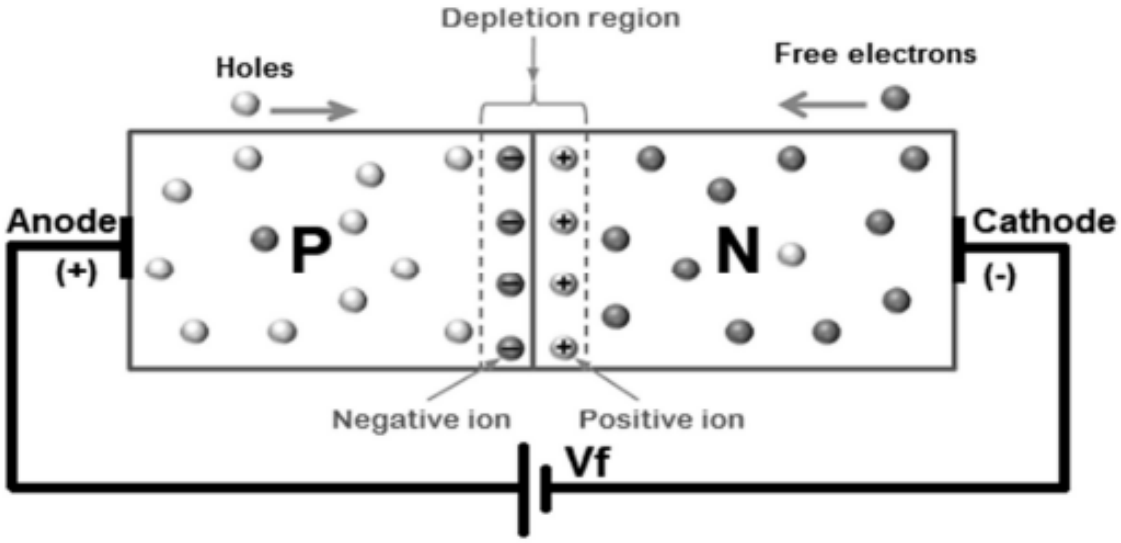


Figure 1.6 Working of Tunnel Diode

The width of a depletion region depends on the number of impurities added. Impurities are the atoms introduced into the p-type and n-type semiconductor to increase electrical conductivity. If a small number of impurities are added to the p-n junction diode (p-type and n-type semiconductor), a wide depletion region is formed. On the other hand, if large number of impurities are added to the p-n junction diode, a narrow depletion region is formed. In tunnel diode, the p-type and n-type semiconductor is heavily doped which means a large number of impurities are introduced into the p-type and n-type semiconductor. This heavy doping process produces an extremely narrow depletion region. The concentration of impurities in tunnel diode is 1000 times greater than the normal p-n junction diode. In normal p-n junction diode, the depletion width is large as compared to the tunnel diode. This wide depletion layer or depletion region in normal diode opposes the flow of current. Hence, depletion layer acts as a barrier.

To overcome this barrier, need to apply sufficient voltage. When sufficient voltage is applied, electric current starts flowing through the normal p-n junction diode. Unlike the normal p-n junction diode, the width of a depletion layer in tunnel diode is extremely narrow. So applying a small voltage is enough to produce electric current in tunnel diode. Tunnel diodes are capable of remaining stable for a long duration of time than the ordinary p-n junction diodes. They are also capable of high-speed operations.

1.4 PHOTO DIODE:

A photodiode is a diode working in reverse polarization and having a window where the light can enter and hit directly the p-n junction. As in the case of the LED, the energy level of the impurities has been chosen to allow electrons to jump from valence to conduction band. In the absence of light the leakage current is negligible, but when light is present, the leakage current increases to measurable values.

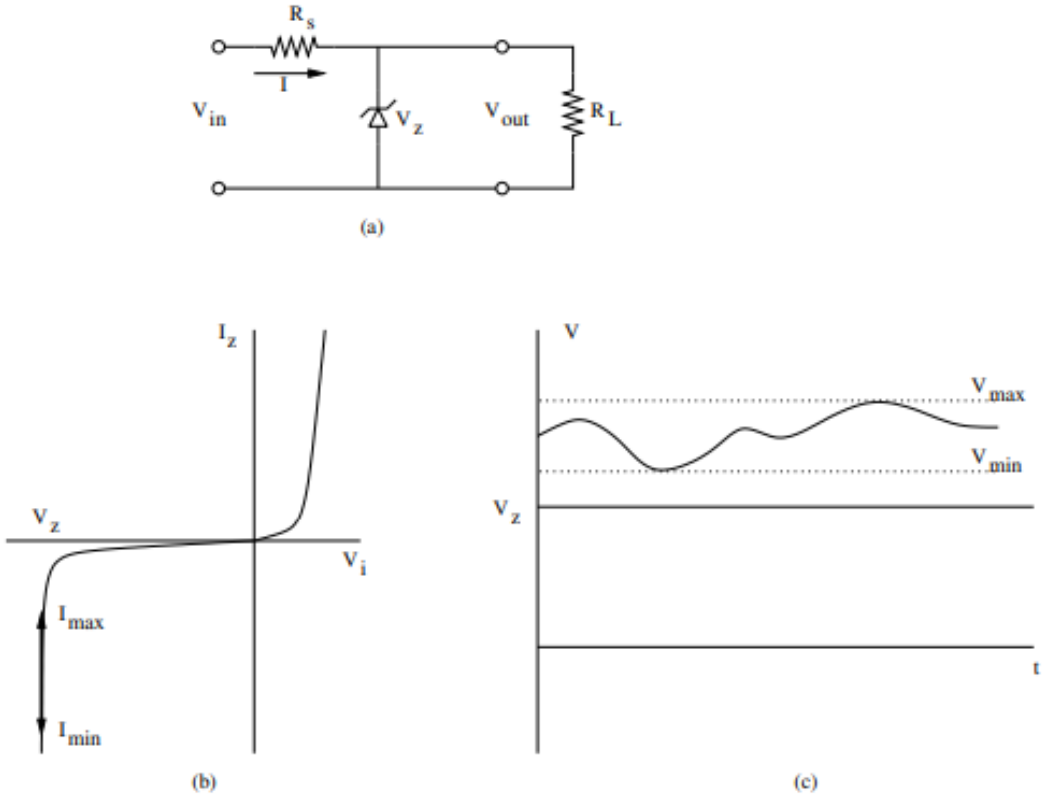


Figure 1.7: (a) A voltage regulator is inserted between the input voltage and the load. (b) The Zener current Iz varies between Imax and Imin in response to the varying input voltage so as to keep the load current and the load voltage constant (c).

1.5 TRANSISTORS:

The transistor is the main building block “element” of electronics. It is a semiconductor device and it comes in two general types: the Bipolar Junction Transistor (BJT) and the Field Effect Transistor (FET). Here we will describe the system characteristics of the BJT configuration and explore its use in fundamental signal shaping and amplifier circuits.

The BJT is a three terminal device and it comes in two different types. The np-n BJT and the p-np BJT. The BJT symbols and their corresponding block diagrams are shown on Figure. The BJT is fabricated with three separately doped regions. The np-n device has one p region between two n regions and the p-np device has one n region between two p regions.

The BJT has two junctions (boundaries between the n and the p regions). These junctions are similar to the junctions we saw in the diodes and thus they may be forward biased or reverse biased. By relating these junctions to a diode model the p-np BJT may be modeled as shown on Figure 1.8.

The three terminals of the BJT are called the Base (B), the Collector (C) and the Emitter (E).

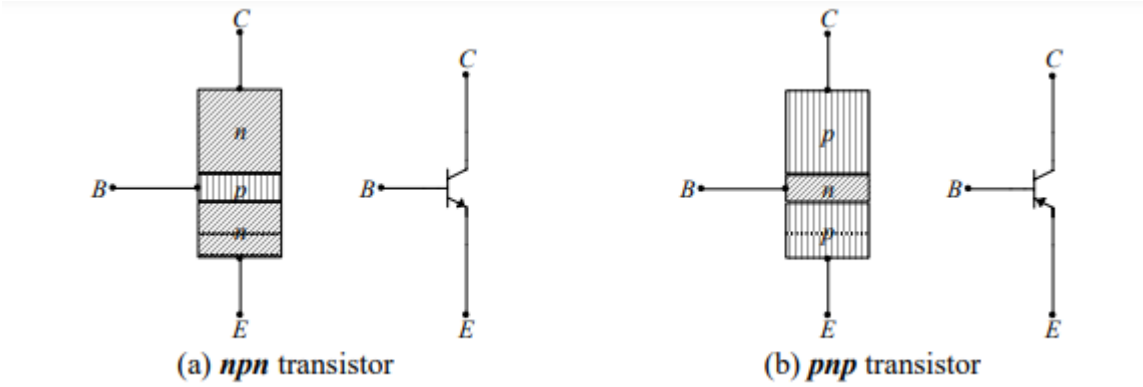


Figure 1.8 BJT schematics and structures. (a) np-n transistor, (b) p-np transistor

Since each junction has two possible states of operation (forward or reverse bias) the BJT with its two junctions has four possible states of operation. For a detailed description of the BJT structure see: Jaeger and Blalock, Microelectronic Circuit Design, McGraw Hill.

Here it is sufficient to say that the structure as shown on Figure .8 is not symmetric. The n and p regions are different both geometrically and in terms of the doping concentration of the regions. For example, doping concentrations in the collector, base and emitter may be, and respectively. Therefore, the behavior of the device is not electrically symmetric and the two ends cannot be interchanged.

Before proceeding let's consider the BJT np-n structure shown on Figure 1.9.

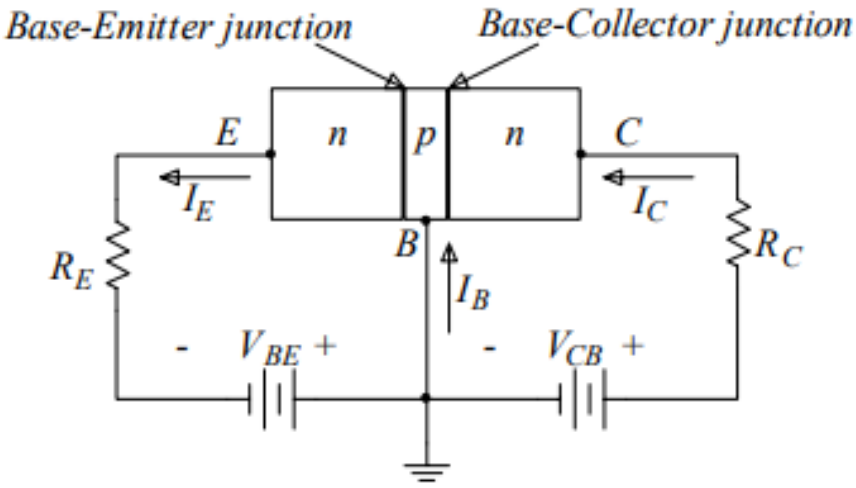


Figure 1.9 Biasing voltages of np-n transistor

1.6 SILICON CONTROLLED RECTIFIER:

A silicon *controlled rectifier is a semiconductor *device that acts as a true electronic switch. It can change alternating current into direct current and at the same time can control the amount of power fed to the load. Thus, SCR combines the features of a rectifier and a transistor.

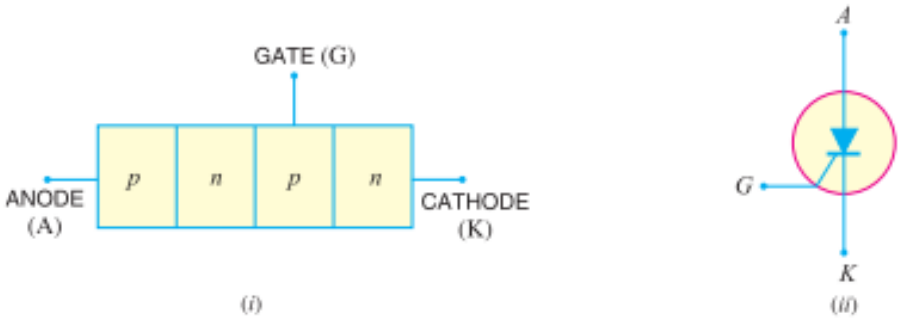


Figure 1.10 Silicon Controlled Rectifier

Constructional details: When a p-n junction is added to a junction transistor, the resulting three p-n junction device is called a silicon-controlled rectifier. Figure 1.10 (i) shows its construction. It is clear that it is essentially an ordinary rectifier (p-n) and a junction transistor (np-n) combined in one unit to form p-np-n device. Three terminals are taken; one from the outer p-type material called anode A, second from the outer n-type material called cathode K and the third from the base of transistor section and is called gate G. In the normal operating conditions of SCR, anode is held at high positive potential w.r.t. cathode and gate at small positive potential w.r.t. cathode. Figure 1.10 (ii) shows the symbol of SCR. The silicon-controlled rectifier is a solid state equivalent of thyratron. The gate, anode and cathode of SCR correspond to the grid, plate and cathode of thyratron. For this reason, SCR is sometimes called thyristor.

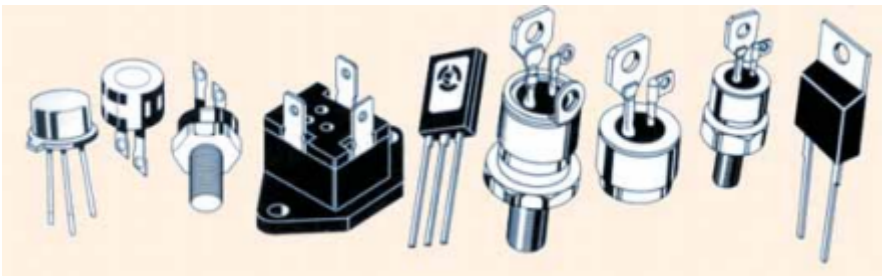


Figure 1.11 Typical SCR Packages

1.6.1 Working of SCR:

In a silicon-controlled rectifier, load is connected in series with anode. The anode is always kept at positive potential w.r.t. cathode. The working of SCR can be studied under the following two heads:

(i) When gate is open the SCR circuit with gate open i.e. no voltage applied to the gate. Under this condition, junction J2 is reverse biased while junctions J1 and J3 are forward biased. Hence, the situation in junctions J1 and J3 is just as in a np-n transistor with base open. Consequently, no current flows through the load RL and the SCR is cut off. However, if the applied voltage is gradually increased, a stage is reached when * reverse biased junction J2 breaks down. The SCR now conducts ** heavily and is said to be in the ON state. The applied voltage at which SCR conducts heavily without gate voltage is called Breakover voltage.

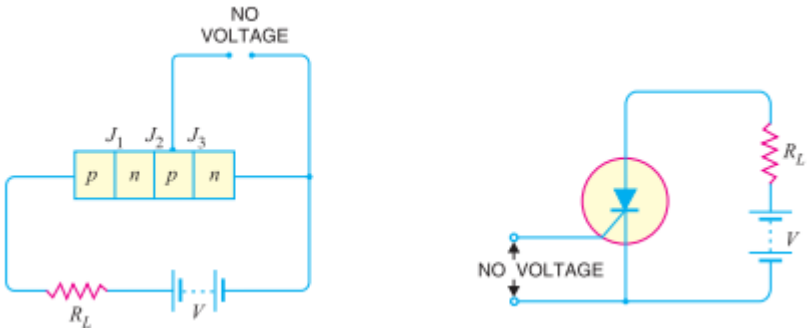


Figure 1.12 SCR Circuit with gate open

(ii) When gate is positive w.r.t. cathode. The SCR can be made to conduct heavily at smaller applied voltage by applying a small positive potential to the gate as shown in Figure 1.13. Now junction J3 is forward biased and junction J2 is reverse biased. The electrons from n-type material start moving across junction J3 towards left whereas holes from p-type towards the right. Consequently, the electrons from junction J3 are attracted across junction J2 and the current gate starts flowing. As soon as the gate current flows, anode current increases. The increased anode current in turn makes more electrons available at junction J2. This process continues and in an extremely small time, junction J2 breaks down and the SCR starts conducting heavily. Once SCR starts conducting, the gate (the reason for this name is obvious) loses all control. Even if gate voltage is removed, the anode current does not decrease at all. The only way to stop conduction (i.e. bring SCR in off condition) is to reduce the applied voltage to zero.

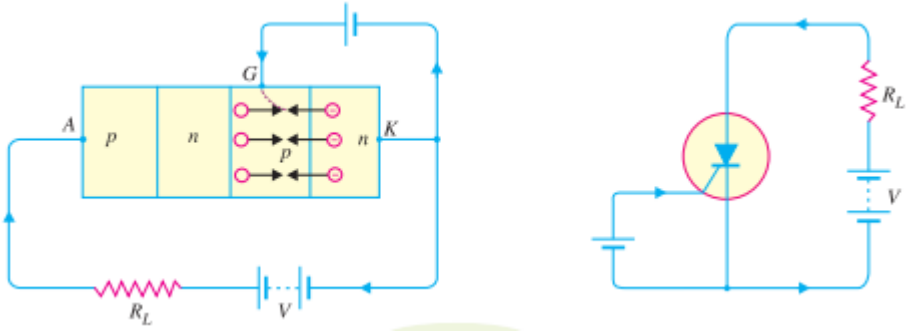


Figure 1.13 Applying small potential to the gate

1.7 SUMMARY:

This study provides an in-depth analysis of semiconductor devices, including diodes, transistors, and silicon-controlled rectifiers (SCRs), focusing on their structure, working principles, and applications. It covers various types of diodes such as p-n junction diodes, tunnel diodes, and photodiodes, explaining their role in rectification, amplification, and switching. The study highlights the impact of doping on the depletion region and how it influences device performance. It also explores the function of transistors, specifically bipolar junction transistors (BJTs), in signal processing and amplification. Additionally, the study examines the operation of SCRs in power control applications, explaining their ability to switch and regulate electrical power efficiently. Special emphasis is given to the tunneling effect in tunnel diodes and the photoelectric effect in photodiodes. Understanding these semiconductor devices is crucial for advancements in modern electronics, as they form the foundation of circuits used in communication, computing, and power management systems.

1.8 TECHNICAL TERMS:

Diodes, Tunnel diode, Photo diode, Transistor and Silicon controlled rectifier.

1.9 SELF-ASSESSMENT QUESTIONS:

Essay Question:

1) What is the working principle of a p-n junction diode under forward and reverse bias conditions?
2) How does a tunnel diode differ from a conventional p-n junction diode, and what is the significance of the tunneling effect?
3) What are the main differences between a Bipolar Junction Transistor (BJT) and a Field Effect Transistor (FET)?

Short Notes:

1) How does a photodiode operate, and what role does light play in its functionality?
2) What is the structure and working principle of a Silicon Controlled Rectifier (SCR), and how does it control power?
3) Why is doping important in semiconductor devices, and how does it affect the depletion region in different diodes?

1.10 SUGGESTED READINGS:

1) "Solid State Electronic Devices" – Ben G. Streetman and Sanjay Kumar Banerjee.A fundamental book covering semiconductor physics and various electronic devices like diodes, transistors, and SCRs.

2) "Microelectronic Circuits" – Adel S. Sedra and Kenneth C. Smith. A comprehensive book explaining the design and analysis of electronic circuits using semiconductor devices.

3) "Semiconductor Physics and Devices" – Donald A. Neamen. Detailed insights into the physics of semiconductor materials and devices, including p-n junctions, transistors, and tunnel diodes.

4) "Electronic Devices and Circuit Theory" – Robert L. Boylestad and Louis Nashelsky. A widely used textbook covering the fundamentals of diodes, BJTs, FETs, and other semiconductor devices.

5) "Power Electronics: Circuits, Devices, and Applications" – Muhammad H. Rashid. A great resource for understanding power semiconductor devices like SCRs, thyristors, and their applications in power control.

Prof. Sandhya Cole

LESSON-2

EVOLUTION AND ADVANCEMENTS IN SEMICONDUCTOR DEVICES

2.0 AIM AND OBJECTIVES:

The aim of this document is to provide a comprehensive understanding of semiconductor devices, focusing on Unijunction Transistors (UJT), Field Effect Transistors (FET), Junction Field Effect Transistors (JFET), Metal Oxide Semiconductor Field Effect Transistors (MOSFET), and Complementary Metal Oxide Semiconductor (CMOS) technology. The objective is to explain their construction, working principles, characteristics, and applications in modern electronics. The document explores the fundamental differences between BJTs and FETs, highlighting the advantages of FETs, such as high input impedance and low noise. It also discusses the significance of MOSFETs in integrated circuits and digital logic design, emphasizing the role of CMOS technology in power-efficient semiconductor applications. By detailing these semiconductor components, this document aims to serve as a valuable resource for students, researchers, and professionals seeking a deeper understanding of transistor technology and its impact on electronic circuit design.

STRUCTURE:

2.1 UNI JUNCTION TRANSISTOR:

Unijunction transistor (abbreviated as UJT), also called the double-base diode is a 2- layer, 3-terminal solid-state (silicon) switching device. The device has a unique characteristic that when it is triggered, its emitter current increases re generatively (due to negative resistance

characteristic) until it is restricted by emitter power supply. Since the device has one p n junction and three leads p n junction and three leads

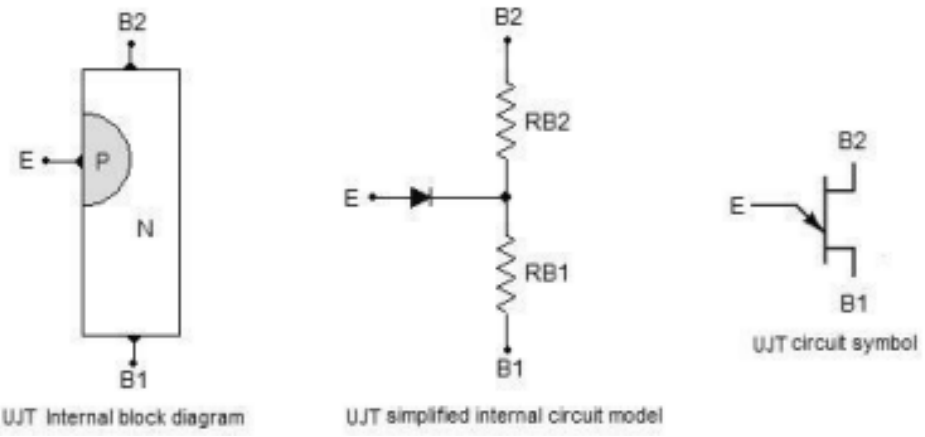


Figure 2.1 UJT Structure, Equivalent Circuit and Symbol

2.1.1 Construction of a UJT:

The basic structure of a unijunction transistor is shown in figure. It essentially consists of a lightly-doped N-type silicon bar with a small piece of heavily doped P-type material alloyed to its one side to produce single P-N junction. The single P-N junction accounts for the terminology unijunction. The silicon bar, at its ends, has two ohmic contacts designated as base-1 (B1) and base-2 (B2), as shown and the P-type region is termed the emitter (E). The emitter junction is usually located closer to base-2 (B2) than base-1 (B1) so that the device is not symmetrical, because symmetrical unit does not provide optimum electrical characteristics for most of the applications.

The symbol for unijunction transistor is shown in figure 2.1. The emitter leg is drawn at an angle to the vertical line representing the N-type material slab and the arrowhead points in the direction of conventional current when the device is forward-biased, active or in the conducting state. The basic arrangement for the UJT is shown in figure. A complementary UJT is formed by diffusing an N-type emitter terminal on a P-type base. Except for the polarities of voltage and current, the characteristics of a complementary UJT are exactly the same as those of a conventional UJT.

- The device has only one junction, so it is called the unijunction device.
- The device, because of one P-N junction, is quite similar to a diode but it differs from an ordinary diode as it has three terminals.
- The structure of a UJT is quite similar to that of an N-channel JFET. The main difference is that P-type (gate) material surrounds the N-type (channel) material in case of JFET and the gate surface of the JFET is much larger than emitter junction of UJT.

- In a unijunction transistor the emitter is heavily doped while the N-region is lightly doped, so the resistance between the base terminals is relatively high, typically 4 to 10 kilo Ohm when the emitter is open.
- The N-type silicon bar has a high resistance and the resistance between emitter and base-1is larger than that between emitter and base-2. It is because emitter is closer to base-2 than base-1.
- UJT operates with emitter junction forward- biased while the JFET is normally operated with the gate junction reverse-biased.
- UJT does not have ability to amplify but it has the ability to control a large ac power with a small signal. It exhibits a negative resistance characteristic and so it can be employed as an oscillator.

2.1.2 UJT Characteristics:

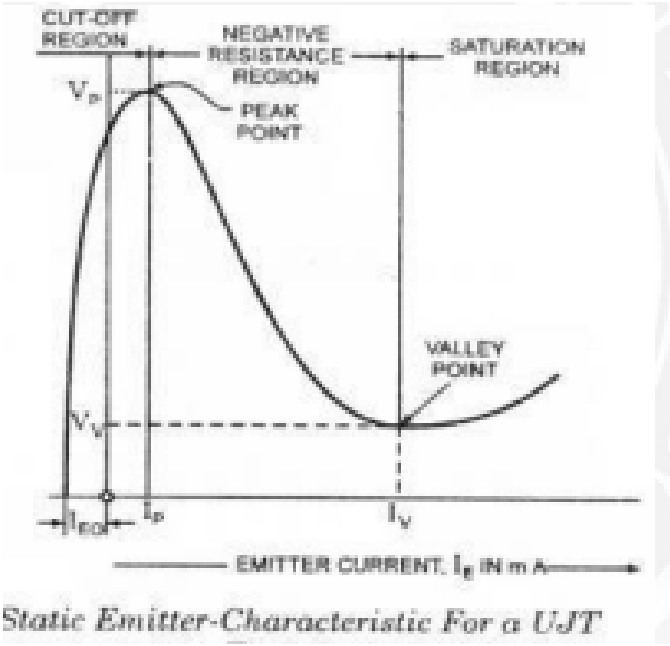


Figure 2.2 Static Emitter Characteristic for a UJT

The static emitter characteristic (a curve showing the relation between emitter voltage VE and emitter current IE) of a UJT at a given inter base voltage VBB is shown in figure. From figure it is noted that for emitter potentials to the left of peak point, emitter current IE never exceedsIEo. The current IEo corresponds very closely to the reverse leakage current ICo of the conventional BJT. This region, as shown in the figure, is called the cut-off region. Once conduction is established at VE = VP the emitter potential VE starts decreasing with the increase in emitter current IE. This Corresponds exactly with the decrease in resistance RB for increasing current IE. This device, therefore, has a negative resistance region which is stable enough to be used with a great

deal of reliability in the areas of applications listed earlier. Eventually, the valley point reaches, and any further increase in emitter current IE places the device in the saturation region, as shown in the figure 2.2. Three other important parameters for the UJT are IP, VV and IV and are defined below:

a) **Peak-Point Emitter Current Ip**: It is the emitter current at the peak point. It represents the minimum current that is required to trigger the device (UJT). It is inversely proportional to the inter base voltage VBB.
b) **Valley Point Voltage V:** The valley point voltage is the emitter voltage at the valley point. The valley voltage increases with the increase in inter base voltage VBB.
c) **Valley Point Current IV:** The valley point current is the emitter current at the valley point. It increases with the increase in inter-base voltage VBB.

2.1.3 Applications of UJT:

- Relaxation oscillators.
- Switching Thyristors like SCR, TRIAC etc.
- Magnetic flux sensors.
- Voltage or current limiting circuit.
- Bistable oscillators.
- Voltage or current regulators.
- Phase control circuits.

2.2 FIELD EFFECT TRANSISTOR:

The field effect transistor (FET) has, by virtue of its construction and biasing, large input impedance which may be more than 100 megaohms. The FET is generally much less noisy than the ordinary or bipolar transistor. The rapidly expanding FET market has led many semiconductor marketing managers to believe that this device will soon become the most important electronic device, primarily because of its integrated-circuit applications. In this chapter, we shall focus our attention on the construction, working and circuit applications of field effect transistors.

2.2.1 Types of Field Effect Transistors:

A bipolar junction transistor (BJT) is a current controlled device i.e., output characteristics of the device are controlled by base current and not by base voltage. However, in a field effect transistor (FET), the output characteristics are controlled by

input voltage (i.e., electric field) and not by input current. This is probably the biggest difference between BJT and FET. There are two basic types of field effect transistors:

(i) Junction field effect transistor (JFET) (ii) Metal oxide semiconductor field effect transistor (MOSFET) To begin with, we shall study about JFET and then improve form of JFET, namely; MOSFET.

2.3 JFET:

A junction-field effect transistor is a three-termsemi-conductor device in which current conduction is by one type of carrier i.e., electrons or holes.

The JFET was developed at about the same time as the transistor, but it came into general use only in the late 1960s. In a JFET, the current conduction is either by electrons or holes and is controlled by means of an electric field between the gate electrode and the conducting channel of the device. The JFET has high input impedance and low noise level. Constructional details. A JFET consists of a p-type or n-type silicon bar containing two pn junctions at the sides as shown in Figure 2.3. The bar forms the conducting channel for the charge carriers. If the bar is of n-type, it is called n-channel JFET as shown in Figure2.3 (i) and if the bar is of p-type, it is called a p-channel JFET as shown in Figure2.3 (ii). The two pn junctions forming diodes are connected *internally and a common terminal called gate is taken out. Other terminals are source and drain taken out from the bar as shown. Thus, a JFET has essentially three terminals viz., gate (G), source (S) and drain (D).

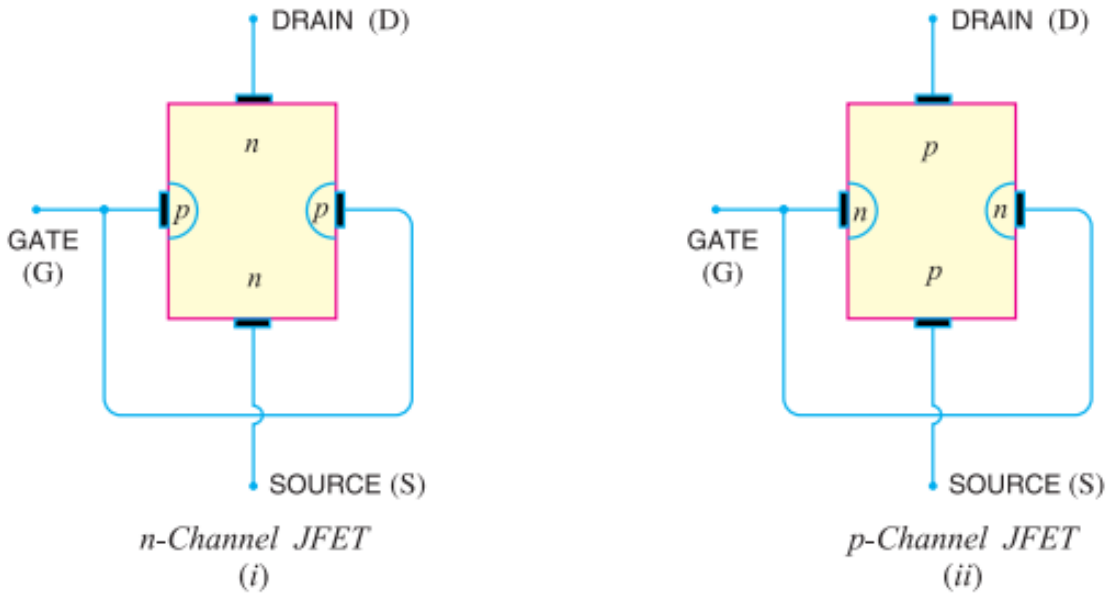


Figure 2.3: JFET Channels

2.3.1 Principle and Working of JFET:

Figure 2.4 shows the circuit of n-channel JFET with normal polarities. Note that the gate is reverse biased.

a. Principle. The two pn junctions at the sides form two depletion layers. The current conduction by charge carriers (i.e. free electrons in this case) is through the channel between the two depletion layers and out of the drain. The width and hence *resistance of this channel can be controlled by changing the input voltage VGS. The greater the reverse voltage of VGS, the wider the depletion layers will be and the narrowerthe conducting channel will be. The narrower channel means greater resistance and hence sources to drain current decreases. The reverse will happen should VGS decrease. Thus, JFET operates on the principle that width and hence resistance of the conducting channel can be varied by changing the reverse voltage VGS. In other words, the magnitude of drain current (ID) can be changed by altering VGS.

Working: The working of JFET is as under:

(i) When a voltage VDS is applied between drain and source terminals and voltage on the gate is zero [See Figure 2.4 (i)], the two pn junctions at the sides of the bar establish depletion layers. The electrons will flow from source to drain through a channel between the depletion layers. The size of these layers determines the width of the channel and hence the current conduction through the bar.

(ii) When a reverse voltage VGS is applied between the gate and source [See Figure 2.4 (ii)], the width of the depletion layers is increased. This reduces the width of conducting channels, thereby increasing the resistance of n-type bar. Consequently, the current from source to drain is decreased. On the other hand, if the reverse voltage on the gate is decreased, the width of the depletion layers also decreases. This increases the width of the conducting channel and hence sources drain current.

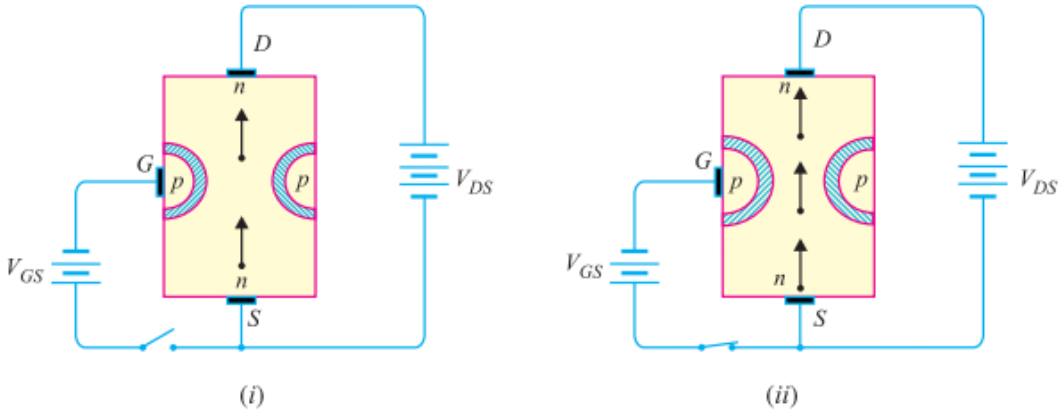


Figure 2.4: Principle and Working of JFET

It is clear from the above discussion that current from source to drain can be controlled by the application of potential (i.e. electric field) on the gate. For this reason, the device is called a field effect transistor. It may be noted that a p-channel JFET operates in the same manner as an n -channel JFET except that channel current carriers will be the holes instead of electrons and the polarities of VGS and VDS are reversed.

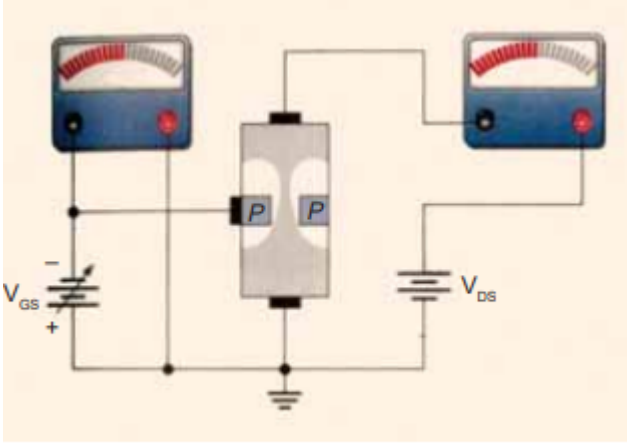


Figure 2.5 JFET biased for Conduction

2.3.2 Schematic Symbol of JFET:

Figure 1.5 shows the schematic symbol of JFET. The vertical line in the symbol may be thought of as channel and source (S) and drain (D) connected to this line. If the channel is n-type, the arrow on the gate points towards the channel as shown in Figure 2.6 (i). However, for p-type channel, the arrow on the gate points from channel to gate [See Figure 2.6 (ii)].

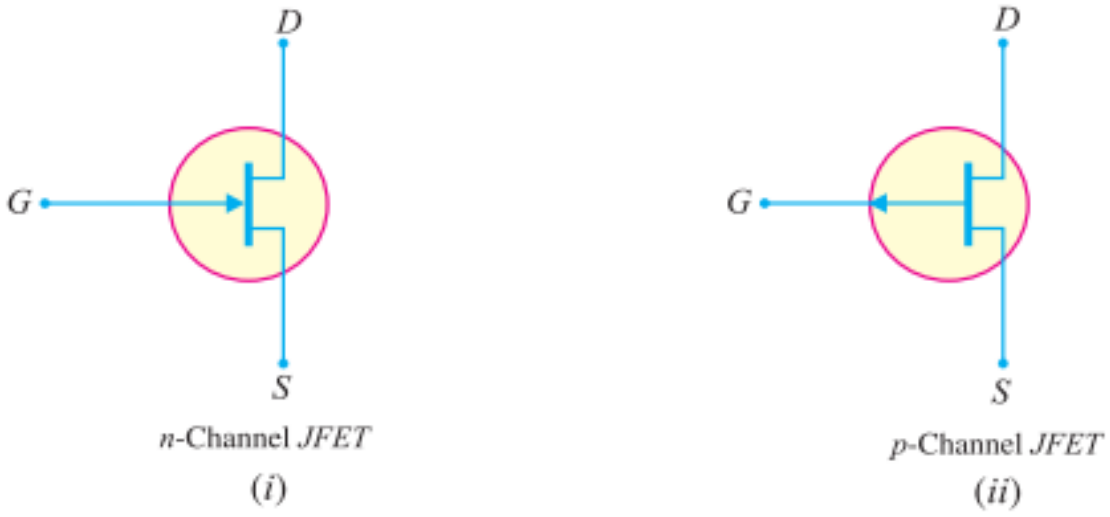


Figure 2.6 Schematic Symbol of JFET

2.4 MOSFET:

MOSFET's operate the same as JFET's but have a gate terminal that is electrically isolated from the conductive channel.

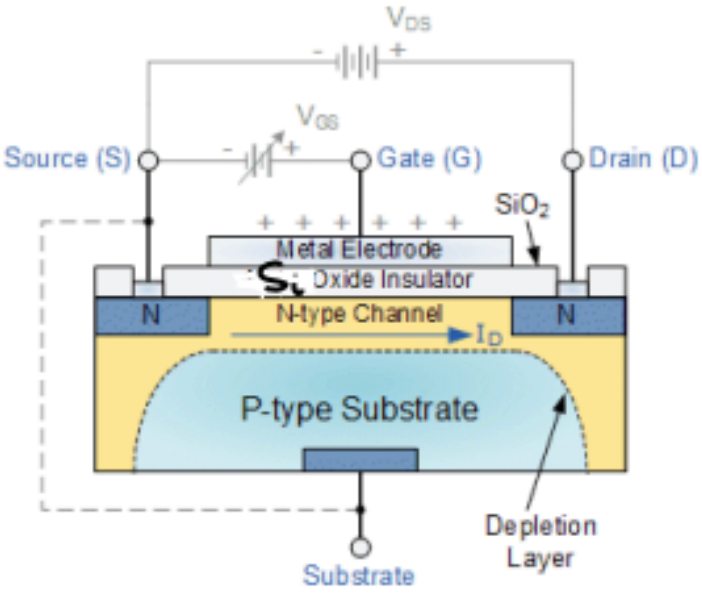


Figure 2.7: JFET Gate Terminal

As well as the Junction Field Effect Transistor (JFET), there is another type of Field Effect Transistor available whose Gate input is electrically insulated from the main current carrying channel and is therefore called an Insulated Gate Field Effect Transistor. The most common type of insulated gate FET which is used in many different types of electronic circuits is called the Metal Oxide Semiconductor Field Effect Transistor or MOSFET for short.

The IGFET or MOSFET is a voltage-controlled field effect transistor that differs from a JFET in that it has a "Metal Oxide" Gate electrode which is electrically insulated from the main semiconductor n-channel or p-channel by a very thin layer of insulating material usually silicon dioxide, commonly known as glass. This ultra-thin insulated metal gate electrode can be thought of as one plate of a capacitor. The isolation of the controlling Gate makes the input resistance of the MOSFET extremely high way up in the Mega-ohms (MΩ) region thereby making it almost infinite.

As the Gate terminal is electrically isolated from the main current carrying channel between the drain and source, "NO current flows into the gate" and just like the JFET, the MOSFET also acts like a voltage-controlled resistor where the current flowing through the main channel between the Drain and Source is proportional to the input voltage. Also like the JFET, the MOSFETs' very high input resistance can easily accumulate large amounts of static charge resulting in the MOSFET becoming easily damaged unless carefully handled or protected.

Like the previous JFET tutorial, MOSFETs are three terminal devices with a Gate, Drain and Source and both P-channel (PMOS) and N-channel (NMOS) MOSFETs are available. The main difference this time is that MOSFETs are available in two basic forms:

Depletion Type – the transistor requires the Gate-Source voltage, (V) to switch the device "OFF". The depletion mode MOSFET is equivalent to a "Normally Closed" switch.

Enhancement Type – the transistor requires a Gate-Source voltage, (V) to switch the device "ON". The enhancement mode MOSFET is equivalent to a "Normally Open" switch.

The symbols and basic construction for both configurations of MOSFETs are shown below.

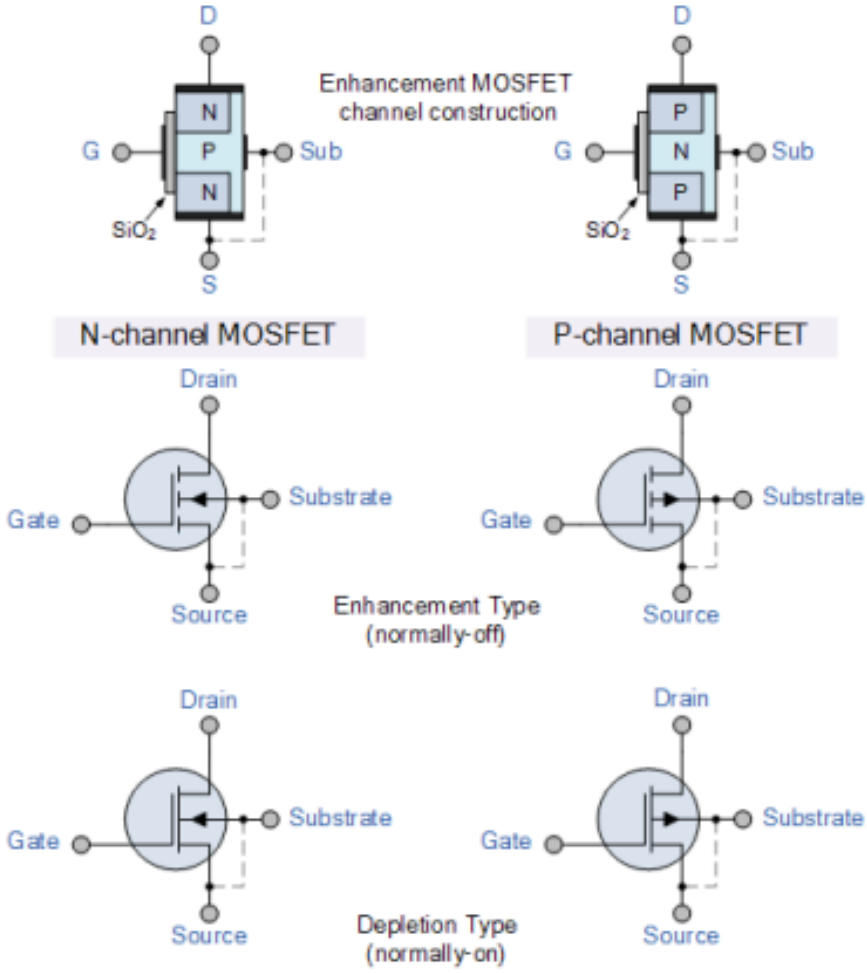


Figure 2.8: The Symbols and Basic Construction for Both Configurations of MOSFETs

The four MOSFET symbols above show an additional terminal called the Substrate and is not normally used as either an input or an output connection but instead it is used for grounding the substrate. It connects to the main semiconductive channel through a diode junction to the body or metal tab of the MOSFET.

2.4.1 Basic MOSFET Structure and Symbol:

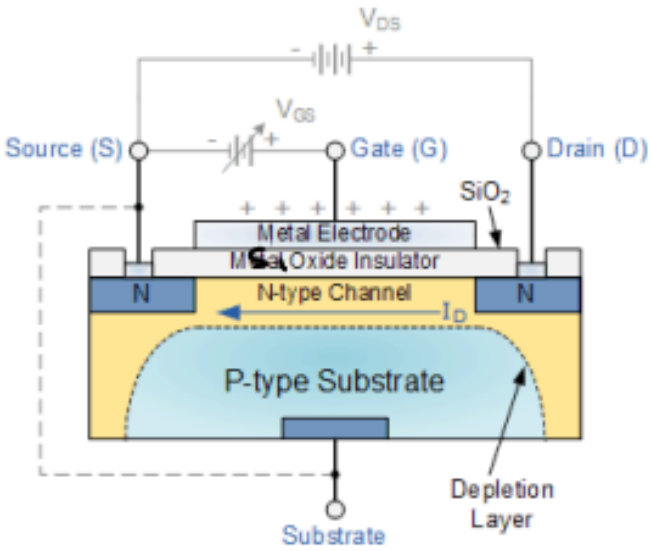


Figure 2.9 Basic MOSFET structure and Symbol

The construction of the Metal Oxide Semiconductor FET is very different to that of the Junction FET. Both the Depletion and Enhancement type MOSFETs use an electrical field produced by a gate voltage to alter the flow of charge carriers, electrons for n-channel or holes for P-channel, through the semiconductive drain-source channel. The electrode gate is placed on top of a very thin insulating layer and there are a pair of small n-type regions just under the drain and source electrodes. We saw in the previous tutorial that the gate of a junction field effect transistor, JFET must be biased in such a way as to reverse-bias the pn-junction.

With a insulated gate MOSFET device no such limitations apply so it is possible to bias the gate of a MOSFET in either polarity, positive (+ve) or negative (-ve). This makes the MOSFET device especially valuable as electronic switches or to make logic gates because with no bias they are normally non-conducting, and this high gate input resistance means that very little or no control current is needed as MOSFETs are voltage-controlled devices. Both the p-channel and the n-channel MOSFETs are available in two basic forms, the Enhancement type and the Depletion type.

2.5 CMOS:

2.5.1 CMOS Working Principle and Applications:

The term CMOS stands for “Complementary Metal Oxide Semiconductor”. CMOS technology is one of the most popular technologies in the computer chip design industry and broadly used today to form integrated circuits in numerous and varied applications. Today’s

computer memory, CPUs and cell phones make use of this technology due to several key advantages. This technology makes use of both P channel and N channel semiconductor devices.

One of the most popular MOSFET technologies available today is the Complementary MOS or CMOS technology. This is the dominant semiconductor technology for microprocessors, microcontroller chips, memory like RAM, ROM, EEPROM and application specific integrated circuits (ASICs).

2.5.2 CMOS (Complementary Metal Oxide Semiconductor):

The main advantage of CMOS over NMOS and BIPOLAR technology is the much smaller power dissipation. Unlike NMOS or BIPOLAR circuits, a Complementary MOS circuit has almost no static power dissipation. Power is only dissipated in case the circuit actually switches. This allows integrating more CMOS gates on an IC than in NMOS or bipolar technology, resulting in much better performance. Complementary Metal Oxide Semiconductor transistor consists of P-channel MOS (PMOS) and Nchannel MOS (NMOS). Please refer the link to know more about the fabrication process of CMOS transistor.

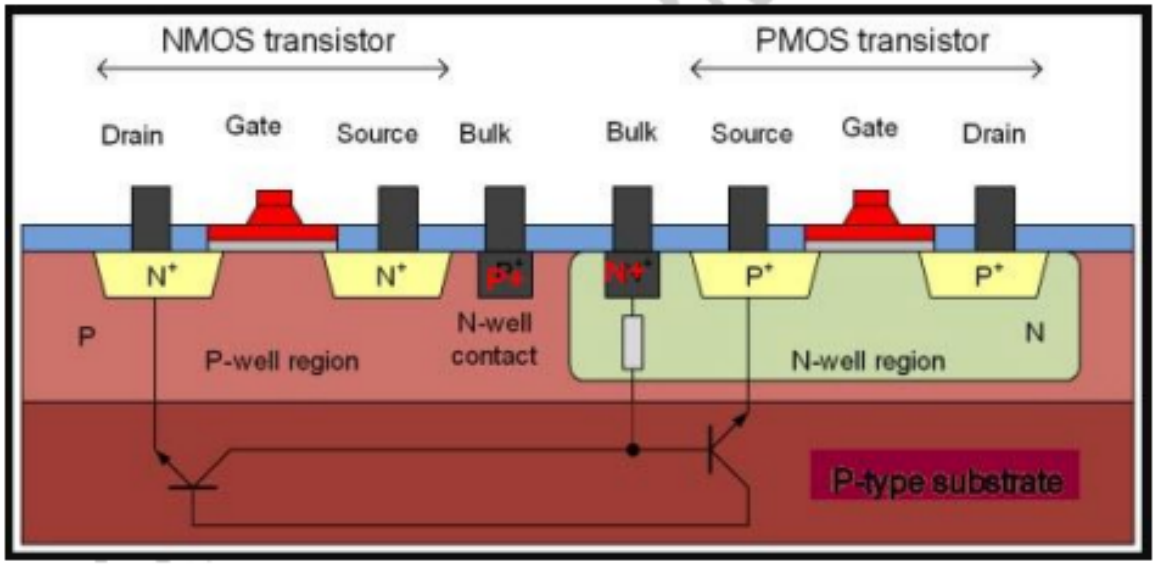


Figure 2.10: NMOS Transistor and PMOS Transistor

2.5.2.1 NMOS:

NMOS is built on a p-type substrate with n-type source and drain diffused on it. In NMOS, the majority carriers are electrons. When a high voltage is applied to the gate, the NMOS will conduct. Similarly, when a low voltage is applied to the gate, NMOS will not conduct. NMOS are faster than PMOS, since the carriers in NMOS, which are electrons, travel twice as fast as the holes.

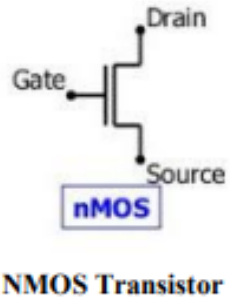


Figure 2.11 NMOS Transistor

2.5.2.2 PMOS:

P-channel MOSFET consists of P-type Source and Drain diffused on an N-type substrate. Majority carriers are holes. When a high voltage is applied to the gate, the PMOS will not conduct. When a low voltage is applied to the gate, the PMOS will conduct. The PMOS devices are more immune to noise than NMOS devices.

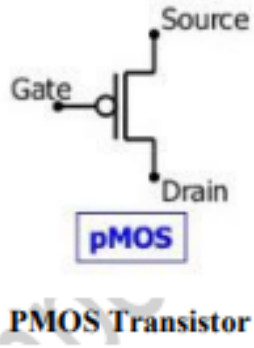


Figure 2.12 PMOS Transistor

2.5.3 CMOS Working Principle:

In CMOS technology, both N-type and P-type transistors are used to design logic functions. The same signal which turns ON a transistor of one type is used to turn OFF a transistor of the other type. This characteristic allows the design of logic devices using only simple switches, without the need for a pullup resistor. In CMOS logic gates a collection of n-type MOSFETs is arranged in a pull-down network between the output and the low voltage power supply rail (Vss or quite often ground).

Instead of the load resistor of NMOS logic gates, CMOS logic gates have a collection of p-type MOSFETs in a pull-up network between the output and the higher-voltage rail (often named Vdd). Thus, if both a p-type and n-type transistor have their gates connected to the same input, the p-type MOSFET will be ON when the n-type MOSFET is OFF, and vice-versa. The networks are arranged such that one is ON and the other OFF for any input pattern as shown in the figure below.

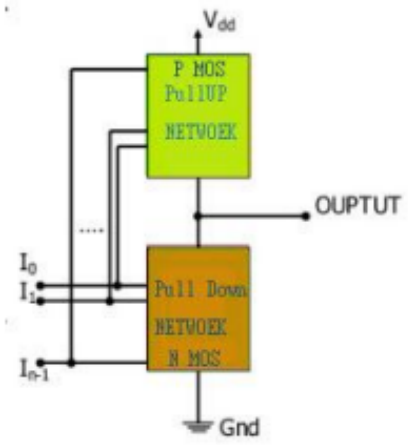


Figure 2.13: MOSFET Arranged Networks

2.5.4 CMOS Logic Gate using Pull-Up and Pull-Down Networks:

CMOS offers relatively high speed, low power dissipation, high noise margins in both states, and will operate over a wide range of source and input voltages (provided the source voltage is fixed). Furthermore, for the better understanding of the Complementary Metal Oxide Semiconductor working principle, we need to discuss in brief about CMOS logic gates as explained below.

2.6 SUMMARY:

This document provides a detailed overview of semiconductor devices, focusing on Unijunction Transistors (UJT), Field Effect Transistors (FET), Junction Field Effect Transistors (JFET), Metal Oxide Semiconductor Field Effect Transistors (MOSFET), and Complementary Metal Oxide Semiconductor (CMOS) technology. It explains their construction, working principles, characteristics, and applications in modern electronics. The document highlights the differences between Bipolar Junction Transistors (BJT) and FETs, emphasizing FETs' advantages such as high input impedance and low noise. It also explores the critical role of MOSFETs in integrated circuits and digital logic, particularly in CMOS technology, which enables power-efficient semiconductor applications. By covering these essential semiconductor components, this document serves as a valuable resource for students, researchers, and professionals interested in electronic circuit design. Understanding these devices is crucial for advancements in microelectronics, computing, and communication technologies, making this study an essential foundation for future innovations in the semiconductor industry.

2.7 TECHNICAL TERMS:

JFET, UJT, FET, CMOS, MOSFET

2.8 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are the key differences between Bipolar Junction Transistors (BJT) and Field Effect Transistors (FET)?
2) How does Unijunction Transistor (UJT) function, and what are its primary applications?
3) What are the advantages of using a Junction Field Effect Transistor (JFET) over a BJT?

Short Answer Questions:

1) How does a Metal Oxide Semiconductor Field Effect Transistor (MOSFET) operate, and why is it widely used in modern electronics?
2) What is Complementary Metal Oxide Semiconductor (CMOS) technology, and how does it contribute to power efficiency in digital circuits?
3) Why is understanding semiconductor devices essential for advancements in microelectronics and communication technologies?

2.9 SUGGESTED READINGS:

1) "Semiconductor Physics and Devices" – Donald A. Neamen.
2) "Solid State Electronic Devices" – Ben G. Streetman & Sanjay Kumar Banerjee.
3) "Microelectronic Circuits" – Adel S. Sedra & Kenneth C. Smith.
4) "Fundamentals of Microelectronics" – Behzad Razavi.
5) "Physics of Semiconductor Devices" – Simon M. Sze & Kwok K. Ng.
6) "Electronic Devices and Circuit Theory" – Robert L. Boylestad & Louis Nashelsky.
7) "Introduction to Semiconductor Devices" – Kevin F. Brennan.

Prof. Sandhya Cole

LESSON-3

OPTO ELECTRONIC DEVICES

3.0 AIM AND OBJECTIVES:

The aim of this study is to explore the working principles, construction, and materials used in solar cells, photo detectors, and LEDs, focusing on their significance in energy conversion and optical communication. Solar cells, which convert light energy into electrical energy through the photovoltaic effect, are primarily made using semiconductor materials like silicon, GaAs, CdTe, and $CuInSe_2$, chosen for their suitable band gaps. Photodetectors, particularly PIN and avalanche photodiodes (APD), play a crucial role in fiber-optic communication by converting light signals into electrical signals with high sensitivity. LEDs, operating on the principle of electroluminescence, emit visible or infrared light based on the material used, with applications ranging from display systems to optical sensors. Understanding the relationship between material properties and device performance is essential for optimizing efficiency in renewable energy generation, optical data transmission, and solid-state lighting. This study aims to provide a comprehensive understanding of these semiconductor devices, highlighting their construction, functionality, and advancements that contribute to energy-efficient technologies and high-speed communication systems.

STRUCTURE:

3.1 SOLAR CELLS:

A solar cell (also known as a photovoltaic cell or PV cell) is defined as an electrical device that converts light energy into electrical energy through the photovoltaic effect. A solar cell is basically a p-n junction diode. Solar cells are a form of photoelectric cell, defined as a device whose electrical characteristics – such as current, voltage, or resistance – vary when exposed to light.

Individual solar cells can be combined to form modules commonly known as solar panels. The common single junction silicon solar cell can produce a maximum opencircuit voltage of approximately 0.5 to 0.6 volts. By itself this isn't much – but remember these solar cells are tiny. When combined into a large solar panel, considerable amounts of renewable energy can be generated.

3.1.1 Construction of Solar Cell:

A solar cell is basically a junction diode, although its construction is little bit different from conventional p-n junction diodes. A very thin layer of p-type semiconductor is grown on a relatively thicker n-type semiconductor. We then apply a few finer electrodes on the top of the p-type semiconductor layer.

These electrodes do not obstruct light to reach the thin p-type layer. Just below the ptype layer there is a p-n junction. We also provide a current collecting electrode at the bottom of the n-type layer. We encapsulate the entire assembly by thin glass to protect the solar cell from any mechanical shock.

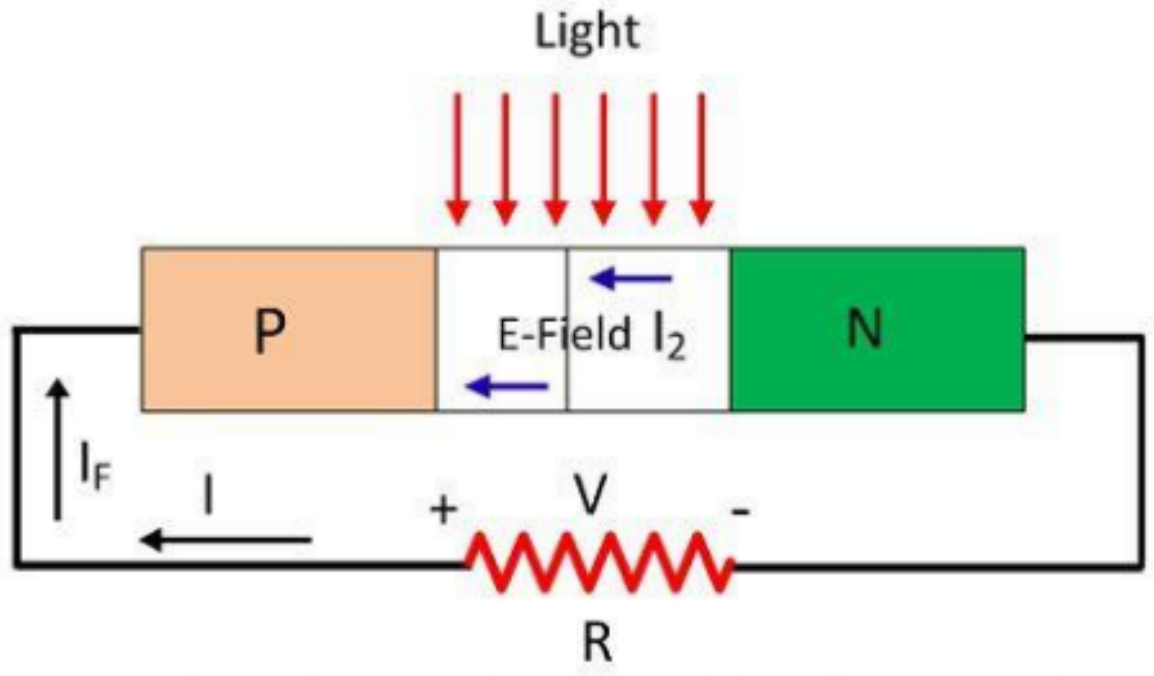


Figure 3.1 P-N Junction Solar Cell with Resistive Load

3.1.2 Working Principle of Solar Cell:

When light reaches the p-n junction, the light photons can easily enter in the junction, through a very thin p-type layer. The light energy, in the form of photons, supplies sufficient energy to the junction to create several electron-hole pairs. The incident light breaks the thermal equilibrium condition of the junction. The free electrons in the depletion region can quickly come to the n-type side of the junction.

Similarly, the holes in the depletion can quickly come to the p-type side of the junction. Once, the newly created free electrons come to the n-type side, cannot further cross the junction because of barrier potential of the junction.

Similarly, the newly created holes once come to the p-type side cannot further cross the junction became of same barrier potential of the junction. As the concentration of electrons become higher in one side, i.e. n-type side of the junction and concentration of holes becomes more in another side, i.e. the p-type side of the junction, the p-n junction will behave like a small battery cell. A voltage is set up which is known as photo voltage. If we connect a small load across the junction, there will be a tiny current flowing through it.

3.1.3 Materials Used in Solar Cell:

The materials which are used for this purpose must have a band gap close to 1.5ev. The commonly used materials are

1) Silicon
2) GaAs
3) CdTe
4) CuInSe2

3.2 PHOTO DETECTORS:

- A photo detector is a device which converts the light signals to electrical signals.
- The two main photo detectors used for fiber communication system are,
  1) PIN photo diode
  2) Avalanche photo diode (APD)

3.2.1 Pin Photo Diode:

3.2.1.1 Principle:

- Reverse bias is applied to a diode.
- Light is allowed to fall on the neutral (or) intrinsic‘i’ region, electron hole pairs are generated.
- These electrons and holes are accelerated by the external electric field, which results in photo current.
- Thus, light is converted into electrical signals.

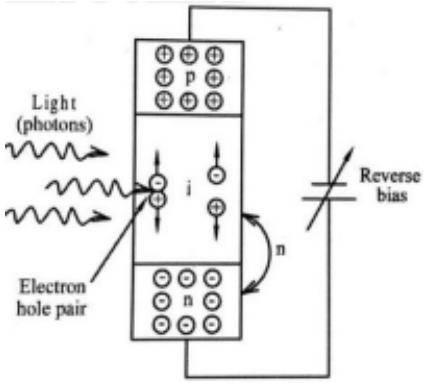


Figure 3.2: PIN Photo Diode

3.2.1.2 Construction:

1) It consists of 'p' and 'n' regions separated by an intrinsic region 'i'.
2) 'p' and 'n' regions are made by silicon, germanium and their alloys and they are heavily doped.
3) i' region is lightly doped by 'n' material and made as large as possible to absorb more photons.
4) So, it is called as positive-intrinsic-negative (PIN) photo diode.

3.2.1.3 Working:

1) When a high reverse bias voltage is applied, the width of the depletion (neutral) region gets increased.
2) When a photon of energy is greater than or equal to the band gap energy of the photo diode incidents on the depletion region, the electron hole pairscreated due to the absorption of photon.
3) The mobile charges are accelerated by the applied voltage which gives photo current in external circuit.
4) It acts as a linear device because the photo current is directly proportional to the optical power incident.

3.3 LEDs:

Light Emitting Diode (LED) works only in forward bias condition. When Light Emitting Diode (LED) is forward biased, the free electrons from n-side and the holes from p-side are pushed towards the junction.

When free electrons reach the junction or depletion region, some of the free electrons recombine with the holes in the positive ions. We know that positive ions have less electrons than protons. Therefore, they are ready to accept electrons. Thus, free electrons recombine with holes in the depletion region. In a similar way, holes from the p-side recombine with electrons in the depletion region.

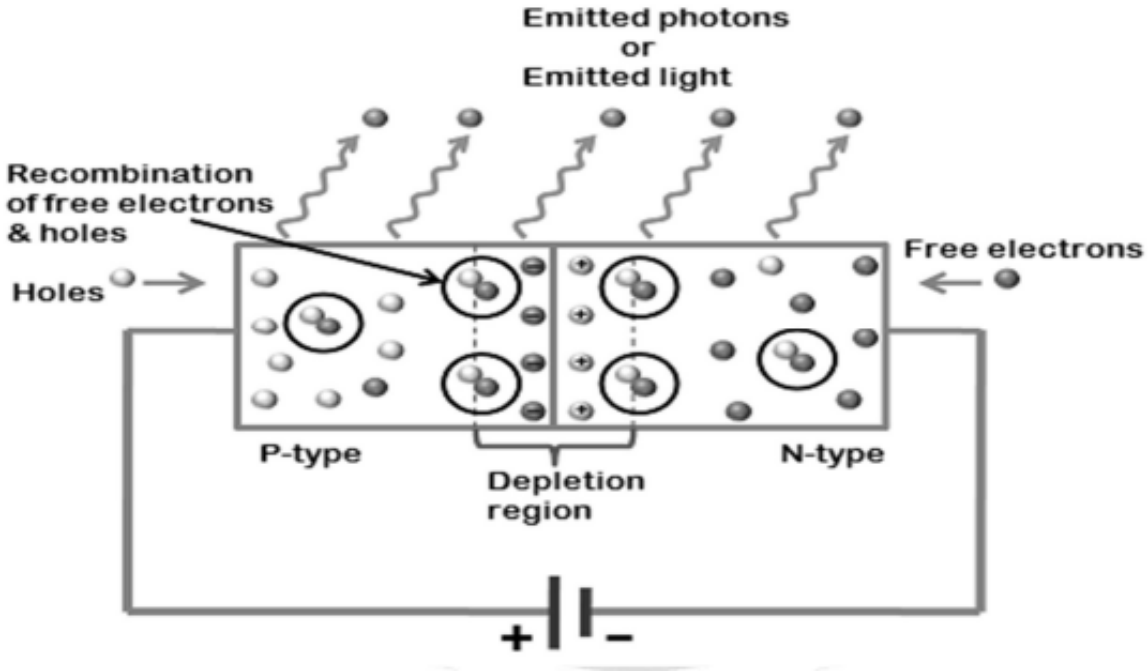


Figure 3.3: Working of Light Emitting Diode

Because of the recombination of free electrons and holes in the depletion region, the width of depletion region decreases. As a result, more charge carriers will cross the p-n junction.

Some of the charge carriers from p-side and n-side will cross the p-n junction before they recombine in the depletion region. For example, some free electrons from n-type semiconductors cross the p-n junction and recombine with holes in p-type semiconductor. In a similar way, holes from p-type semiconductors cross the p-n junction and combine with free electrons in the n-type semiconductor.

Thus, recombination takes place in depletion region as well as in p-type and n-type semiconductor. The free electrons in the conduction band releases energy in the form of light before they recombine with holes in the valence band.

In silicon and germanium diodes, most of the energy is released in the form of heat and emitted light is too small. However, in materials like gallium arsenide and gallium phosphide the emitted photons have sufficient energy to produce intense visible light. When external voltage is applied to the valence electrons, they gain sufficient energy and break the bonding with the parent atom. The valence electrons which break bonding with the parent atom are called free electrons. When the valence electron left the parent atom, they leave an

empty space in the valence shell at which valence electron left. This empty space in the valence shell is called a hole. The energy level of all the valence electrons is almost the same.

Grouping the range of energy levels of all the valence electrons is called valence band. In a similar way, energy level of all the free electrons is almost same. Grouping the range of energy levels of all the free electrons is called conduction band. The energy level of free electrons in the conduction band is high compared to the energy level of valence electrons or holes in the valence band. Therefore, free electrons in the conduction band need to lose energy to recombine with the holes in the valence band. The free electrons in the conduction band do not stay for long periods. After a short period, the free electrons lose energy in the form of light and recombine with the holes in the valence band. Each recombination of charge carrier will emit some light energy.

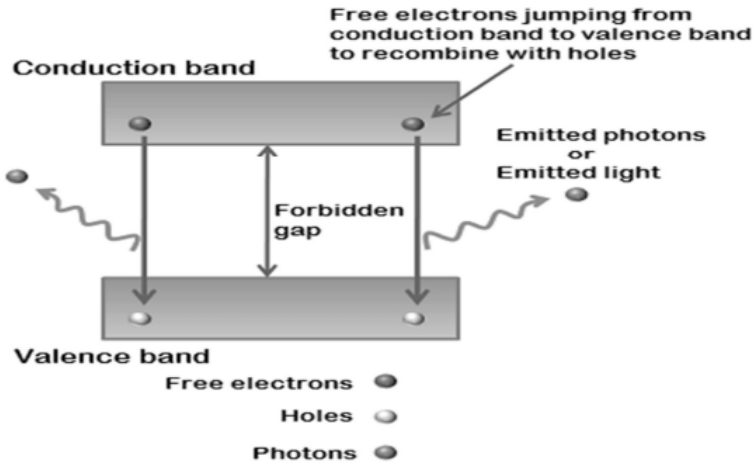


Figure 3.4 Process of Light Emission in LED

The energy loss of free electrons or the intensity of emitted light depends on the forbidden gap or energy gap between conduction band and valence band. The semiconductor device with large forbidden gap emits high intensity light whereas the semiconductor device with small forbidden gap emits low intensity light. In other words, the brightness of the emitted light depends on the material used for constructing LED and forward current flow through the LED. In normal silicon diodes, the energy gap between conduction band and valence band is less.

Hence, the electrons fall only a short distance. As a result, low energy photons are released. These low energy photons have low frequency which is invisible to the human eye. In LEDs, the energy gap between conduction band and valence band is very large so the free electrons in LEDs have greater energy than the free electrons in silicon diodes. Hence, the free electrons fall to a large distance. As a result, high energy photons are released. These high energy photons have high frequency which is visible to the human eye. The efficiency of the generation of light in LED increases with increase in injected current and with a decrease in temperature. In light emitting diodes, light is produced due to the recombination process.

Recombination of charge carriers takes place only under forward bias conditions. Hence, LEDs operate only in forward bias condition.

When light emitting diode is reverse biased, the free electrons (majority carriers) from nside and holes (majority carriers) from p-side moves away from the junction. As a result, the width of depletion region increases, and no recombination of charge carriers occurs. Thus, no light is produced. If the reverse bias voltage applied to the LED is highly increased, the device may also be damaged. All diodes emit photons or light, but not all diodes emit visible light. The material in an LED is selected in such a way that the wavelength of the released photons falls within the visible portion of the light spectrum. Light emitting diodes can be switched ON and OFF at a very fast speed of 1 ns.

3.3.1 Light Emitting Diode (LED) Symbol:

The symbol of LED is like the normal p-n junction diode except that it contains arrows pointing away from the diode indicating that light is being emitted by the diode.

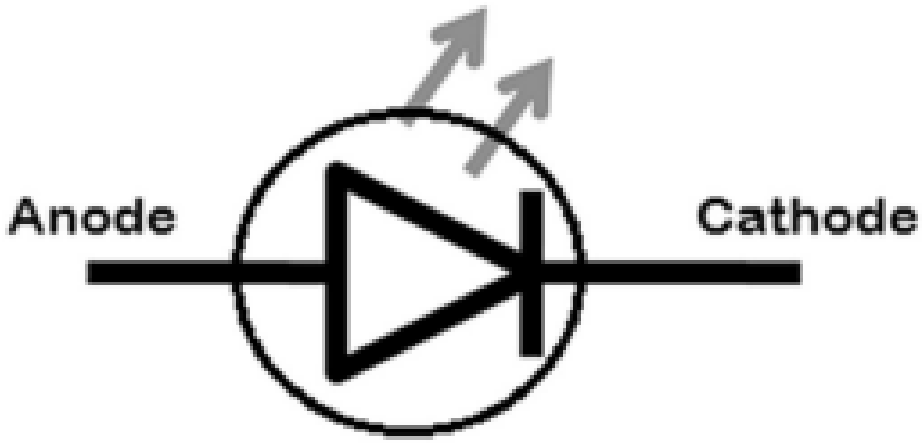


Figure 3.5 Symbol of Light Emitting Diode

LEDs are available in different colors. The most common colors of LEDs are orange, yellow, green and red. The schematic symbol of LED does not represent the color of light. The schematic symbol is the same for all colors of LEDs. Hence, it is not possible to identify the color of LED by seeing its symbol.

3.3.2 LED Construction:

One of the methods used to construct LED is to deposit three semiconductor layers on the substrate. The three semiconductor layers deposited on the substrate are n-type semiconductor, p-type semiconductor and active region. Active region is present in between the n-type and p-type semiconductor layers.

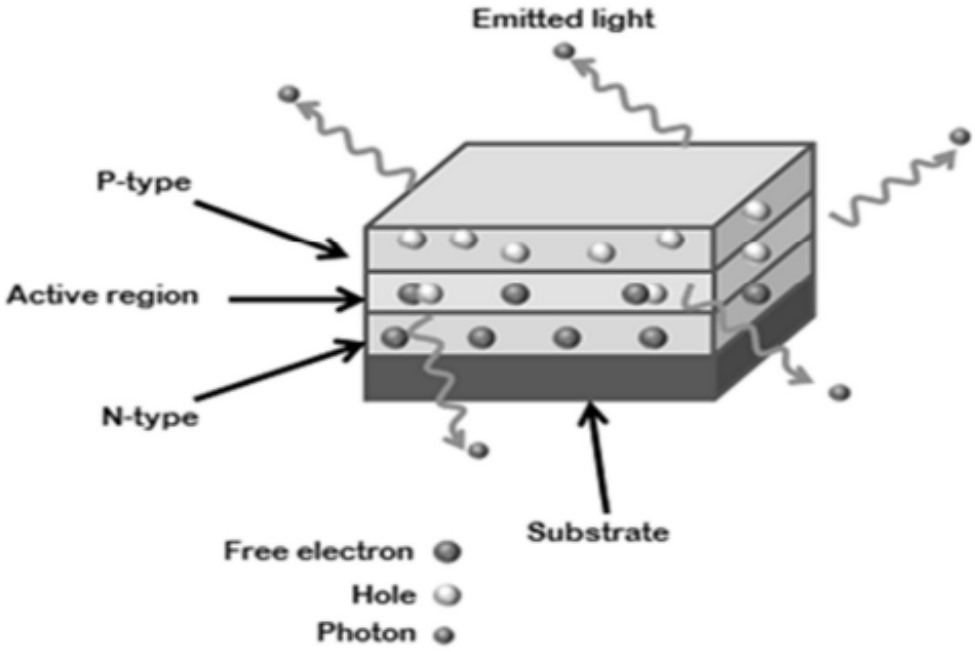


Figure 3.6 Construction of Light Emitting Diode

When LED is forward biased, free electrons from n-type semiconductor and holes from p-type semiconductor are pushed towards the active region. When free electrons from n-side and holes from p-side recombine with the opposite charge carriers (free electrons with holes or holes with free electrons) in active region, an invisible or visible light is emitted. In LED, most of the charge carriers recombine at active region. Therefore, most of the light is emitted by the active region. The active region is also called depletion region.

3.3.3 Output Characteristics of LED:

The amount of output light emitted by the LED is directly proportional to the amount of forward current flowing through the LED. The more forward the current, the greater the output light is emitted. The graph of forward current vs output light is shown in figure 3.7.

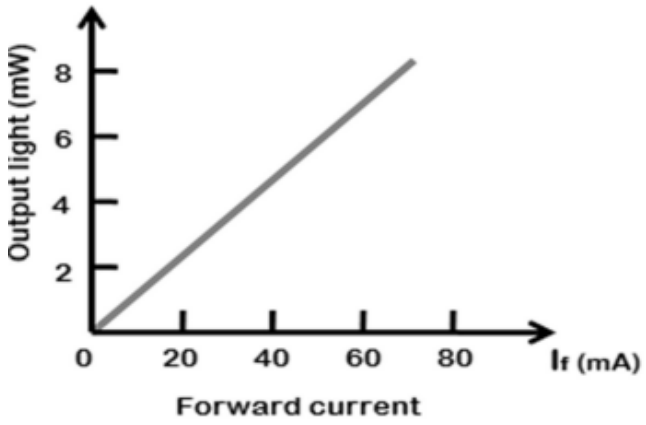


Figure 3.7 Characteristics of Light Emitting Diode

3.4 VISIBLE LED'S AND INVISIBLE LED'S:

LEDs are mainly classified into two types: visible LEDs and invisible LEDs. Visible LED is a type of LED that emits visible light. These LEDs are mainly used for display or illumination where LEDs are used individually without photosensors. Invisible LED is a type of LED that emits invisible light (infrared light). These LEDs are mainly used with photosensors such as photodiodes.

3.4.1 Determines the Color of an LED:

The material used for constructing LED determines its color. In other words, the wavelength or color of the emitted light depends on the forbidden gap or energy gap of the material. Different materials emit different colors of light. Gallium arsenide LEDs emit red and infrared light. Gallium nitride LEDs emit bright blue light. Yttrium aluminium garnet LEDs emit white light. Gallium phosphide LEDs emit red, yellow and green light. Aluminium gallium nitride LEDs emit ultraviolet light. Aluminum gallium phosphide LEDs emit green light.

3.5 SUMMARY:

This study explores the principles, construction, and materials used in solar cells, photo detectors, and LEDs, highlighting their importance in energy conversion and optical communication. Solar cells, based on the photovoltaic effect, generate electricity from light using materials like silicon, GaAs, CdTe, and $CuInSe_2$. Photodetectors, including PIN and avalanche photodiodes, convert light signals into electrical signals, playing a vital role in fiber-optic communication. LEDs function through electroluminescence, emitting visible or infrared light depending on the semiconductor material used, with applications in displays, lighting, and sensors. The efficiency of these devices is influenced by material properties such as band gaps and doping. Understanding their working mechanisms and material selection helps enhance performance in renewable energy, high-speed communication, and solid-state lighting. This study provides a detailed insight into these semiconductor devices, emphasizing their role in advancing modern technology and promoting energy-efficient solutions.

3.6 TECHNICAL TERMS:

Photovoltaic Effect, Band Gap, Electroluminescence, Depletion Region, Quantum Efficiency

3.7 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What is the photovoltaic effect, and how does it enable solar cells to generate electricity?

2) How does the band gap of a semiconductor material influence the efficiency of solar cells, photodetectors, and LEDs?

3) What is electroluminescence, and why is it essential for the functioning of LEDs?

Short Answer Questions:

1) How does the depletion region in a p-n junction semiconductor affect the performance of solar cells and photodetectors?

2) What are the key differences between PIN photodiodes and avalanche photodiodes in optical communication systems?

3) How does quantum efficiency impact the performance of solar cells and photodetectors?

3.8 SUGGESTED READINGS:

1) S.M. Sze and K.K. Ng, Physics of Semiconductor Devices, 3rd Ed. Hoboken, NJ, USA: Wiley, 2007.

2) J. Nelson, The Physics of Solar Cells. London, UK: Imperial College Press, 2003.

3) Rogalski, Semiconductor Optoelectronic Devices: Introduction to Physics and Applications, 2nd ed. Boca Raton, FL, USA: CRC Press, 2020.

4) M.A. Green, Solar Cells: Operating Principles, Technology and System Applications. Englewood Cliffs, NJ, USA: Prentice Hall, 1982.

5) G. Streetman and S. Banerjee, Solid State Electronic Devices, 7th Ed. Upper Saddle River, NJ, USA: Pearson, 2014.

Prof. Sandhya Cole

LESSON-4

OPERATIONAL AMPLIFIERS

4.0 AIM AND OBJECTIVES:

The aim of this lesson is to understand the fundamental concepts, configurations, and applications of operational amplifiers (Op-Amps), focusing on differential amplifiers as their core building blocks. The lesson explores various circuit configurations, including dual-input balanced output differential amplifiers, and provides both DC and AC analysis to understand their operational behavior. It also covers critical parameters such as common-mode rejection ratio (CMRR), inverting and non-inverting inputs, and voltage gain calculations. By the end of this lesson, learners will be able to analyze and design differential amplifier circuits, comprehend the working principles of Op-Amps, and apply this knowledge to real-world applications. Additionally, students will gain insights into the block diagram representation of a typical Op-Amp and its practical significance in analog electronics, including amplification, filtering, and signal processing.

STRUCTURE:

4.1 DIFFERENTIAL AMPLIFIERS:

The differential amplifier, also called a difference amplifier, as the name implies, amplifies the difference between two signals. Because of its balanced nature and symmetry, it can amplify very small signals. It usually requires a minimum number of capacitors and can operate without bypassing and coupling capacitors. It is the basic building block of operational amplifiers, which are most widely used in integrated circuits.

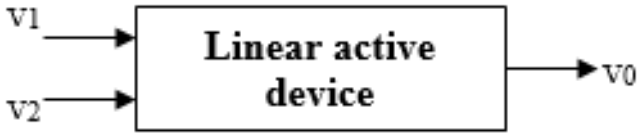


Figure 4.1 Schematic Diagram of a Differential Amplifier

For a linear active device with two input signals v1 and v2 and the output signals v0 each measured with respect to ground, we have

$$v_0 = A\,(v_1 - v_2) \qquad \text{------ (4.1)}$$

where A is the voltage gain of the differential amplifier. In actual practice, the output depends not only upon the difference of the two input signals but also upon the average level. In symmetrical circuits we talk about the in-phase signals (called common mode (CM) Signals v_c) and the difference or anti-phase signals (called differential mode (DM) signals v_d). They are defined as

$$v_c = \frac{1}{2}(v_1+v_2), \text{ and } vd = (v_1 - v_2) \qquad \text{------ (4.2)}$$

The output v_0 can be expressed as linear combination of the two input voltages, as

$$v_0 = A_1 v_1 + A_2 v_2 \qquad \text{------ (4.3)}$$

Where A_1 and A_2 are the voltage amplifications from input 1and 2 respectively. From equations (4.2)

and (4.3) we get

$$v_o = A_1(v_c + \frac{1}{2}v_d) + A_2(v_c - \frac{1}{2}v_d) = (A_1 + A_2)v_c + \frac{1}{2}(A_1 - A_2)\,v_d = A_c v_c + A_d v_d$$

----------(4.4)

where $A_c = A_1 + A_2$ and $A_d = \frac{1}{2}(A_1 - A_2)$, are voltage gains for the signals in common mode and differential modes respectively. They may be defined as

$$A_d = \left(\frac{v_0}{v_d}\right)_{V_C = 0} \quad \text{and} \quad A_c = \left(\frac{v_0}{v_c}\right)_{V_d = 0} \qquad \text{----------(4.5)}$$

Thus, A_d can be measured directly by setting $v_c = 0$, or $v_2 = - v_1$.

The A_c can be measured by setting $v_d = 0$, or $v_2 = v_1$, generally the desired signals in differential amplifier are DM and undesired signals are CM. The figure of merit for differential amplifiersin defined as

$$\rho = \left[\frac{A_d}{A_c}\right] \qquad \text{---------- (4.6)}$$

This is called the common-mode rejection ratio (CMRR) and is also sometimes referred to as the discrimination factor of a differential amplifier. Ideally $A_c=0$, and CMRR= $\propto$. In practice, A/c is non-zero but very small, whereas A_d is very large. The combination of equations (4.4) and (4.6) gives

$$Vo = A_d v_d\left[1+\frac{A_c\, v_c}{A_d\, v_d}\right] = A_d v_d\left[1+\frac{v_c}{v_d}.\frac{1}{CMRR}\right]$$

$$V_0 = A_d v_d \text{ (since CMRR = } \propto) \qquad \text{---------- (4.7)}$$

A basic differential amplifier circuit (ckt), consisting of two interlocked common emitter amplifier stages is shown in figure 4.2. The two stages are linked by having both emitters connected to a constant current generator. As current through one emitter increases, current through the other decreases.

The circuit is symmetrical about the vertical dashed line and the two transistors and resistors R_c form a bridge circuit that is balanced under zero input signal. The resistors and the transistors are simultaneously fabricated in adjacent areas on a small chip. They will be at the same temperature. A simultaneous change in h_{FE} or v_{BE} will produce equal changes in the voltages at a and b and v_0 will not be affected.

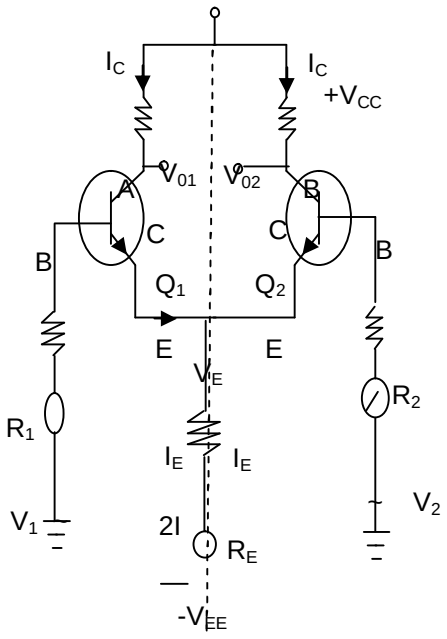


Figure 4.2 Basic Differential Amplifier Circuit Diagram

Consider the circuit operation with no input signals. For $v_1=v_2=0$, an emitter current I_E flows in each BJT. Therefore $I_C = I_E$ and

$$V_{01} = V_{02} = V_{CC} - I_C R_C \qquad \text{----------(4.8)}$$

$$\text{thus the base current } I_B = \frac{I_E}{h_{FE}} \qquad \text{----------(4.9)}$$

$$\text{and } V_E = -I_B R_1 - V_{BE} \qquad \text{----------(4.10)}$$

If v_{CE} is chosen large enough to bias each BJT in the center or the linear operating region, then

$$V_{CC} = V_{EE} + 2 I_E R_E + V_{CE} + I_C R_C \qquad \text{----------(4.11)}$$

The ac equivalent circuit for use differential amplifier is shown in fig (4.3). Here it is assumed

$h_{oe} R_C << 1$ or $h_{oe} << 1/R_C$ and thus h_{oe} is omitted in this figure. The collector current $I_c = h_{fe} I_b$. The voltage $h_{re} v_c$ is neglected in comparison with the $h_{ib} I_b$ drop across h_{ie}

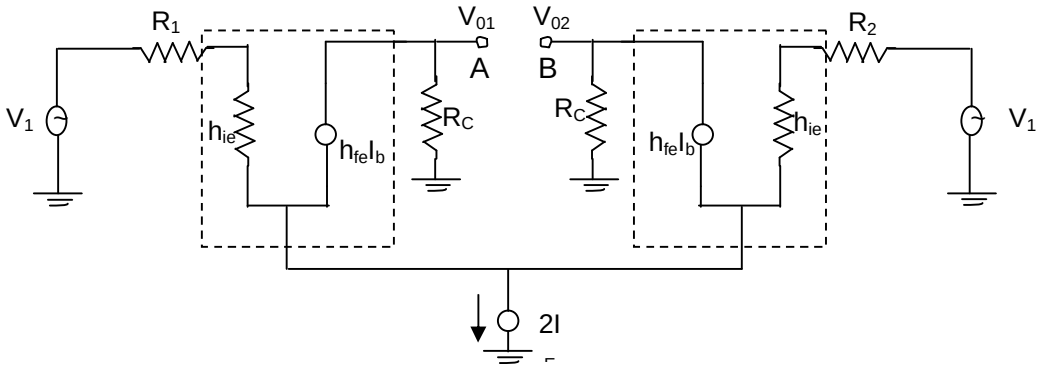


Figure4.3 AC Equivalent Circuit of Basic Differential Amplifier

4.1.1 Common Mode Voltage Gain:

$$\text{Let } v_1 = v_2 = v_s \text{ , user } A_c = \frac{v_o}{v_c} = \frac{v_o}{v_s} .$$

Due to symmetry, each input at the base, sees a common emitter circuit, with an un bypassed emitter resister of $2R_E$ (The emitter resistor is effectively doubled, as it carries the emitter current for both transistors). Thus we have

$$Z_i = R_1 + h_{ie} + 2R_E (1 + h_{fe}) \text{ and } Z_0 = R_c \qquad \text{---------- (4.12)}$$

$$\text{Current gain} = A_i = \frac{-I_c}{I_b} = \frac{-h_{fe} I_b}{I_b} = -h_{fe} \qquad \text{--------- (4.13)}$$

$$\text{Voltage gain } A_v = \frac{A_i Z_0}{Z_i} = \frac{-h_{fe} R_c}{\left(R_1 + h_{ie} + 2R_E\left(1 + h_{fe}\right)\right)} \qquad \text{---------- (4.14)}$$

Since usually $(1+h_{fe})\,2R_E >> h_{ie}$ and $1+h_{fe} = h_{fe}$. The source resistance $R_i << h_{ie}$. Therefore

$$A_v = \frac{-R_C}{2R_E} = \text{common mode voltage gain } A_c.$$

4.1.2 Differential Mode Voltage Gain:

$$\text{Let} -v_2 = v_1 = \frac{v_s}{2};$$

Therefore vd = v_1, $-v_2 = v_s$ and $A_d = \frac{v_0}{v_d} = \frac{v_0}{v_s}$. From the symmetry of fig (4.2) for $v_1 = -v_2$,

the emitter of each transistor is grounded for small signal operation i.e $R_E=0$ and

$$Z_i = 2(R_i + h_{ie}),\ Z_o \cong R_c,\ A_i \cong -h_{fe}. \qquad \text{---------- (4.15)}$$

$$A_v = \frac{A_i Z_0}{Z_i} = \frac{-h_{fe} R_c}{2(R_i + h_{ie})} \qquad \text{---------- (4.16)}$$

= differential mode voltage gains A_d.

Common mode rejection ratio:

$$\text{CMRR} = \frac{A_d}{A_c} = \frac{-h_{fe} R_c}{2(R_i + h_{ie})} \div \left(-\frac{R_c}{2R_E}\right) = \frac{-h_{fe} R_e}{(R_1 + h_{ie})} \qquad \text{---------(4.17)}$$

Thus, we see that CMRR increases with R_E as desirable.

4.1.3 Constant Current Generators:

Instead of resistor R_E, a constant current generator is used. It may be a JFET with its gate tied to its source. A BJT can be used in a similar way with voltage divider bias. In both cases the output current is approximately constant if the voltage across the device is sufficient.

4.1.4 Constant Current Bias:

In the differential amplifier discussed so far, the combination of R_E & V_{EE} is used to set up the dc emitter current. We can also use constant bias to set up the dc emitter current if desired. In fact, the constant current bias is better because it provides current stabilization and in turn, assures a stable operating point for the differential amplifier. Fig (4.4). shows the dual input balanced output differential amplifier using a resistive constant current bias. Notice that the resistor R_E is replaced by a constant current source transistor (Q3) circuit. The dc collector current in the transistor Q_3 is established by resisters R_1, R_2 and R_E and can

be determined as follows. Applying the voltage divider rule, the voltage at the base of transistor Q_3 (neglecting base loading effect) is

$$V_{B3} = \frac{-R_2 V_{EE}}{(R_1 + R2)} \quad \text{----------(4.18)}$$

$$V_{E3} = V_{B3} - V_{BE3} = \left(\frac{-R_2 V_{EE}}{(R_1 + R2)}\right) - V_{BE3} \quad \text{----------(4.19)}$$

$$I_{E3} = I_{C3} = \frac{(V_{E3} - (-V_{EE}))}{R_E} \quad \text{----------(4.20)}$$

$$I_{C3} = v_{EE} - \left(\frac{-R_2 V_{EE}}{(R_1 + R2)}\right) . \frac{V_{BE3}}{R_E} \quad \text{----------(4.21)}$$

Because two halves of the differential amplifiers are symmetrical, each has half of current I_{ce3}.

That is

$$I_{E1} = I_{E2} = \frac{I_{C3}}{2} = V_{EE} - \left(\frac{R_2 V_{EE}}{(R_1 + R2)}\right) . \frac{V_{BE3}}{2R_E} \quad \text{----------(4.22)}$$

The collector current I_{c3} in transistor Q_3 is fixed and must be invariant because no signal is injected into either the emitter or the base of Q_3. Thus, the transistor Q_3 is a source of constant emitter current for transistors Q_1 and Q_2 of the differential amplifiers. Besides supplying constant emitter current, the constant current bias also provides a very high source resistance since the ac equivalent of the dc current source is ideally an open circuit.

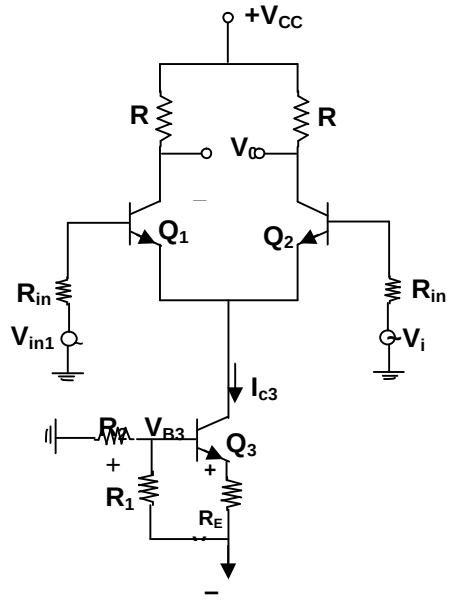


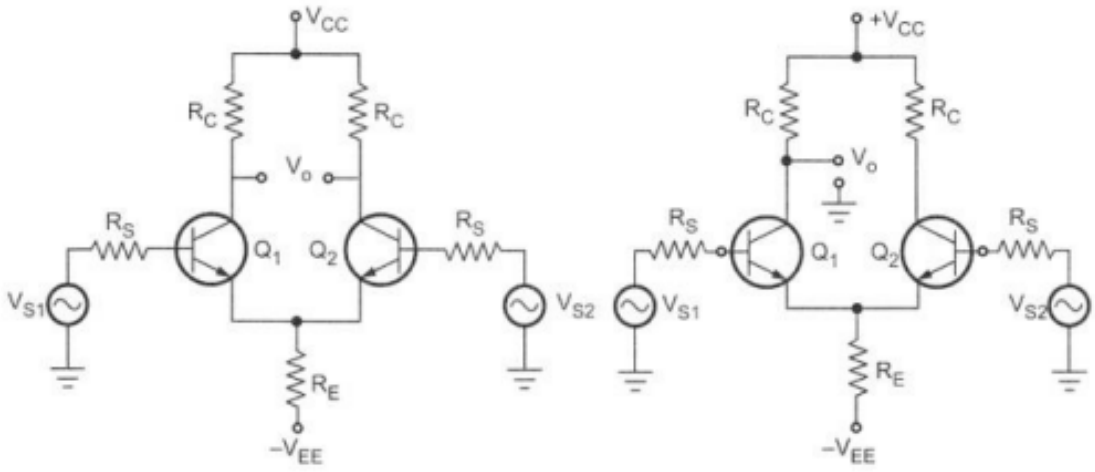
Figure 4.4 Differential Amplifier Using Constant Current Bias

4.2 CIRCUIT CONFIGURATIONS:

The differential amplifier, in the different amplifier stage in the op-amp, can be used in four configurations:

- Dual input balanced output differential amplifier.
- Dual input, unbalanced output differential amplifier.
- Single input, balanced output differential amplifier.
- Single input, unbalanced output differential amplifier

The differential amplifier uses two transistors in common emitter configuration. If output is taken between the two collectors, it is called balanced output or double ended output. While if the output is taken between one collector with respect to ground it is called unbalanced output or single ended output. If the signal is given to both the input terminals it is called dual input, while if the signal is given to only one input terminal and other terminal is grounded it is called single input or single ended input. Out of these four configurations the dual input, balanced output is the basic differential amplifier configuration. This is shown in Figure 4.1. (a). The dual input, unbalanced output differential amplifier is shown in Figure 4.1.(b). The single input, balanced output differential amplifier is shown in the Figure 4.5 (c) and the single input, unbalanced output differential amplifier is shown in the Figure 4.5. (d).



(a) Dual input balanced output

(b) Dual input unbalanced output

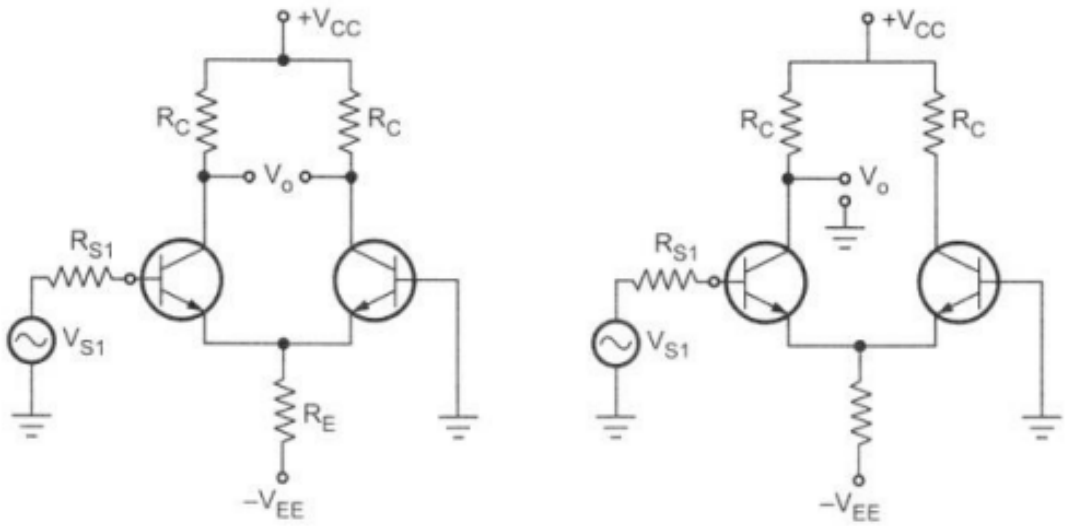


Figure 4.5 The dual input, unbalanced output differential amplifier is shown in the Fig (b). The single input, balanced output differential amplifier is shown in the Fig (c) and the single input, unbalanced output differential amplifier is shown in the Fig. (d).

4.3 DUAL INPUT- BALANCED OUTPUT DIFFERENTIAL AMPLIFIER:

The circuit shown below is a dual-input balanced-output differential amplifier. Here in this circuit, the two input signals (dual input), vin1 and vin2, are applied to the bases B1 and B2 of transistors Q1 and Q2.The output vo is measured between the two collectors C1 and C2 which are at the same dc potential.Because of the equal dc potential at the two collectors with respect to ground, the output is referred as a balanced output.

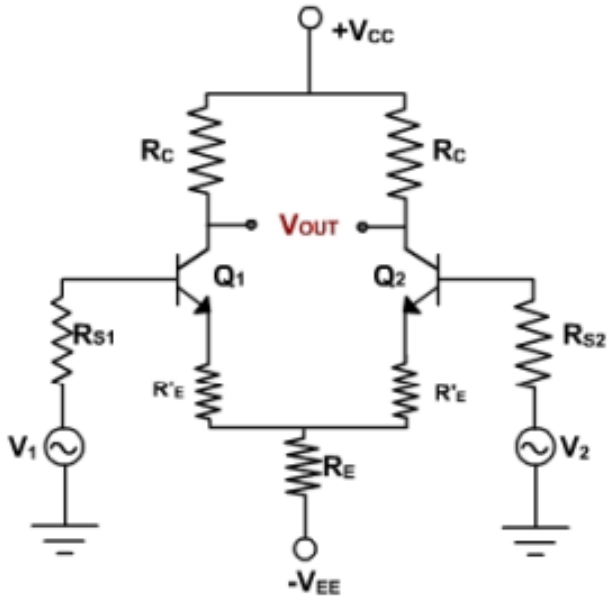


Figure 4.6: Circuit Diagram

4.4 DC ANALYSIS:

To determine the operating point values (ICQ and VCEQ) for the differential amplifier, we need to obtain a dc equivalent circuit. The dc equivalent circuit can be obtained simply by reducing the input signals vin1 and vin2 to zero. The dc equivalent circuit thus obtained is shown in fig below. Note that the internal resistances of the input signals are denoted by Rinbecause Rin1 = Rin2.Since both emitter biased sections of the differential amplifier are symmetrical (matched in all respects), we need to determine the operating point collector current ICQ and collector to emitter voltage VCEQ for only one section.We shall determine the ICQ and VCEQ values for transistor Q1 only.These ICQ and VCEQ values can then be used for transistor Q2 also.

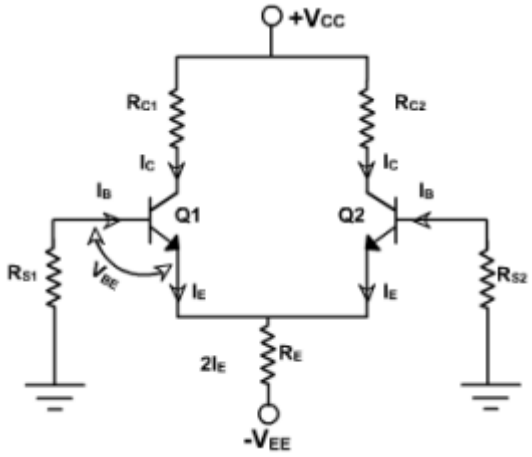


Figure 4.7 DC Equivalent Circuit for Dual-Input Balanced Output Differential Amplifier

Applying Kirchhoff's voltage law to the base-emitter loop of the transistor Q1,

RinIB- VBE - RE(2IE)+VEE = 0 (1)

But

IB = IE/Bdc since IC = IE

Thus the emitter current through Q1 is determined directly from eqn (1) as follows:

IE = (VEE - VBE)/(2RE + Rin/Bdc) (2)

Where VBE = 0.6V for silicon transistors

VBE =0.2V for germanium transistors

Generally, Rin/Bdc<< 2RE.Therefore, eqn (2) can be rewritten as

ICQ=IE = (VEE - VBE)/2RE (3)

From eqn (3) we see that the value of RE sets up the emitter current in transistors Q1 and Q2 for a given value of VEE. In other words, by selecting a proper value of RE, we can obtain a desired value of emitter current for a known value of –VEE. Notice that the emitter current in transistors Q1 and Q2 is independent of collector resistance RC.

Next we shall determine the collector to emitter voltage VCE. The voltage at the emitter of transistor Q1 is approximately equal to VBE if we assume the voltage drop across Rin to be negligible. Knowing the value of emitter current IE(=IC),we can obtain the voltage at the collector VCC as follows:

$$\mathbf{VC = VCC - RCIC}$$

Thus the collector to emitter voltage VCE is

VCE = VC - VE = (VCC – RCIC) – (-VEE)

VCEQ=VCE = VCC+VBE - RCIC (4)

Thus for both transistors we can determine the operating point values by using the eqns (2) and (4), respectively, because at the operating point IE=ICQ and VCEQ=VCE

Remember that the dc analysis eqns (2) and (4) are applicable for all 4 differential amplifier configurations as long as we use the same biasing arrangement for each of them.

4.5 AC ANALYSIS:

To perform ac analysis to derive the expression for the voltage gain Ad and input resistance Ri of adifferential amplifier:

1) Set the dc voltages +VCC and –VEE at 0

2) Substitute the small signal T equivalent models for the transistors

Figure below shows resulting ac equivalent circuit of the dual input balanced output differential Amplifier.

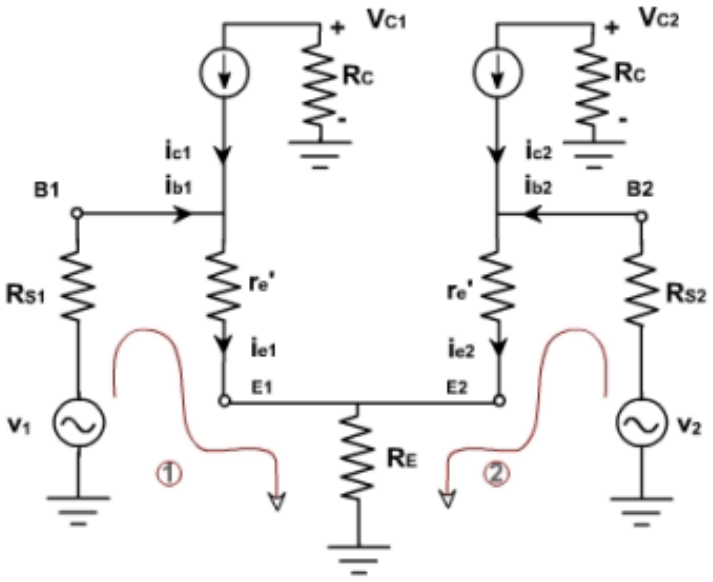


Figure 4.8AC Equivalent Circuit for Dual-Input Balanced Output Differential Amplifier

4.6 INVERTING AND NON-INVERTING INPUTS CMRR:

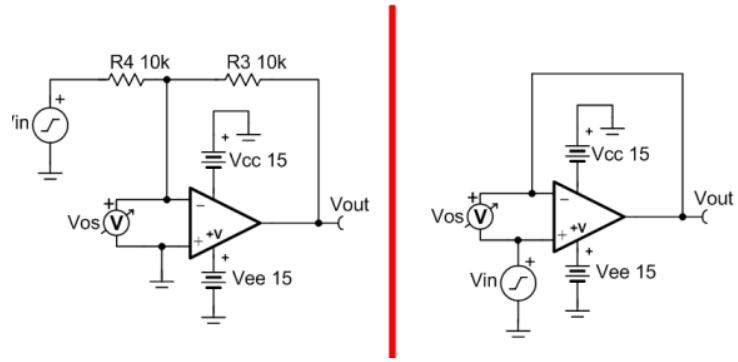


Figure 4.9: Inverting and Non-Inverting Inputs of CMRR

4.6.1 Inverting Configuration:

In this configuration input is supplied on the inverting terminal of the op-amp, so called inverting configuration. R2 closes loop around the op-amp, so acts as negative feedback.

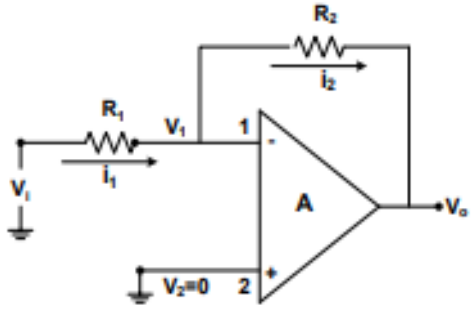


Figure 4.9: Op-Amp Inverting Configuration

4.6.2 Non-Inverting Configuration:

Here, input is fed into the non-inverting terminal-2 of an op-amp, so called non-inverting configuration.

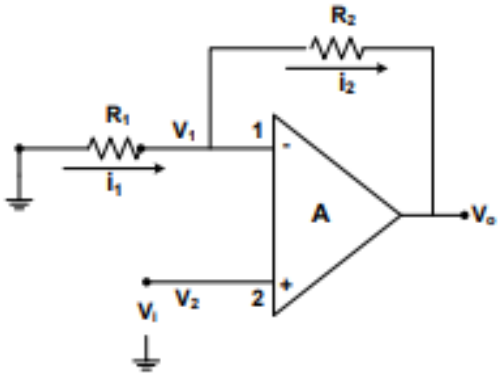


Figure 4.10 Op-Amp Non-Inverting Configuration

4.7 BLOCK DIAGRAM OF A TYPICAL OP-AMP-ANALYSIS:

An operational amplifier is a direct-coupled high gain amplifier usually consists of one or more differential amplifiers and usually followed by a level translator and an output stage. The output stage is generally a push pull or push pull complementary symmetry pair. An operational amplifier is available as a single integrated circuit package. The operational amplifier is a versatile device that can be used to amplify dc as well as ac input signals and was originally designed for computing such mathematical functions as addition, subtraction, multiplication and integration. Thus, the name operational amplifier stems from its original use for doing these mathematical operations and so is abbreviated to op-amp. With the addition of suitable external feedback component, the modern-day operational amplifier can be used for a variety of applications. Such as ac and dc signal amplification, active filters, oscillators, comparators, regulators and others.

4.7.1 Block Diagram Representation of a Typical Op-Amp:

Since an op-amp is a multistage amplifier. It can be represented by a block diagram shown in figure (4.11). The stage is the dual input balanced output differential amplifier. This stage generally provides most of the voltage gain of the amplifier and also establishes the input resistance of the op-amp. The intermediate stage is usually another differential amplifier, which is driven by the output of the first stage. In most amplifiers, the intermediate stage is dual input unbalanced (single ended) output. Because direct coupling is used, the dc voltage at the output of the intermediate stage is well above ground potential. Therefore, the level translator circuit is used after the intermediate stage to shift the dc level at the output of the intermediate stage downward to zero voltage with respect to ground. The final stage is usually a push pull complementary amplifier output stage. The output stage increases the output voltage swing and raises the current supplying capabilities of the op-amp. The well-designed output stage also provides low output resistance.

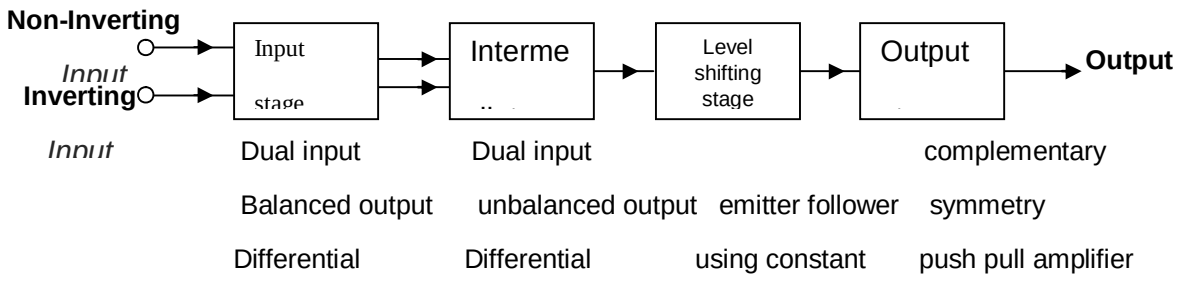


Fig 4.11 Block Diagram of a Typical Op-Amp

4.7.2 Schematic Symbol:

The most widely used symbol for a circuit with two inputs and one output is shown in figure 4.12.

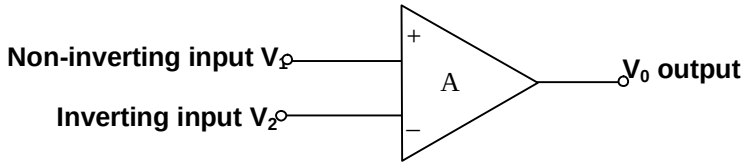


Figure 4.12 Schematic Symbol of Op-Amp

In fig (4.12)

v_1 = voltage at the non-inverting input (volts)

v_2 = voltage at the inverting input (volts)

v_o = output voltage (voltage)

All these are measured w.r.t. ground

A = large signal voltage gain that is specified on the data sheets for an op-amp

For amplifiers, power supply and other pin connections are omitted. Since the input differential amplifier stage of the op-amp is designed to be operated in differential mode, the differential inputs are designated by the (+) and (-) notations, the (+) input is used for non-inverting input. An ac signal (or dc voltage) applied to this input produces an in-phase (or same polarity) signal at the output. On the other hand, the (-) input is the inverting input because an ac signal (or dc voltage) applied to this input produces an 180 out of phase (or opposite polarity) signal at the output.

4.7.3 Ideal Op-Amp:

An ideal op-amp exhibits the following electrical characteristics:

(1) Infinite voltage gain A_V.

(2) Infinite input resistance R_i, so that, almost any signal source can drive it and there is no loading of the preceding stage.

(3) Zero output resistance R_o, so that, output can drive an infinite number of other devices

(4) Zero output voltage when the input voltage is zero.

(5) Infinite bandwidth, so that any frequency signal from 0 to $\propto$ Hz can be amplified without attenuation.

(6) Infinite common mode rejection ratio so that output common mode noise voltage is zero.

(7) Infinite slew rate so that voltage changes occur simultaneously with input voltage changes.

There are practical op-amps that can be made to achieve some of these characteristics using a negative feedback arrangement. In particular, the input resistance, output resistance, and bandwidth can be brought close to ideal values by this method.

4.7.4 Equivalent Circuit of An Op-Amp: Figure 4.13 shows an equivalent circuit of an op-amp. The circuit includes important values from the data sheets: A, R_i and R_o.

Note that A v_{id} is an equivalent Thevenin voltage source, and R_o is the Thevenin equivalent resistance looking back into the output terminal of an op-amp.

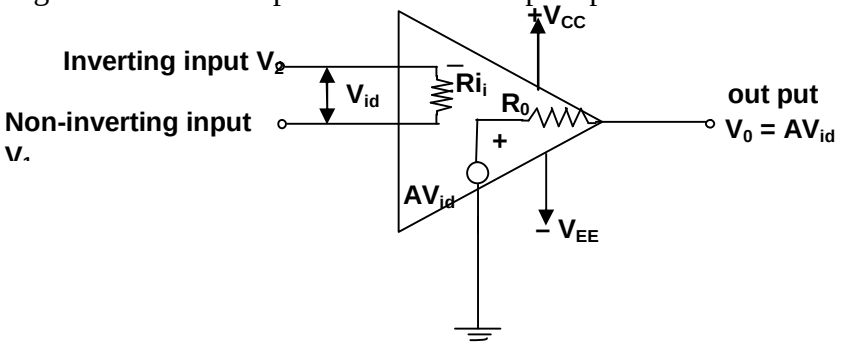


Figure 4.13 Equivalent Circuit of Op-Amp

The equivalent circuit is useful in analyzing the basic operating principles of op-amps and in observing the effectiveness of feedback arrangement. For the circuit shown in fig. (1.7), the output voltage is

$$v_0 = Av_{id} = A\,(v_1 - v_2) \qquad \text{----------(4.23)}$$

Where A = large- signal voltage gain

v_{id} = difference input voltage

v_1 = voltage at the non-inverting input terminal w.r.t. ground.

v_2 = voltage at the inverting input terminal w.r.t. ground.

Eqn. 4.23 indicates that the output voltage v_0 is directly proportional to the algebraic difference between the two input voltages. In other words, the op-amp amplifies the difference between the two input voltages; it does not amplify the input signal voltages themselves. For this reason, the polarity of the output voltage depends upon the polarity of the different voltages.

4.8 SUMMARY:

This lesson provides a comprehensive understanding of operational amplifiers (Op-Amps) with a focus on differential amplifiers, their configurations, and applications. It begins with an introduction to Op-Amps, explaining their importance in analog electronics and their role in signal processing. The lesson explores different types of differential amplifier circuits, including dual-input balanced output configurations, with detailed DC and AC analysis. Key parameters such as common-mode rejection ratio (CMRR), input offset voltage, and gain calculations are examined to understand their impact on circuit performance. Additionally, the lesson covers the block diagram representation of an Op-Amp, highlighting its essential components and functions. By the end, students will have gained the knowledge required to analyze, design, and apply differential amplifier circuits in various applications such as amplification, filtering, and instrumentation. This understanding forms a strong foundation for further studies in electronic circuit design and applications of Op-Amps in modern technology.

4.9 TECHNICAL TERMS

Common-Mode Rejection Ratio (CMRR), Input Offset Voltage, Gain Bandwidth Product (GBP), Slew Rate, Differential Mode Gain (Ad)

4.10 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are the key characteristics of an operational amplifier (Op-Amp)?
2) How does the Common-Mode Rejection Ratio (CMRR) affect the performance of a differential amplifier?
3) What is the significance of the Gain Bandwidth Product (GBP) in an Op-Amp?

Short Answer Questions:

1) How does input offset voltage impact the accuracy of an Op-Amp circuit?
2) Why is slew rate important in high-speed applications, and how can it be improved?
3) What are the practical applications of differential mode gain in signal processing?

4.11 SUGGESTED READINGS:

1) Sedra, A.S., & Smith, K.C. (2018). Microelectronic Circuits (8th ed.). Oxford University Press.

2) Gray, P.R., Hurst, P.J., Lewis, S.H., & Meyer, R.G. (2009). Analysis and Design of Analog Integrated Circuits (5^{th} ed.). Wiley.

3) Franco, S. (2014). Design with Operational Amplifiers and Analog Integrated Circuits (4^{th} ed.). McGraw-Hill Education.

4) Razavi, B. (2016). Fundamentals of Microelectronics (2^{nd} ed.). Wiley.

5) Texas Instruments. (2021). Op Amps for Everyone Design Guide (Rev. B).

Prof. Sandhya Cole

LESSON-5

AMPLIFIERS AND FEEDBACK

5.0 AIM AND OBJECTIVES:

The aim of this lesson is to provide a comprehensive understanding of amplifiers and feedback in operational amplifier (op-amp) circuits. Amplifiers play a crucial role in signal processing, and feedback mechanisms significantly influence their performance. This lesson focuses on both open-loop and closed-loop configurations of op-amps, explaining different types of amplifiers, including differential, inverting, and non-inverting amplifiers. It also covers the concept of negative feedback, its impact on gain, input resistance, output resistance, and bandwidth, and how it improves stability while reducing distortion. Additionally, the lesson explores voltage series feedback and its role in controlling amplifier characteristics. A special focus is given to the voltage follower, its function as a unity gain amplifier, and its application as a buffer circuit. By the end of this lesson, learners will have a clear understanding of amplifier design principles, the importance of feedback in circuits, and practical applications of op-amps in signal amplification.

STRUCTURE:

5.1 OPEN LOOP CONFIGURATION INVERTING AND NON-INVERTING AMPLIFIERS:

5.1.1 Open-Loop Op-Amp Configurations:

In the case of amplifiers, the term “open-loop” indicates that no connections either direct or via another network exists between the o/p and i/p terminals. That is, the o/p signal is not fed-back in any form as part of the input signal and the “loop” that would have been formed with feedback is open.

When connected in open loop configuration the op-amp simply functions as a high gain amplifier. There are three open loop op-amp configurations:

1) The differential amplifier.
2) The inverting amplifier.
3) The non-inverting amplifier.

5.1.2 The Differential Amplifier:

Fig 5.1 shows the open loop differential amplifier in which input signals v_{is1} and v_{is2} are applied to the positive and negative input terminals. Since the op-amp amplifies the difference between the two input signals, this configuration is called the differential amplifier.

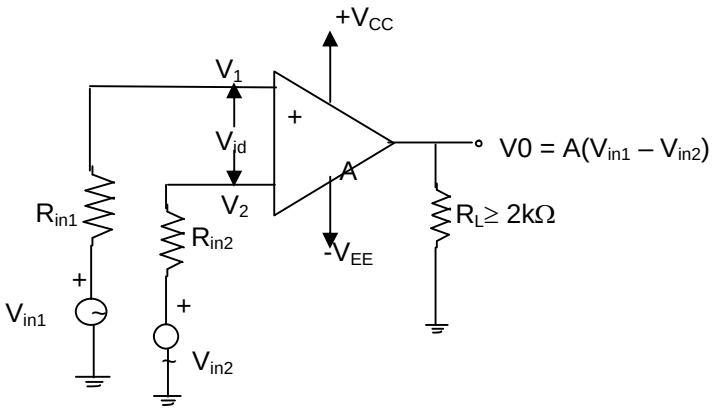


Figure 5.1: Open Loop Differential Amplifier

The op-amp is a versatile device because it amplifies both ac and dc input signals. This means that v_{is1} and v_{is2} could be either ac or dc voltages. The source resistance R_{in1} and R_{in2} are normally negligible compared to the input resistance R_i. Therefore, the voltage drops across the resistors can be assumed to be zero, which then implies that $v_1 = v_{is1}$, and $v_2 = v_{is2}$. Substituting these values of v_1 and v_2 in equation for output voltage, we get

$$\mathbf{v_0 = A(v_{in1}-v_{in2})}$$

Thus, as expected the output voltage is equal to voltage gain A times the difference between the two input voltages. Also notice that the polarity of the output voltage is dependent on the polarity of the input difference voltage (v_{in1}-v_{in2}). In open loop configurations gain A is commonly referred to as open loop gain.

5.1.3 The Inverting Amplifier:

In the inverting amplifier only one input is applied and that, to the inverting input terminal.

The non-inverting input terminal is grounded. Since $v_1 = 0$ and $v_2 = v_{in}$.

$V_0 = - AV_{in}$

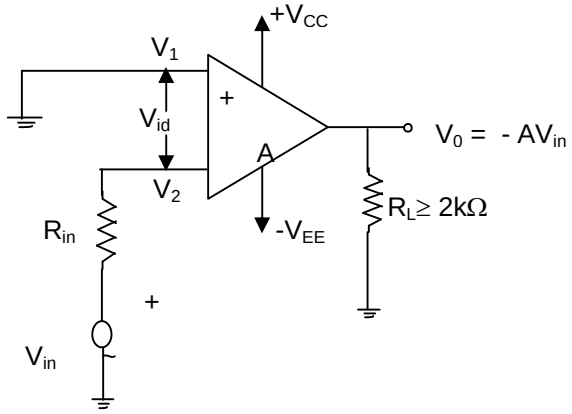


Figure 5.2 Inverting Amplifier

The negative sign indicates that the output voltage is out of phase with respect to input by 180^0or is of opposite polarity. Thus, in the inverting amplifier the input signal is amplified by gain A and is also inverted at the output.

5.1.4 The Non-Inverting Amplifier:

Fig 5.3 shows the open loop configurations of non-inverting amplifier in this configuration the input is applied to the non-inverting input terminal and the inverting input terminal is grounded.

In the circuit shown below $v_1=v_{in}$ and $v_2=0v$. Therefore, according to equation $v_0 = Av_{in}$. This means that the output voltage is larger than the input voltage. By gain A and is in phase with input signal.

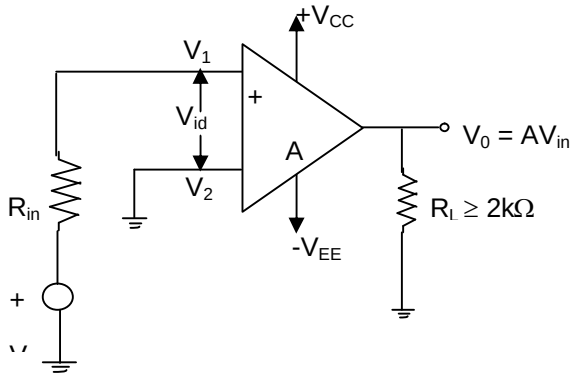


Figure 5.3 non-inverting amplifier

In all the three open loop configurations any input signal that is only slightly greater than zero drives the output to saturation level. This results from the very high gain A of the op-amp. Thus, when operated in an open loop, the output of the op-amp is either at negative

saturation or positive saturation or switches between positive and negative saturation levels. For this reason, the open loop configurations are not used in linear applications.

5.2 OP-AMP WITH NEGATIVE FEEDBACK:

Returning to the circuit of Kirchoff's voltage equation for the input loop is

$$v_{id} = v_{in} - v_f \qquad \text{--------(5.1)}$$

Where v_{in} = input voltage

V_f = feedback voltage

V_{id} =difference voltage

However, recall that an op-amp always amplifies the difference input voltage v_{id}. From eqn (5.1) the difference voltage is equal to input voltage v_{in} minus the feedback voltage v_f, in other words the feedback voltage opposes the input voltage (or is out of phase by 180^0 with respect to the input voltage). Hence the feedback is said to be negative. Returning to the analysis of voltage series feedback amplifier we should note that, it is performed by computing closed loop voltage gain, input resistance, output resistance, and bandwidth.

5.3 VOLTAGE SERIES FEEDBACK:

The schematic diagram of the voltage series feedback amplifier is shown in fig (5.4). The op-amp is represented by its schematic symbol including the large signal voltage gain A, and the feedback circuit is composed of two resistances R_1 and R_F.

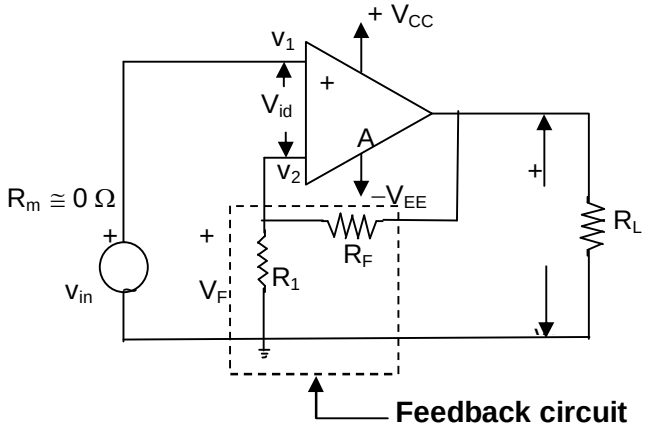


Figure 5.4 voltage series feedback amplifier

The circuit shown in fig (5.4) is commonly known as a non- inverting amplifier with feedback (or closed loop non-inverting amplifier) because it uses feedback, and the input signal is applied to the non-inverting input terminal of the op-amp.

Before proceeding, it is necessary to define some important terms for the voltage series feedback amplifier of fig (5.4). Especially the voltage gain with and without feedback, and the gain of the feedback circuit are defined as follows.

Open loop voltage gain (or gain without feedback) $A = \frac{v_o}{v_{id}}$.

Closed-loop voltage gain (or gain with feedback) $A_f = \frac{v_o}{v_{in}}$

Gain of the feedback current $B = \frac{v_f}{v_o}$

5.4 EFFECT OF FEEDBACK ON CLOSED LOOP GAIN INPUT RESISTANCE OUTPUT RESISTANCE BANDWIDTH AND OUTPUT OFFSET VOLTAGE:

5.4.1 Input Resistance with Feedback:

A voltage series feedback amplifier with the op-amp equivalent circuit is shown in fig (5.5).In this circuit R_i is the input resistance (open-loop) of the op-amp, and R_{iF} is the input resistance of the feedback amplifier. The input resistance with feedback is defined as

$$R_{if} = \frac{v_{in}}{i_{in}} = \frac{v_{in}}{v_{id}/R_i} \qquad \text{---------- (5.2)}$$

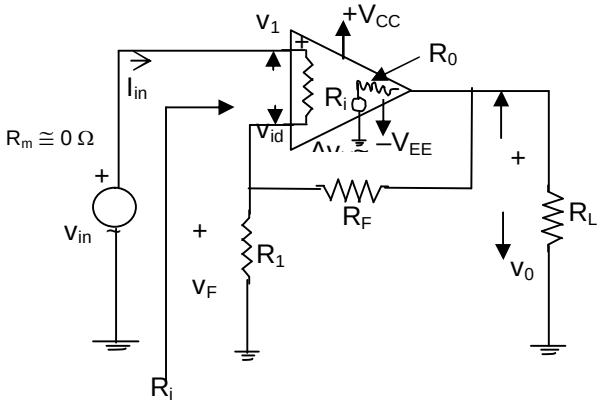


Figure 5.5 Derivation of Input Resistance with Feedback

However,

$$V_{id} = \frac{v_0}{A} \quad \text{and} \quad v_0 = \frac{A}{1+AB} v_{in} \qquad \text{---------- (5.3)}$$

$$R_{if} = R_i . \frac{v_{in}}{v_0 / A} \qquad \text{---------- (5.4)}$$

$$R_{if} = AR_i \frac{v_{in}}{\frac{Av_{in}}{(1+AB)}} = R_i (1+AB) \qquad \text{---------- (5.5)}$$

It means, the input resistance of the op-amp with feedback is (1+AB) times that without feedback.

5.4.2 Output Resistance with Feedback:

Output resistance is the resistance determined looking back into the feedback amplifier from the output terminal, as shown in figure 5.6. This resistance can be obtained by using Thevenin's theorem for dependent sources. Specifically, to find output resistance with feed back R_{oF}, reduce the independent source v_{in} to zero, apply an external voltage v_0 and then calculate the resulting current i_o in short, the R_{oF} is defined as follows:

$$R_{oF} = \frac{v_o}{i_o}.$$

Writing Kirchoff currents equation at the output node N, we get

$i_o = i_a + i_b$ ----------(5.6)

Since $R_F + R_1 \parallel R_i >> R_0$, $i_a >> i_b$ therefore

$i_o \cong i_a$ ----------(5.7)

The current i_o can be found by writing Kirchoff's voltage equation for the output loop

$v_0 - R_0 i_0 - Av_{id} = 0$ ---------- (5.8)

$$i_0 = \frac{v_0 - Av_{id}}{R_0}$$

$$v_{id} = v_1 - v_2 = -v_f = -\frac{R_1 v_0}{R_1 + R_F} = -Bv_o$$

therefore $i_0 = \frac{v_0 + A.Bv_0}{R_0}$

substituting the value of i_o in equation, we get

$$R_{oF} = \frac{v_0}{(v_0 + ABv_0)/R_0} = \frac{R_0}{1+AB} \qquad \text{---------- (5.9)}$$

This result shows that the output resistance of the voltage series feedback amplifier is 1/(1+AB) times the output resistance R_o of the op-amp. That is, the output resistance of the op-amp with feedback is much smaller than the output resistance without feedback.

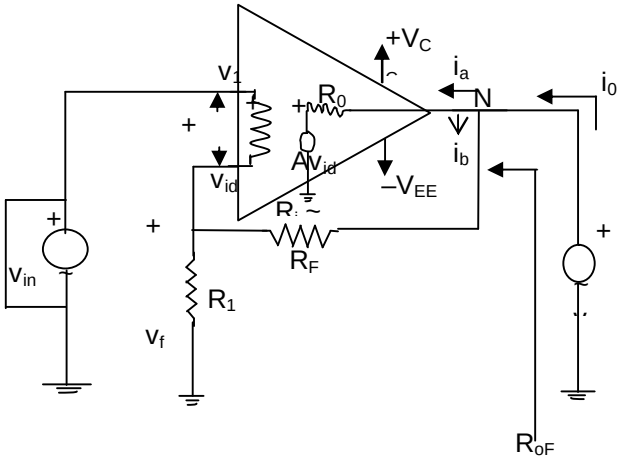


Figure 5.6 Derivation of Output Resistance with Feedback

5.4.3 Bandwidth with feedback:

The bandwidth of an amplifier is defined as the band (range) of frequencies for which the gain remains constant. Manufacturers generally specify either the gain bandwidth product or supply open loop gain versus frequency curve for the op-amp. For the 741 op-amp the latter is typical.

The open loop gain versus frequency curve of the 741C op-amp is shown in Figure 2.4. From this curve for a gain of 200,000 the bandwidth is approximately 5Hz. In other extreme, the bandwidth is approximately 1MHz when the gain is unity. The frequency at which gain equals 1 is known as unity gain-band width. It is the maximum frequency the op-amps can be used for. Furthermore, the gain bandwidth product obtained from the open loop gain versus frequency curve of fig 5.6 is equal to the unity gain bandwidth of an op-amp. However this holds true only for those op-amps like 741, which have just one break frequency below unity gain bandwidth.

Since the gain bandwidth is constant, obviously the higher the gain the smaller the bandwidth and vice versa. As we have seen if negative feedback is used gain A decreases to A/1+AB. Therefore, to obtain the closed loop bandwidth, the open loop bandwidth must be multiplied by the same factor, by which the gain is divided, that is, by the factor (1+AB). In short.

Band width with feedback = (band width without feedback) (1+AB)

$$f_F = f_0 (1+AB) \qquad \text{---------- (5.10)}$$

or alternatively,

f_F = unity gain band width/closed loop gain = U.G.B.W /A_F ----------(5.11)

The closed loop bandwidth can also be determined from the open loop gain versus frequency plot. To do this we locate the closed loop voltage gain value on the gain axis and draw a parallel line through this value to the frequency axis. Then we project the point of inter section of the line with the curve on the frequency axis and read the value of the closed loop bandwidth. Using this procedure in figure 5.7. The bandwidth is approximately 100 KHz. for a closed loop gain of 10.

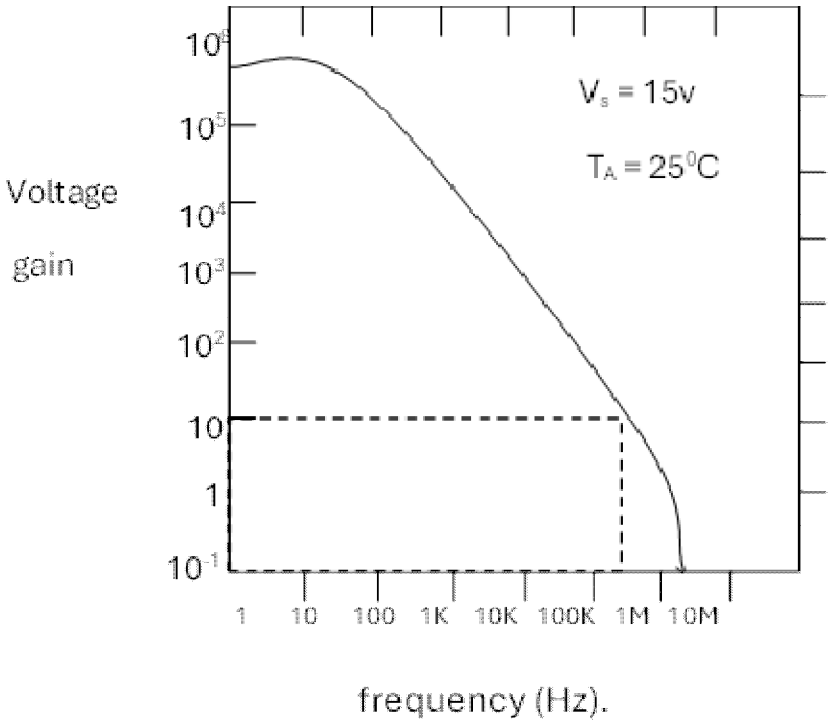


Figure 5.7 Open Loop Gain Versus Frequency Curve

5.5 VOLTAGE FOLLOWER:

The lowest gain that can be obtained from a non-inverting amplifier with feedback is 1. When the non-inverting amplifier is configured for unity gain, it is called a voltage follower because the output voltage is equal to and in phase with the input. In other words, in the voltage follower, the output follows the input.

Although it is similar to the discrete emitter follower, the voltage follower is preferred because it has much higher input resistance, and the output amplitude is exactly equal to the input.

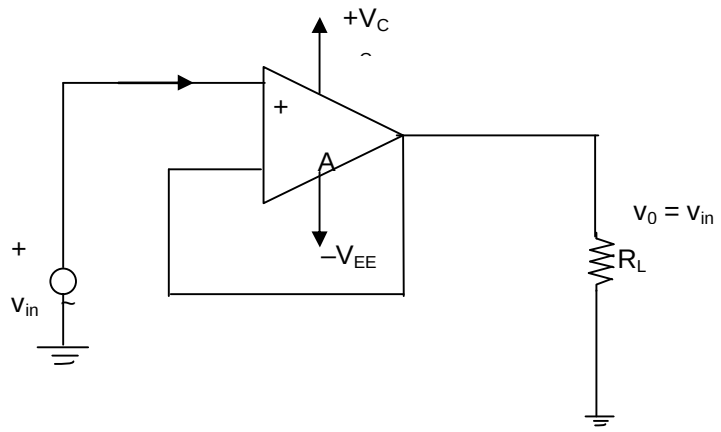


Figure 5.8 Voltage Follower

To obtain the voltage follower from the non-inverting amplifier of figure (5.8), simply open R_1 and short R_F. The resulting circuit is shown in figure 5.8. In this figure all the output voltage is feedback into the inverting terminal of the op-amp; consequently, the gain of the feedback circuit is 1. ($B = A_F = 1$).

Since the voltage follower is a special case of the non-inverting amplifier all the formulae developed for the latter are indeed applicable to the former except that the gain of the feedback circuit is 1 (B=1) the applicable formulae are:-

$A_F = 1$

$R_{iF} = A\ R_i$

$R_{oF} = R_o/A$

$f_F = Af_o$

$$voT = \frac{\pm v_{sat}}{A}$$

Since $(1+A) \cong A$

The voltage follower is also called a non-inverting buffer because when placed between two networks, it removes the loading on the first network.

5.6 SUMMARY:

This lesson explores amplifiers and feedback in operational amplifier (op-amp) circuits, focusing on both open-loop and closed-loop configurations. Open-loop op-amp configurations include differential, inverting, and non-inverting amplifiers, which exhibit extremely high gain but are unsuitable for linear applications due to saturation issues. Introducing negative feedback improves amplifier stability, reduces distortion, and controls

gain, input resistance, output resistance, and bandwidth. Voltage series feedback, commonly used in non-inverting amplifiers, enhances performance by reducing output resistance and increasing input resistance. The lesson also covers the voltage follower, a special case of a non-inverting amplifier with unity gain, used as a buffer to prevent loading effects between circuits. Understanding these principles helps in designing efficient amplification circuits for various applications in electronics, ensuring stable signal processing and improved performance of op-amp-based systems.

5.7 TECHNICAL TERMS:

Open-loop gain, Negative feedback, Differential amplifier, Voltage follower, Bandwidth, Input resistance

5.8 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are the key differences between open-loop and closed-loop configurations of an operational amplifier?
2) How does negative feedback improve the performance of an op-amp circuit?
3) What is the role of a differential amplifier in signal processing?

Short Answer Questions:

1) Why is the voltage follower used as a buffer in electronic circuits?
2) How does feedback affect the gain, bandwidth, and input resistance of an amplifier?
3) What are the advantages of using a non-inverting amplifier with voltage series feedback?

5.9 SUGGESTED READINGS:

1) Sedra, A.S., & Smith, K.C. (2016). Microelectronic Circuits (7^{th} ed.). Oxford University Press.
2) Gayakwad, R.A. (2000). Op-Amps and Linear Integrated Circuits (4^{th} ed.). Pearson Education.
3) Malvino, A., & Bates, D.J. (2016). Electronic Principles (8^{th} ed.). McGraw-Hill Education.
4) Franco, S. (2014). Design with Operational Amplifiers and Analog Integrated Circuits (4^{th} ed.). McGraw-Hill Education.
5) Millman, J., & Grabel, A. (2009). Microelectronics (2^{nd} ed.). McGraw-Hill Education.

Prof. R.V.S.S.N. Ravi Kumar

LESSON-6

PRACTICAL OP-AMPS

6.0 AIM AND OBJECTIVES:

The aim of this lesson is to provide a comprehensive understanding of practical operational amplifiers (op-amps), their characteristics, and their various applications in electronic circuits. Op-amps are essential components in analog electronics, widely used in signal processing, control systems, and instrumentation. This lesson focuses on key parameters such as input offset voltage, input bias current, input offset current, and total output offset voltage, which affect the performance of an op-amp. Additionally, it explores the Common Mode Rejection Ratio (CMRR) and its frequency response, crucial for minimizing noise and enhancing signal integrity. The objectives of this lesson are to explain the fundamental properties of op-amps and their real-world limitations, analyze how input bias current and offset voltage impact circuit performance, and understand techniques for offset voltage compensation. Furthermore, students will learn about summing amplifiers, scaling and averaging amplifiers, and their applications in signal processing. The lesson also covers the design and working principles of integrator and differentiator circuits, enabling students to apply op-amps effectively in various analog computing and signal conditioning tasks.

STRUCTURE:

6.1 INPUT OFFSET VOLTAGE:

6.1.1 Input Offset Voltage: Input offset voltage is the voltage that must be applied between the two input terminals of an op-amp to null the output as shown in figure6.1. In figure6.1V_{dc1} and V_{dc2} are dc voltages and R_s represents source resistance. We denote input offset voltage by V_{i0}. This voltage V_{i0} could be positive or negative.

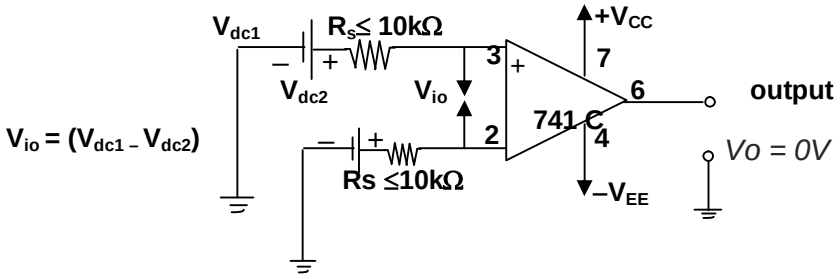


Figure 6.1: Defining Input Offset Voltage V_{i0}

For 741C, the maximum value of V_{i0} is 6 mV DC. The smaller the value of v_{i0}, the better the input terminals are matched.

6.2 INPUT BIAS CURRENT:

Input bias current I_B is the average of the currents that flow into the inverting and non-inverting input terminals of the op-amp. In equation form

$$\mathbf{I_B} = \frac{I_{B1} + I_{B2}}{2} \qquad \rightarrow \mathbf{(6.1)}$$

I_B = 500 nA maximum for the 741C, whereas I_B for the precision 741C is ± 7 nA

Note: that the input currents I_{B1} and I_{B2} are actually the base currents of the first differential amplifier stage.

6.3 INPUT OFFSET CURRENT:

The algebraic difference between the currents into the inverting and non-inverting terminals is referred to as input offset current I_{i0} (see fig.6.2). In the form of an equation.

$$\mathbf{I_{i0} = |I_{B1} - I_{B2}|} \qquad \rightarrow \mathbf{(6.2)}$$

Where $I_{b1}$is the current into the non-inverting input I_{b2} is the current into the inverting input

The offset current input for the 741C is 200 nA maximum. As the matching between the two input terminals is improved, the difference between I_{B1} and I_{B2} becomes smaller; that is, the I_{i0} value decreases further.

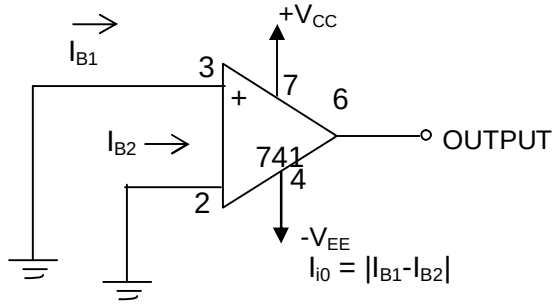


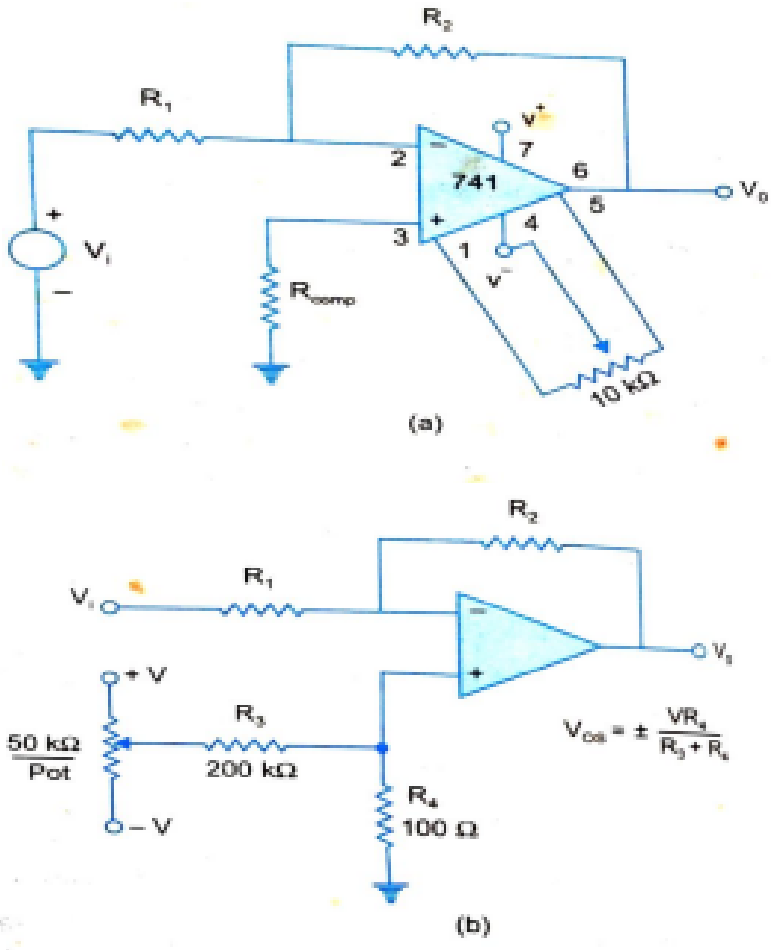
Figure 6.2: Defining Input Offset Current I_{i0}

6.4 TOTAL OUTPUT OFFSET VOLTAGE:

The total output offset voltage VOT could be either more or less than the offset voltage produced at the output due to input bias current (IB) or input offset voltage alone(Vos).

Total output offset voltage with feedback = (Total output offset voltage without feedback) / (1+AB)

Total output offset voltage Compensation:



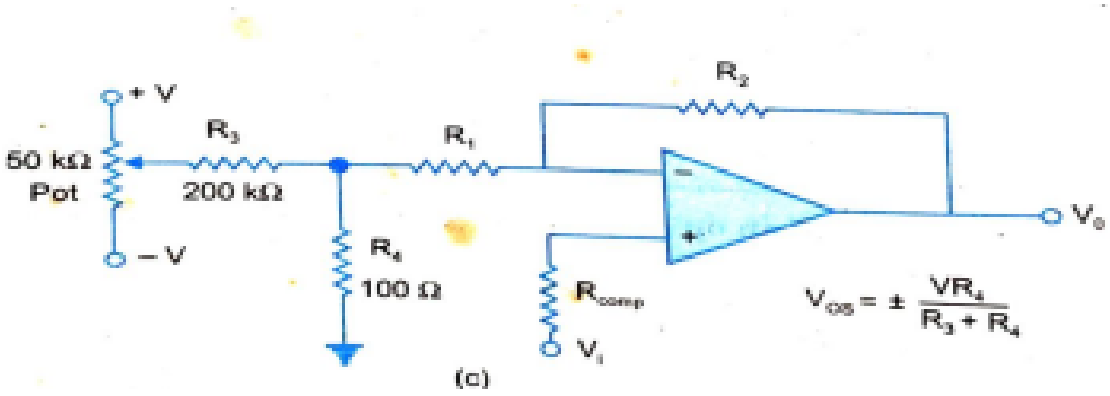


Figure 6.3 (a) Offset Null Pin Connection for µA741. Balancing Circuit for

(b) Inverting Amplifier (c) Non-Inverting Amplifier

6.5 CMRR FREQUENCY RESPONSE:

6.5.1 Common mode rejection ratio: The common-mode rejection ratio (CMRR) is defined in several essentially equivalent ways by the various manufactures. Generally, it can be defined as the ratio of the differential voltage gain A_d to the common mode voltage gain A_{cm}. That is

$$\textbf{CMRR} = \frac{A_d}{A_{cm}} \qquad \rightarrow \textbf{(6.3)}$$

The differential voltage gain A_d is the same as the large signal voltage gain A, which is specified on the data sheets; The common mode voltage gain can be determined from the circuit.

$$\mathbf{A_{cm}} = \frac{V_{ocm}}{V_{cm}} \qquad \rightarrow \textbf{(6.4)}$$

V_{0cm} = o/p common mode voltage.

V_{cm} = i/p common mode voltage.

A_{cm} = common mode voltage gain.

Generally, the A_{cm} is very small and $A_d = A$ is very large; therefore the CMRR is very large, being a large value, CMRR is most often expressed is decibels (dB) for the 741C, CMRR is 90 dB typically.

6.6 SUMMING AMPLIFIER:

The Summing Amplifier is another type of operational amplifier circuit configuration that is used to combine the voltages present on two or more inputs into a single output voltage.

We saw previously in the inverting operational amplifier that the inverting amplifier has a single input voltage, (Vin) applied to the inverting input terminal. If we add more input resistors to the input, each equal in value to the original input resistor, (Rin) we end up with another operational amplifier circuit called a Summing Amplifier, "summing inverter" or even a "voltage adder" circuit as shown below.

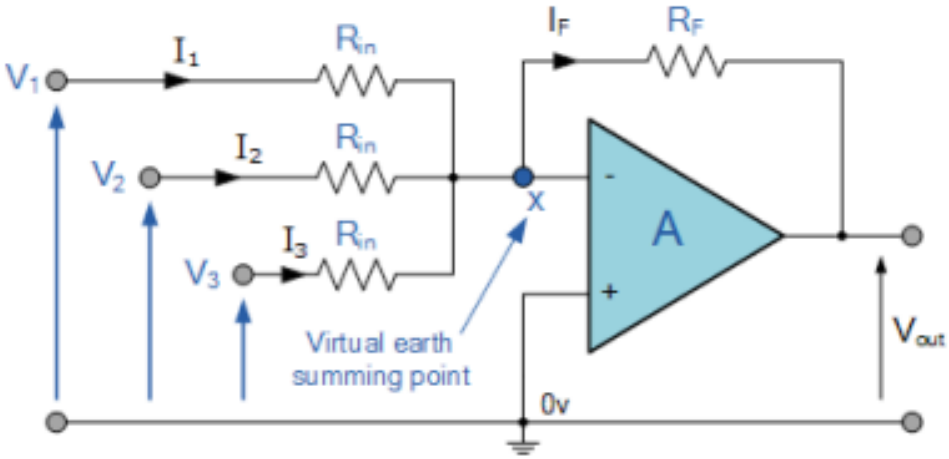


Figure 6.4: Summing Amplifier Circuit

In this simple summing amplifier circuit, the output voltage, (Vout) now becomes proportional to the sum of the input voltages, V, V , V , etc. Then we can modify the original equation for the inverting amplifier to take account of these new inputs thus:

$$\mathbf{IF = I_1+I_2+I_3 = -\left[\frac{V1}{Rin} + \frac{V2}{Rin} + \frac{V3}{Rin}\right]} \qquad \rightarrow \mathbf{(6.5)}$$

Inverting Equation: Vout = $\mathbf{-\frac{Rf}{Rin}^{x}}$ **Vin**

$$\mathbf{Then, -V\ out = \left[\frac{R_F}{Rin}V1 + \frac{R_F}{Rin}V2 + \frac{R_F}{Rin}V3\right]} \qquad \rightarrow \mathbf{(6.6)}$$

However, if all the input impedances, (R) are equal in value, we can simplify the above equation to give an output voltage of:

6.6.1 Summing Amplifier Equation:

$$\mathbf{-Vout = \frac{R_F}{R_{IN}}\,[V1+V2+V3\ldots.etc]} \qquad \rightarrow \mathbf{(6.7)}$$

We now have an operational amplifier circuit that will amplify each individual input voltage and produce an output voltage signal that is proportional to the algebraic "SUM" of the three individual input voltages V, V and V. We can also add more inputs if required as each individual input "see's" their respective resistance, Rin as the only input impedance. This is

because the input signals are effectively isolated from each other by the "virtual earth" node at the inverting input of the op-amp. A direct voltage addition can also be obtained when all the resistances are of equal value and Rf is equal to Rin. Note that when the summing point is connected to the inverting input of the op-amp the circuit will produce the negative sum of any number of input voltages. Likewise, when the summing point is connected to the non-inverting input of the op-amp, it will produce a positive sum of the input voltages.

6.6.2 Summing Amplifier Applications:

So, what can we use summing amplifiers for, either inverting or non-inverting. If the input resistances of a summing amplifier are connected to potentiometers the individual input signals can be mixed by varying amounts.

For example, measuring temperature, you could add a negative offset voltage to make the output voltage or display read "0" at the freezing point or produce an audio mixer for adding or mixing individual waveforms (sounds) from different source channels (vocals, instruments, etc) before sending them combined to an audio amplifier.

6.7 SCALING AND AVERAGING AMPLIFIERS:

6.7.1 Scaling and Weighted Amplifiers:

If each input voltage is amplified by a different factor, in other words, weighted differently at the output, the circuit is then called a scaling or weighted amplifier. This condition can be accomplished if R_a, R_b and R_c are different in value. Thus, the output voltage of the scaling amplifier is

$$\mathbf{V_o = -\left(\frac{R_F}{R_a}V_a + \frac{R_F}{R_b}V_b + \frac{R_F}{R_c}V_c\right)} \qquad \rightarrow \ \mathbf{(6.8)}$$

Were

$$\mathbf{\frac{R_F}{R_a} \neq \frac{R_F}{R_b} \neq \frac{R_F}{R_c}}$$

6.7.2 Averaging Amplifiers:

The output voltage is equal to the average of all input voltages times the gain of the circuit $(1+\frac{R_F}{R_1})$, hence the name averaging amplifier. Depending on the application requirement, the gain $(1+\frac{R_F}{R_1})$ can be set to specific value. Obviously, if the gain is 1, the output voltage will be equal to the average of all input voltages.

Note that there are two basic differences between this averging amplifier and that using the inverting configuration;

1) No sign change or phase reversal occurs between the average of the inputs and outputs.
2) The non inverting input voltage V_1 is the average of all inputs, whereas in the inverting averaging amplifier the output is the average of all inputs, with a negative sign.

6.8 INTEGRATOR AND DIFFERENTIATOR:

The electronic circuits which perform mathematical operations such as differentiation and integration are called differentiator and integrator, respectively. Here we will discuss in detailthe p-amp based differentiator and integrator.

6.8.1 Differentiator:

A differentiator is an electronic circuit that produces an output equal to the first derivative of its input. An op-amp based differentiator produces an output, which is equal to the differential of input voltage that is applied to its inverting terminal. The circuit diagram of an op-amp based differentiator is shown in the following figure

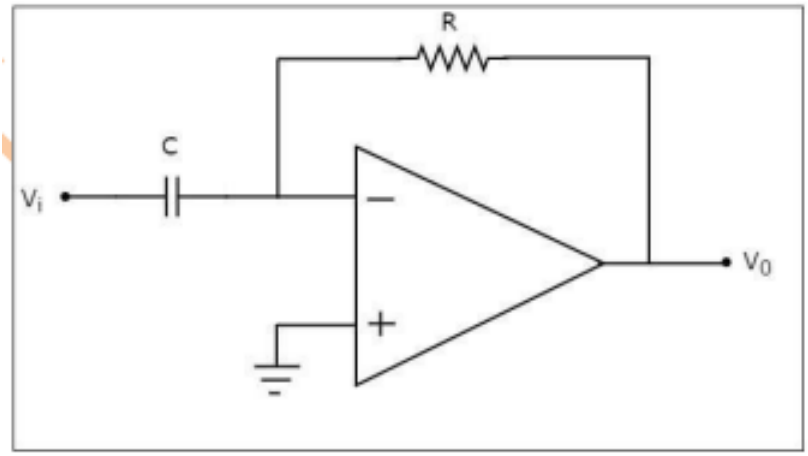


Figure 6.5: Op-Amp Based Differentiator

In the above circuit, the non-inverting input terminal of the op-amp is connected to ground. That means zero volts is applied to its non-inverting input terminal. According to the virtual short concept, the voltage at the inverting input terminal of op-amp will be equal to the voltage present at its non-inverting input terminal. So, the voltage at the inverting input terminal of op-amp will be zero volts. The nodal equation at the inverting input terminal's node is

$$C\,\frac{d(0-V_i)}{dt} + \frac{0-V_i}{R} = 0 \qquad \rightarrow (6.9)$$

$$-C\frac{dV_i}{dt} = \frac{V_0}{R}$$

$$V_0 = -RC\frac{dV_i}{dt} \qquad \rightarrow (6.10)$$

If RC = 1 sec, then the output voltage V_o will be-

$$V_0 = -\frac{dV_i}{dt} \qquad \rightarrow (6.11)$$

Thus, the op-amp based differentiator circuit shown above will produce an output, which is the differential of input voltage Vi, when the magnitudes of impedances of resistor and capacitor are reciprocal to each other. Note that the output V0 is having a negative sign, which indicates that there exists a 180o phase difference between the input and the output.

6.8.2 Integrator:

An integrator is an electronic circuit that produces an output that is the integration of the applied input. This section discusses the op-amp based integrator. An op-amp based integrator produces an output, which is an integral of the input voltage applied to its inverting terminal. The circuit diagram of an op-amp based integrator is shown in the following figure:

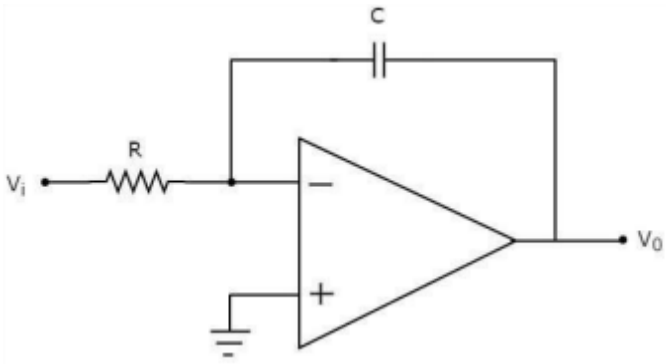


Figure 6.6: Op-Amp Based Integrator

In the circuit shown above, the non-inverting input terminal of the op-amp is connected to ground. That means zero volts is applied to its non-inverting input terminal. According to virtual short concept, the voltage at the inverting input terminal of op-amp will be equal to the voltage present at its noninverting input terminal. So, the voltage at the inverting input terminal of opamp will be zero volts.

The nodal equation at the inverting input terminal is:

$$\frac{0-V_i}{R} + C\frac{d(0-V_i)}{dt} = 0 \qquad \rightarrow (6.12)$$

$$\frac{-V_i}{R} = C\frac{dV_0}{dt}$$

$$\frac{dV_0}{dt} = \frac{-V_i}{RC}$$

$$dV_0 = \left(\frac{-V_i}{RC}\right)dt \qquad \rightarrow (6.13)$$

Integrating both sides of the equation shown above, we get-

$$\int dV_0 = \left(\frac{-V_i}{RC}\right)dt$$

$$V_0 = -\frac{1}{RC}\int V_i\, dt \qquad \rightarrow (6.14)$$

If RC=1sec RC= 1sec, then the output voltage, V_o will be-

$$V_0 = -\int V_i\, dt \qquad \rightarrow (6.15)$$

So, the op-amp based integrator circuit discussed above will produce an output, which is the integral of input voltage Vi, when the magnitude of impedances of resistor and capacitor are reciprocal to each other.

6.9 SUMMARY:

This lesson provided an in-depth understanding of practical operational amplifiers (op-amps) and their real-world characteristics. It covered key parameters such as input offset voltage, input bias current, and input offset current, which influence the accuracy and stability of op-amp circuits. The concept of total output offset voltage and compensation techniques was also discussed. Additionally, the lesson explored the Common Mode Rejection Ratio (CMRR) and its significance in minimizing unwanted noise and improving signal quality. The summing amplifier was introduced as a useful configuration for combining multiple input signals, along with scaling and averaging amplifiers for weighted signal processing. Finally, the lesson explained the principles of integrators and differentiators, which perform mathematical operations on input signals. By understanding these concepts, students can effectively design and implement op-amp circuits for various applications in signal processing, control systems, and instrumentation, ensuring accurate and reliable electronic designs.

6.10 TECHNICAL TERMS:

Input Offset Voltage (Vi0), Input Bias Current (IB), Common Mode Rejection Ratio (CMRR), Summing Amplifier, Integrator, Differentiator.

6.11 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What is input offset voltage (Vi0) in an operational amplifier, and how does it affect circuit performance?
2) Define input bias current (IB) and explain its significance in op-amp circuits.
3) What is the Common Mode Rejection Ratio (CMRR), and why is it important in operational amplifier applications?

Short Answer Questions:

1) How does a summing amplifier work, and what are its practical applications?
2) What is the difference between an integrator and a differentiator, and in which applications are they commonly used?
3) How can total output offset voltage be compensated in an op-amp circuit to improve accuracy?

6.12 SUGGESTED READINGS:

1) **Ramakanth A. Gayakwad**, Op-Amps and Linear Integrated Circuits, Pearson Education – A comprehensive textbook covering operational amplifier fundamentals and applications.
2) **Robert F. Coughlin, Frederick F. Driscoll**, Operational Amplifiers and Linear Integrated Circuits, Pearson – Detailed explanations on op-amp characteristics, configurations, and real-world applications.
3) **Sergio Franco**, Design with Operational Amplifiers and Analog Integrated Circuits, McGraw-Hill – Advanced concepts in op-amp design, including offset voltage compensation and frequency response.
4) **David A. Bell**, Operational Amplifiers and Linear ICs, Oxford University Press – Covers op-amp characteristics, summing amplifiers, integrators, differentiators, and practical circuit design techniques.
5) **Sedra & Smith**, Microelectronic Circuits, Oxford University Press – Discusses the mathematical analysis of op-amp circuits, including CMRR and signal processing applications.
6) **Analog Devices**, Op-Amp Applications Handbook – A practical guide to operational amplifier applications, available online through Analog Devices' technical resources.

Prof. R.V.S.S.N. Ravi Kumar

LESSON-7

OSCILLATORS, MULTIVIBRATORSAND WAVEFORM GENERATORS

7.0 AIM AND OBJECTIVES:

The aim of this study is to analyze the principles, types, and applications of oscillators, including their frequency stability, phase shift, and waveform generation. Oscillators play a crucial role in electronic circuits, generating continuous waveforms such as sinusoidal, square, and triangular signals. This study focuses on different oscillator types, including RC, LC, crystal, Wein bridge, and multivibrators, highlighting their working principles and frequency stability factors. Additionally, it explores the impact of temperature, power supply variations, and circuit components on oscillator performance. The objective is to understand oscillator circuit design, the role of feedback in sustaining oscillations, and practical applications in communication, signal processing, and measurement systems. By examining key equations and circuit configurations, this study aims to enhance knowledge of oscillator functionality and its significance in modern electronics.

STRUCTURE:

7.1 OSCILLATORS PRINCIPLES:

An Oscillator, as such, is a type of feedback amplifier in which part of the output is feedback to the input via a feedback circuit. If the signal feedback is of proper magnitude and phase, the circuit produces alternating currents or voltages. To visualize the requirement of an oscillator, consider the block diagram of figure 7.1.

However, here the input voltage is zero (vin = 0). Besides that, the feedback is positive because most oscillators use positive feedback. Finally, the closed loop gain of the amplifier is denoted by Av rather that AF.

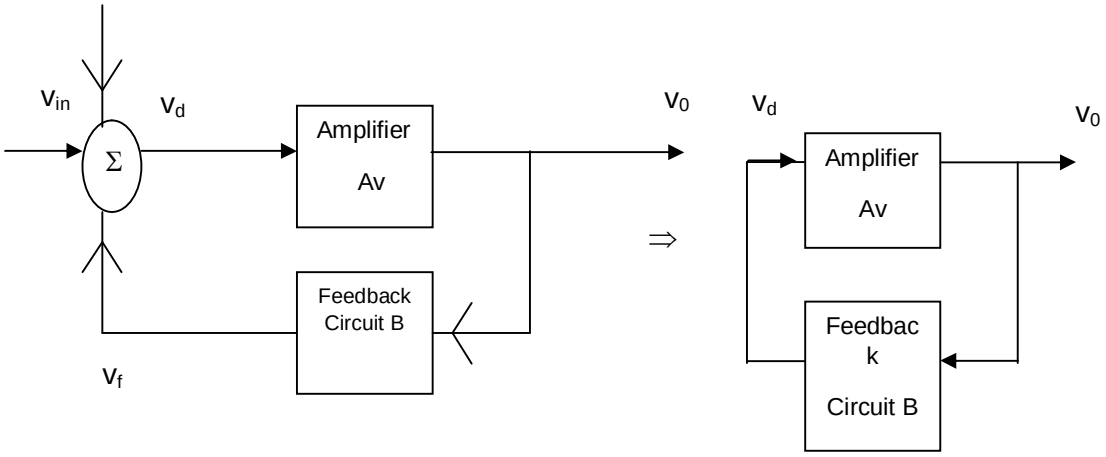


Figure 7.1 Oscillator Block Diagram

In figure1.1

vd = vf + vin ---------- (7.1)

v0 = Av vd ---------- (7.2)

vf = B v0 ---------- (7.3)

Using these relationships, the following equations are obtained ---------- (7.4)

From equations (7.2) & (7.3), we have = AvB ---------- (7.5)

Substitute equation (7.5) is equation (7.4), we have = ---------- (7.6)

However, vin = 0, and v_0 = 0 implies that

Av B = 1

Expressed in polar form

Av B = 1 and Av B = 0 or 3600 ---------- (7.7)

Equation (7.7) gives the two requirements for oscillators:

1) The magnitude of the loop gain Av B must be at least 1, and
2) The total phase of the loop gain Av B must be equal to 00 or 3600. For instance, as indicated in fig (1.1), if the amplifier causes a phase shift of 1800.

The feedback circuit must provide an additional phase shift of 1800, so that the total phase shift around the loop is 3600. The waveforms shown in fig7.1 are sinusoidal and are used to illustrate the circuit's action. The type of waveform generated by an oscillator depends on the components in the circuit and hence, may be sinusoidal, square or triangular. In addition, the frequency of the oscillation is determined by the components in the feedback circuit.

7.2 OSCILLATOR TYPES:

Because of their widespread use, many different types of oscillators are available. These oscillators are summarized as follows.

Types of Oscillator	Frequency of Oscillation	Types of Waveforms Generated
1. RC Oscillator	1. Audio frequency (AF)	1. Sinusoidal
2. LC Oscillator	2. Radio frequency (RF)	2. Square wave
3. Crystal Oscillator		3. Triangular wave
		4. Saw-tooth wave

7.3 FREQUENCY STABILITY:

The ability of the oscillator circuit to produce waves at one exact frequency is called frequency stability. Although there may be a number of factors that cause changes in oscillator frequency, the primary factors are temperature changes and changes in the dc power supply. Temperature and power supply changes cause variations in the op-amp's gain, in junction capacitances and resistances of the transistors in an op-amp, as well as in an external circuit component. In most cases these variations can be kept small by careful design, using regulated power supplies, and by temperature control.

Another important factor that determines frequency stability is the figure of merit Q of the circuit. The higher the Q, the greater the stability. For this reason, crystals oscillators are far more stable than RC and LC oscillators, especially at higher frequencies. LC circuit and crystal are generally used for the generation of high frequency signals, while RC components are most suitable for audio frequency applications.

7.4 THE PHASE SHIFT OSCILLATOR:

Fig. 7.2 shows a phase shift oscillator, which consists of an op-amp, as amplifying stage and three RC cascaded networks in the feedback circuit. The feedback circuit provides feedback voltage from the output back to the input of the amplifier. The op-amp is used in the inverting mode; therefore, any signal that appears at the inverting terminal is shifted by 180^0 at the output. An additional 180^0 phase shift required for oscillation is provided by the cascaded RC networks. Thus, the total shift around the loop is 360^0 (or 0^0).

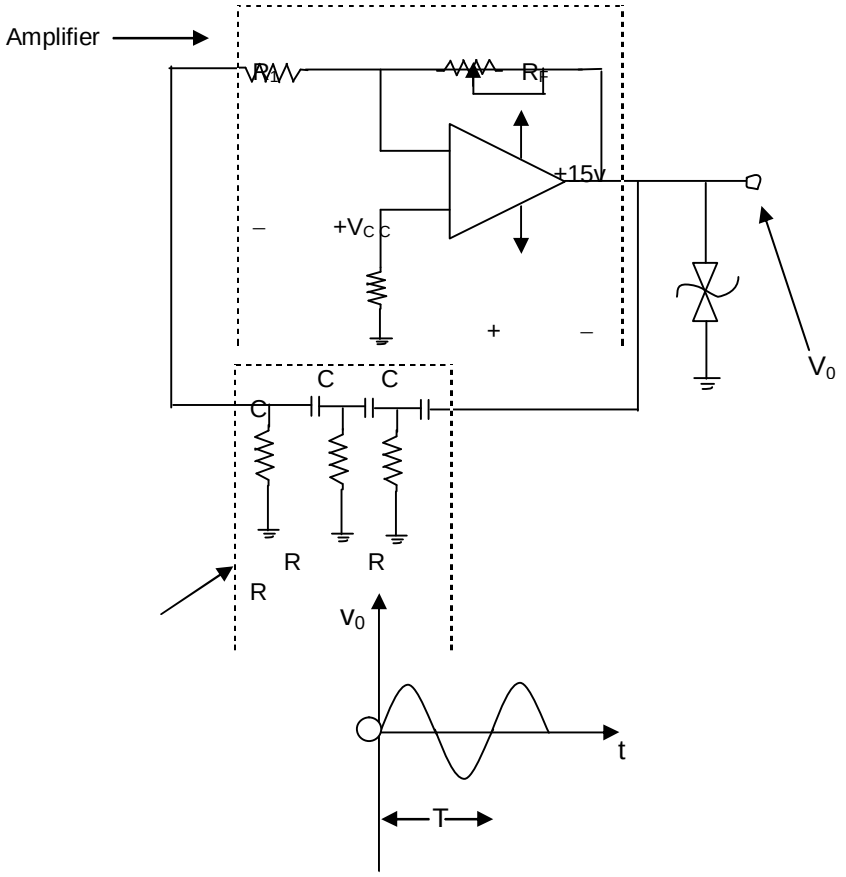


Figure 7.2: Phase Shift Oscillator and Its Output Waveform

At some specific frequency when the phase shift of the cascaded RC networks is exactly 180^0 and the gain of the amplifier is sufficiently large, the circuit will oscillate at that frequency. This frequency is called the frequency of oscillation f_0 and is given by

$$f_0 = \frac{1}{2\pi\sqrt{6}RC} = \frac{0.065}{RC} \qquad \text{---------- (7.8)}$$

At this frequency, the gain A_v must be at least 29. That is

$$\frac{R_F}{R_1} = 29$$

$R_F = 29\ R_1$ ---------- (7.9)

Then the circuit will produce a sinusoidal waveform of frequency f_0, if the gain is 29 and the total phase shift around the circuit is exactly 360^0. For a desired frequency of oscillator, choose a capacitor C_1 and then calculate the value of R from equation (3.8). A desired output amplitude, however, can be obtained with back-to-back Zeners connected at the output terminal.

7.4.1 Analysis:

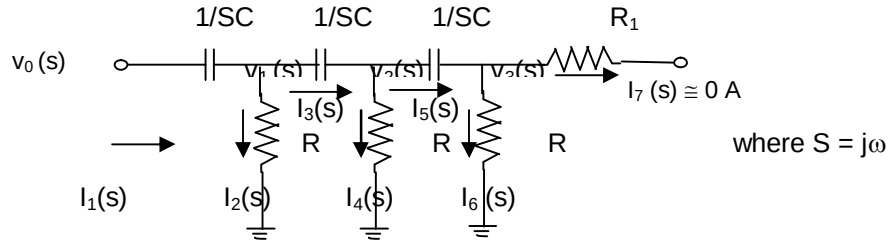


Figure 7.3: RC Network of the Phase Shift Oscillator

Nodal equations are transformed into the s domain

Writing Kirchoff's current law (KCL) at node $v_1(s)$, we get $I_1(s) = I_2(s) + I_3(s)$

$$\frac{v_0(s) - v_1(s)}{1/SC} = \frac{v_1(s)}{R} + \frac{v_1(s) - v_2(s)}{1/SC}$$

$$v_1(s)\left[\frac{1}{R} + 2SC\right] + v_2(s)(-SC) = v_0(s)(SC) \quad \text{---------- (7.10)}$$

Writing KCL at node v_2 (s)

$I_3(s) = I_4(s) + I_5(s)$

$$\frac{v_1(s) - v_2(s)}{1/SC} = \frac{v_2(s)}{R} + \frac{v_2(s) - v_3(s)}{1/SC}$$

$$v_1(s)[-SC] + v_2(s)\left[\frac{1}{R} + 2SC\right] + v_3(s)[-SC] = 0 \quad \text{---------- (7.11)}$$

Writing KCL at node v_3(s)

$I_5(s) = I_6(s)$

$$\frac{v_2(s) - v_3(s)}{1/SC} = \frac{v_3(s)}{R}$$

$$v_2(s)\,[SC] + v_3(s)\left[-\frac{1}{R} - SC\right] = 0 \qquad \text{---------- (7.12)}$$

$$v_2(s) = v_3(s)\left[\frac{1}{RSC} + 1\right]$$

Substituting this in equation (7.11)

$$v_1(s)[-SC] + v_3(s)\left[1 + \frac{1}{SRC}\right]\left[\frac{1}{RSC} + 2\right] + v_3(s).(-SC) = 0$$

$$v_1(s) = v_3(s)\left[-1 + \left(1 + \frac{1}{RSC}\right)\left(2 + \frac{1}{RSC}\right)\right]$$

Substituting this in equation (7.11)

$$v_3(s)\left[-1 + \left(1 + \frac{1}{RSC}\right)\left(2 + \frac{1}{RSC}\right)\right]\left[\frac{1}{R} + 2SC\right] + v_3(s)\left[\frac{1}{RSC} + 1\right](-SC) = v_0(s)(SC)$$

$$v_3(s)\left\{-1 + \left(1 + \frac{1}{RSC}\right)\left(2 + \frac{1}{RSC}\right)\left(\frac{1}{R} + 2SC\right)\right\} + v_3(s)\left((-SC) - \frac{1}{R}\right) = v_0(s)(SC)$$

$$v_3(s)\left[-\frac{1}{R} - 2sC + \left[\frac{1}{RSC} + 1\right]\left[2 + \frac{1}{RSC}\right]\left[\frac{1}{R} + 2SC\right] - SC - \frac{1}{R}\right] = v_0(s)(SC)$$

$$v_3(s)\left[(-3sC - \frac{2}{R}) + \left(\frac{RSC + 1}{RSC}\right)\left(\frac{2RSC + 1}{RSC}\right)\left(\frac{1 + 2RSC}{R}\right)\right] = v_0(s)(SC)$$

$$v_3(s)\left[\frac{\left(-3R^3S^3C^3 - 2R^2S^2C^2\right) + (1 + RSC)\left(1 + 4R^2S^2C^2 + 4RSC\right)}{R^3S^2C^2}\right] = v_0(s)(SC)$$

$$v_3(s)\,[R^3S^3C^3 + 6R^2S^2C^2 + 5RSC + 1] = v_0(s)\,[R^3S^3C^3]$$

$$\frac{v_3(s)}{v_0(s)} = \frac{S^3R^3C^3}{S^3R^3C^3 + 6S^2R^2C^2 + 5SRC + 1} \quad \text{---------- (7.13 (a))}$$

Next consider the op-amp part of the phase shift oscillator, the voltage gain of the op-amp

$$A_v = \frac{v_0(s)}{v_3(s)} = -\frac{R_F}{R_1} \quad \text{---------- (7.13(b))}$$

For an oscillator

$A_v B = + 1.$ ---------- (7.14)

Therefore, using equations 7.13(a) and 7.13(b), we have

Since it is an inverting amplifier $A_v = -\frac{R_F}{R_1}$

$$\left(-\frac{R_F}{R_1}\right)\left[\frac{S^3R^3C^3}{S^3R^3C^3 + 6R^2S^2C^2 + 5SRC + 1}\right] = 1$$

Substituting S = jω, and equating the real and imaginary parts, respectively, we get

$$\left(-\frac{R_F}{R_1}\right)\left[\frac{-j\omega^3R^3C^3}{-j\omega^3R^3C^3 - 6\omega^2R^2C^2 + j5\omega RC + 1}\right] = 1$$

$\left(-\frac{R_F}{R_1}\right)(-j\omega^3R^3C^3) = -j\omega^3R^3C^3 - 6\omega^2R^2C^2 + j5\omega RC + 1$ ----------(7.15)

$-6\omega^2R^2C^2 + 1 = 0$ (Real)

$$\omega^2 = \frac{1}{6R^2C^2} \quad \text{---------- (7.15(a))}$$

$$f = \frac{1}{2\pi\sqrt{6}RC} \quad \text{---------- (7.16)}$$

$\left(-\frac{R_F}{R_1}\right)(-\omega^3R^3C^3) = -\omega^3R^3C^3 + 5\omega RC$ (imaginary)

$$\frac{R_F}{R_1} = -1 + \frac{5}{\omega^2 R^2 C^2}$$ ---------- (7.16 (a))

From equations 7.15(a) and 7.16(a) we have

$$\frac{R_F}{R_1} = 29$$ ---------- (7.17)

7.5 WEIN BRIDGE OSCILLATOR:

The circuit diagram for typical Wien bridge oscillator is shown in figure 7.4. This involves an RC bridge circuit in which a frequency adjusting network is constructed of series combinations of R_1 and C_1 connected from the output of the op-amp to the non-inverting input of the op-amp and of a parallel combination of R_2 and C_2 from the non-inverting input of the op-amp to the ground connections. The resistances R_3 and R_4 from the non-inverting input of the op-amp to the ground connections. The resistors R_3 and R_4 make the feedback path and determine the gain of the amplifier. The bridge circuit causes a phase shift of 360^0 between the output and input of the op-amp.

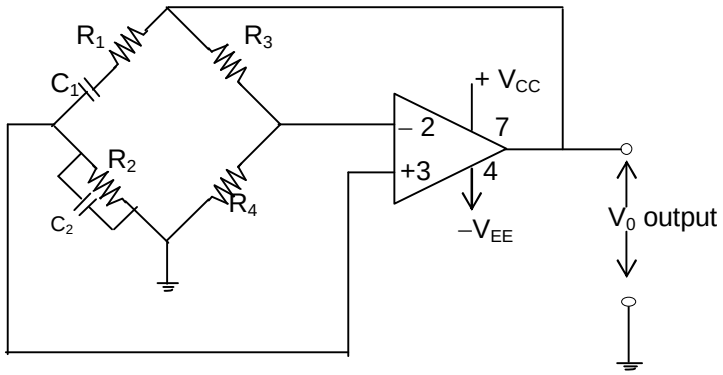


Figure 7.4(a) Circuit Configuration for Wien Bridge Oscillator

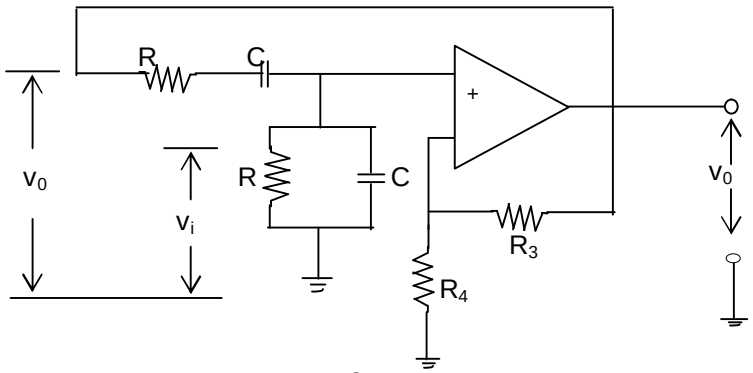


Figure 7.4(b) Redrawn Circuit for $R_1 = R_2 = R$, and $C_1 = C_2 = C$

In the circuit of fig.7.4, the voltage v_i across the parallel combination of R and C is fed back and is expressed as

$$v_i = \frac{v_0 Z_1}{Z_1 + Z_2} \quad \text{---------- (7.18)}$$

Where Z_1 = R||C = $\dfrac{R\frac{1}{j\omega C}}{R+\frac{1}{j\omega C}}$ and Z_2 (R in series with C) = $R + \dfrac{1}{j\omega C}$ ---------- (7.19)

$$v_i = \frac{v_0\left(\dfrac{R\frac{1}{j\omega C}}{R+\frac{1}{j\omega C}}\right)}{\left(\dfrac{R\frac{1}{j\omega C}}{R+\frac{1}{j\omega C}}\right)+\left(R+\frac{1}{j\omega C}\right)} \quad \text{---------- (7.20)}$$

$$v_i = v_0 \frac{R\frac{1}{j\omega C}}{R\frac{1}{jC} + R^2 + 2R.\frac{1}{j\omega C} - \frac{1}{\omega^2 C^2}}$$

$$= v_0 \frac{1}{1 + j\omega RC + 2 - \frac{j}{\omega RC}} = v_0 \frac{1}{3 + j\left(\omega RC - \frac{1}{\omega RC}\right)} \quad \text{---------- (7.21)}$$

In order to have zero phase between v_i and v_0, the j-terms must be zero i.e.,

$$\omega RC - \frac{1}{\omega RC} = 0;$$

$$\omega^2 = \frac{1}{R^2C^2}$$

$$\omega = \frac{1}{RC};\ f = \frac{1}{2\pi RC} \quad \text{---------- (7.22)}$$

Substituting this frequency in equation (7.23)

$$v_i = \frac{1}{3} v_0 \qquad \text{---------- (7.24)}$$

Equation 5.4 shows that there is a loss of voltage gain in the feedback network by a factor of 3. Therefore, the amplifier should have minimum voltage gain as 3 for oscillations.

The input voltage v_1 at the inverting point of the op-amp may be written as

$$v_i = \frac{v_0 R_4}{R_3 + R_4} \qquad \text{---------- (7.25)}$$

Equaling equation (7.24) and (7.25), we get

$$\frac{1}{3} = \frac{R_4}{R_3 + R_4}$$

$$R_3 = 2R_4 \qquad \text{---------- (7.26)}$$

Hence, for a stable operation of the oscillator: R_3 is selected slightly larger than $2R_4$ so that the maximum gain requirement is fulfilled.

7.6 MULTIVIBRATORS:

An electronic circuit that generates square waves (or other non-sinusoidal such as rectangular, saw-tooth waves) is known as a multivibrator. The name multivibrator is derived from the fact that a square wave consists of many (Fourier series analysis) sinusoidal of different frequencies. A multivibrator is a switching circuit which depends on positive feedback for operation. It is basically a two-stage amplifier with an output of one feedback to the input of the other as shown in Figure 7.5.

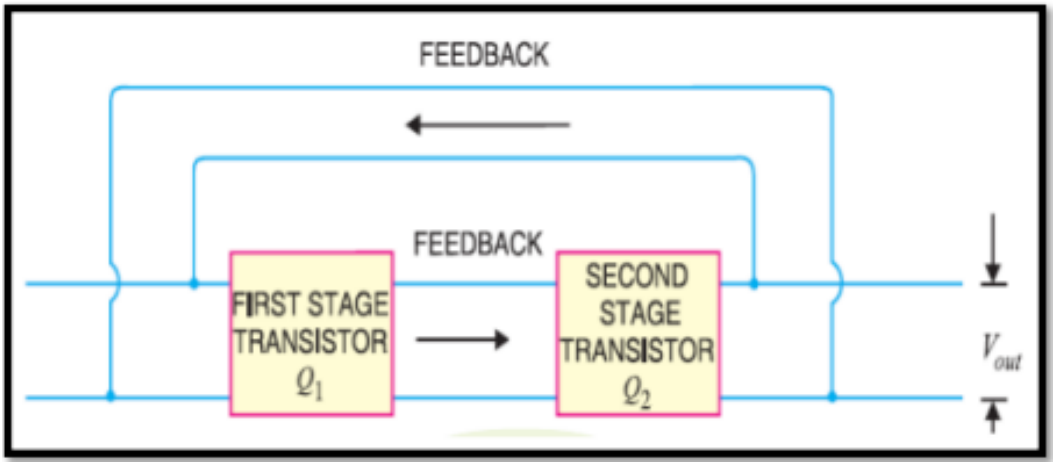


Figure 7.5: Show a Multivibrator

The circuit operates in two states (ON and OFF) controlled by circuit conditions. Each amplifier stage supplies feedback to the other in such a way that will drive the transistor of one stage to saturation (ON state) and the other to cut off (OFF state). After a certain time controlled by circuit conditions, the action is reversed i.e. the saturated stage is driven to cut off and the cut off stage is driven to saturation. The output can be taken across either stage and may be rectangular or square wave depending upon the circuit conditions. It is a two-stage amplifier with 100% positive feedback. Suppose output is taken across transistor Q2. At any instant, one transistor is ON and conducts IC (sat) while the other is OFF. Suppose Q2 is ON and Q1 is OFF. The collector current in Q2 will be IC (sat).

7.6.1 Types of Multivibrators:

A multivibrator is basically a two-stage amplifier with output of one feedback to the input of the other. At any particular instant, one transistor is ON and the other is OFF. After a certain time depending upon the circuit components, the stages reverse their conditions – the conducting stage suddenly cuts off and the non-conducting stage suddenly starts to conduct.

The two possible states of a multivibrator are:

	ON	OFF
First State	Q_1	Q_2
Second State	Q_2	Q_1

Depending upon the way in which the two stages interchange their states, the multivibrators are classified as:

(i) Astable or Free Running Multivibrator

(ii) Monostable or One-Shot Multivibrator

(iii) Bi-Stable or Flip-Flop Multivibrator

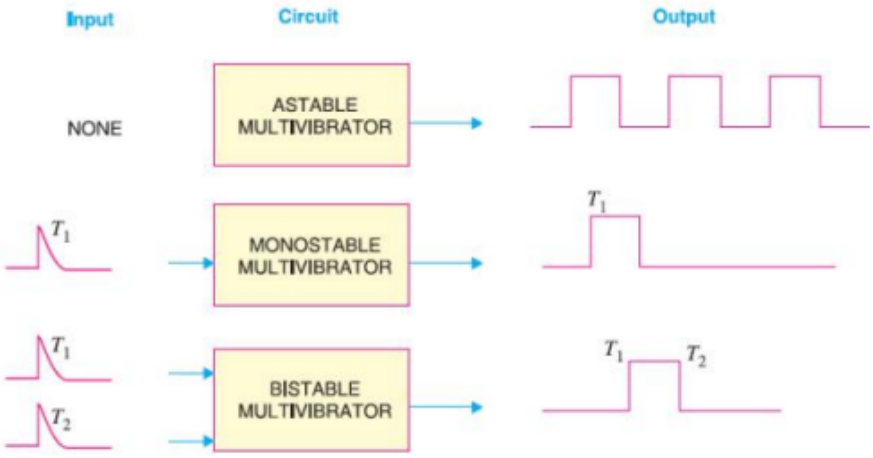


Figure 7.6: Shows the Input/Output Relation for the Three Types of Multivibrators

7.7 MONOSTABLE AND ASTABLE:

7.7.1 Astable Multivibrator:

The astable or free running multivibrator alternates automatically between the two states and remains in each for a time dependent upon the circuit constants. Thus it is just an oscillator since it requires no external pulse for its operation. Of course, it does require a source of d.c. power. Because it continuously produces the square wave output, it is often referred to as a free running multivibrator.

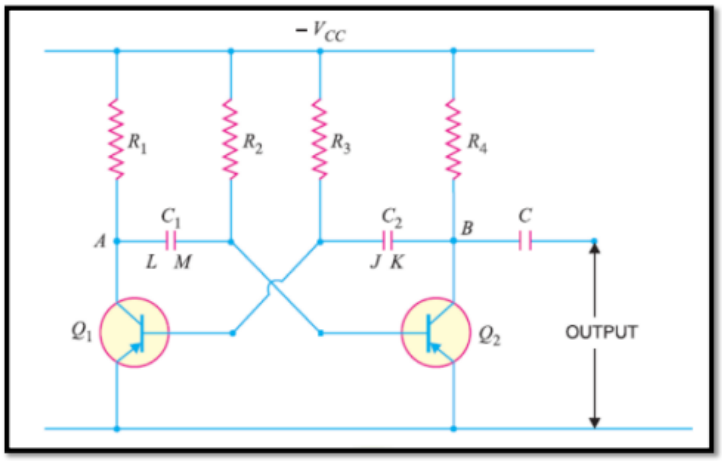


Figure 7.7: Astable Multivibrator

7.7.2 Monostable Multivibrator:

The monostable or one-shot multivibrator has one state stable and one quasistable (i.e. half-stable) state. The application of input pulse triggers the circuit into its quasistable state, in which it remains for a period determined by circuit constants. After this period of time, the circuit returns to its initial stable state, the process is repeated upon the application of each trigger pulse. Since the monostable multivibrator produces a single output pulse for each input trigger pulse, it is generally called oneshot multivibrator.

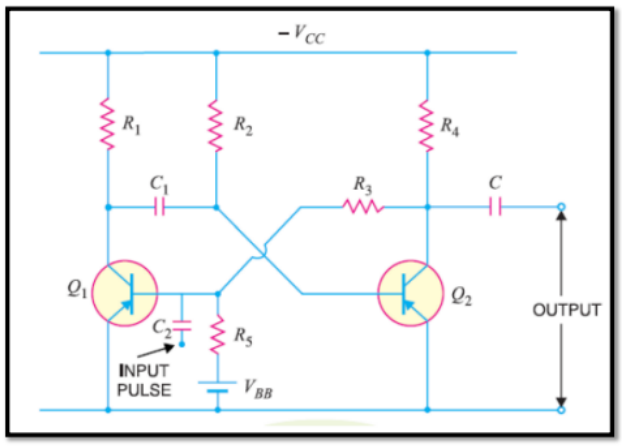


Figure 7.8 Monostable Multivibrator

7.8 COMPARATORS:

Comparators can give precision measurements, with consistent accuracy by eliminating human error. They are employed to find out by how much the dimensions of the given component differ from that of a known datum. If the indicated difference is small, a suitable magnification device is selected to obtain the desired accuracy of measurements. It is an indirect type of instrument and used for linear measurement. If the dimension measured, is less or greater than the standard, then the difference will be shown on the dial. It gives only the difference between actual and standard dimensions of the workpiece. Comparing the height of the job H2 with the standard job of height H1.

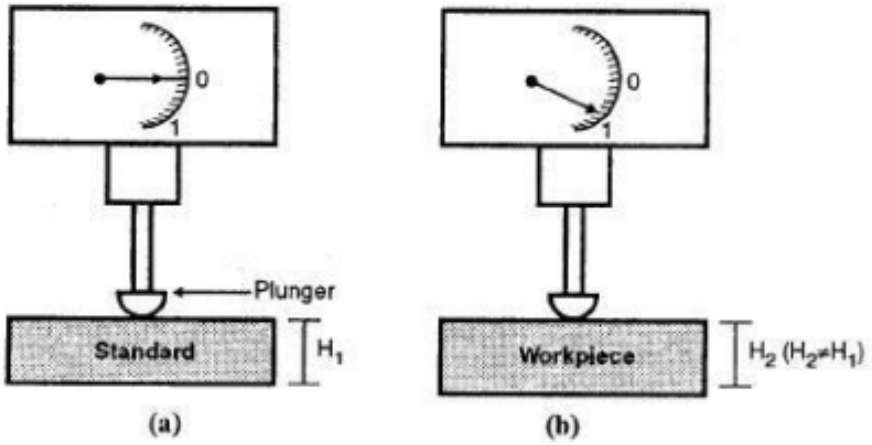


Figure 7.9 Comparator Diagram

Initially, the comparator is adjusted to zero on its dial with a standard job in position as shown in Figure (a). The reading H1 is taken with the help of a plunger. Then the standard job is replaced by the work piece to be checked and the reading H2 is taken. If H1and H2 is different, then the change i~ the dimension will be shown on the dial of the comparator. Thus difference is then magnified 1000 to 3000 X to get the clear variation in the standard and actual job. In short, Comparator is a device which

1) Picks up small variations in dimensions.
2) Magnifies it.
3) Displays it by using indicating devices, by which comparison can be made with some standard value.

7.8.1 Classification:

1) Mechanical Comparator: It works on gears, linkages, levers, springs etc.
2) Pneumatic Comparator: Pneumatic comparator works by using high pressure air, valves, back pressure etc.

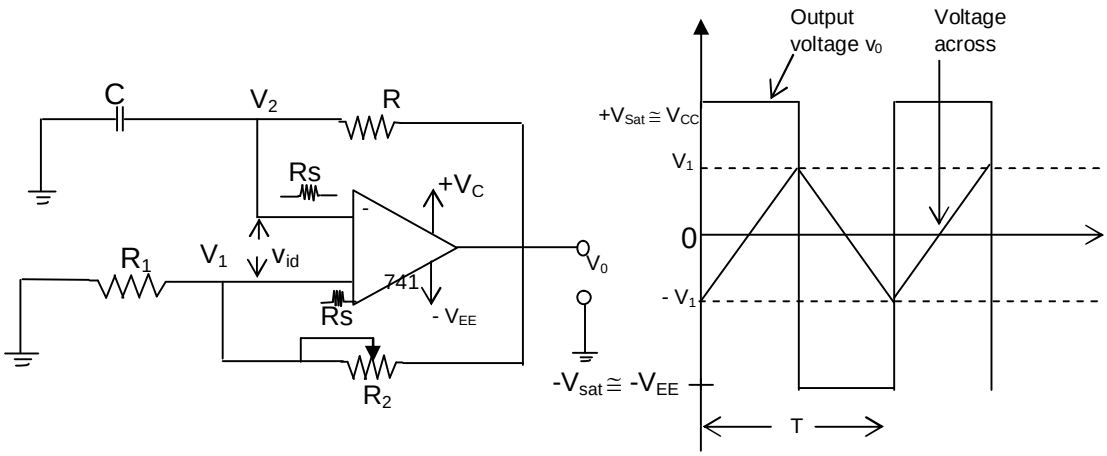
3) Optical Comparator: Optical comparator works by using lens, mirrors, light source etc.
4) Electrical Comparator: Works by using step-up, step-down transformers.
5) Electronic Comparator: It works by using an amplifier, digital signal etc.
6) Combined Comparator: The combination of any two of the above types can give the best result.

7.9 SQUARE WAVE AND TRIANGULAR WAVE GENERATORS:

7.9.1 Square Wave Generator:

In contrast to sine wave oscillator, square wave outputs are generated when the op-amp is forced to operate in the saturated regions, that is, the output of the op-amp is forced to swing repetitively between positive saturation + V_{sat} ($\cong$ + V_{CC}) and negative saturation - V_{sat} ($\cong$ - V_{EE}), resulting in the square wave output. Such a circuit showed in fig (7.10). This square wave generator is also called a free-running multi-vibrator.

Assume that the voltage across the capacitor C is zero volts at the instant the dc supply voltages +V_{CC} and –V_{EE} are applied. This means that the voltage at the inverting terminal is zero initially. At the same instant, however, the voltage v_1 at the non-inverting terminal is very small finite value, that is a function of the offset voltage V_{ooT} and the values of R_1 and R_2 resistors. Thus, the differential input voltage v_{id} is equal to the voltage v_1 at the non-inverting terminal. Although very small, voltage $v_1$will start to drive the op-amp into saturation.



$$|v_1| = \frac{R_1}{R_1 + R_2}\,|V_{sat}|;\ f_0 = \frac{1}{T} = \frac{1}{2RC} \text{ if } R_2 = 1.16R_1$$

Figure 7.10 Square Wave Generator (b) Wave Form of Output Voltage. V_0 and Capacitor Voltage v_2 of the Square Wave Generator

For example, suppose that the output offset voltage v_{001} is positive and that, therefore voltage v_1 is also positive. Since initially the capacitor C acts as a short circuit, the gain of the amplifier is very large (A); hence v_1 drives the output of the op-amp to its positive saturation $+ V_{sat}$ with the output voltage of the op-amp as $+V_{sat}$, the capacitor C_2 starts charging toward $+V_{sat}$ through the resistor, R. However, as soon as the voltage v_2 across the capacitor is slightly more positive than v_1, the output of op-amp is forced to switch to a negative saturation, $-V_{sat}$. With the op-amps output voltage as negative saturation, $-V_{sat}$ the voltage v_1 across R_1 is also negative, since $v_1 = \frac{R_1}{R_1 + R_2}(-V_{sat})$ ---------- (7.27)

Thus, the net differential voltage $v_{id} = v_1 - v_2$ is negative, which holds the output of the op-amp in negative saturation. The output remains in negative saturation until the capacitor C discharges and then recharges to a negative voltage slightly higher than the $-v_1$ [see Fig 7.10(b)]. Now, as soon as the capacitor voltage v_2 becomes more negative than $-v_1$, the net differential voltage v_{id} becomes positive and hence drives the output of the op-amp back to its positive saturation $+V_{sat}$. This completes one cycle. With output at $+V_{sar}$, and voltage v_1, at the non-inverting input is

$$v_1 = \frac{R_1}{R_1 + R}(+v_{sat}) \qquad \text{----------(7.28)}$$

The time period T of the output waveform is given by

$$T = 2RC \ln\left(\frac{2R_1 + R_2}{R_2}\right) \qquad \text{--------- (7.29(a))}$$

$$\text{or} \quad f_0 = \frac{1}{2RC \ln\left(\frac{2R_1 + R_2}{R_2}\right)} \qquad \text{---------- (7.29(b))}$$

Equation 7.29(b) indicates that the frequency of the output f_0 is not only a function of the RC time constant but also of the relationship between R_1 and R_2. For example If $R_2 = 1.16\ R_1$, equation 7.29(b) becomes

$$f_0 = \frac{1}{2RC} \qquad \text{---------- (7.30)}$$

Equation (7.30) shows that the smaller the RC time constant, the higher the output frequency f_0 and vice versa. As in the sine wave oscillator, the highest frequency square wave generated is also set by the slew rate of the op-amp. In practice, each inverting and non-

inverting terminal needs a series resistor R_s to prevent excessive differential current flow because the input of the op-amp is subjected to large differential voltage. R_s should have a value of 100 KΩ or higher.

7.9.2 Triangular Wave Generator:

Recall that the output waveform of the integrator is triangular if its input is a square wave. It means that a triangular wave generator can be formed by simply connecting an integrator to the square wave generator of figure 7.11. The resultant circuit is shown in figure 7.10(a). The circuit requires a dual op-amp, two capacitors and at least five resistors. The frequency of square waves and triangular wave is the same. For fixed R_1, R_2 and C values, the frequency of the square wave, as well as triangular wave, depends on the resistance R. As R is increased or decreased, the frequency of the triangular wave will decrease or increase, respectively. Although the amplitude of the square wave is constant; ($\pm V_{sat}$), the amplitude of the triangular wave decreases with an increase in its frequency, and vice versa.

The input of integrator A_2 is a square wave, while its output is a triangular wave, however, the output of A_2 to be triangular wave requires that $5R_3C_2 > T / 2$, where T is the time period of the square wave input. As a rule, R_3C_2 should be equal to T.

To obtain a stable triangular wave, it may also be necessary to shunt the capacitor C_2 with the resistor $R_4 = 10\ R_3$ and connect an offset voltage compensating network at the non-inverting terminal of A_2. As withany other oscillator, the frequency of the triangular wave generator is limited by slew rate of the op-amp. Therefore, high slew rate op-amps should be used for the generation of relatively higher frequencies.

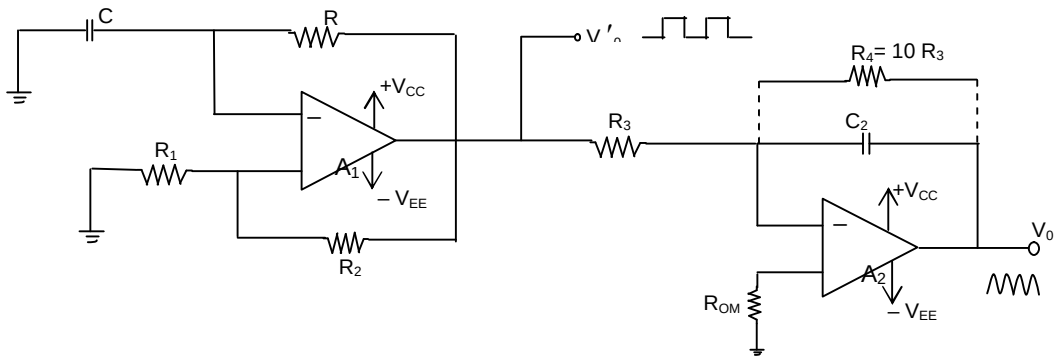


Figure 7.11(a) Triangular Wave Generator Circuit

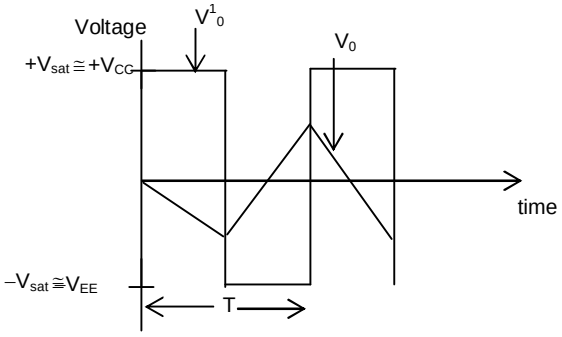


Figure 7.11 (b) Its Output Waveform

7.10 VOLTAGE REGULATORS:

A voltage regulator is used to regulate voltage levels. When a steady, reliable voltage is needed, then the voltage regulator is the preferred device shown in figure 7.12. It generates a fixed output voltage that remains constant for any changes in an input voltage or load conditions. It acts as a buffer for protecting components from damage. A voltage regulator is a device with a simple feed-forward design, and it uses negative feedback control loops.

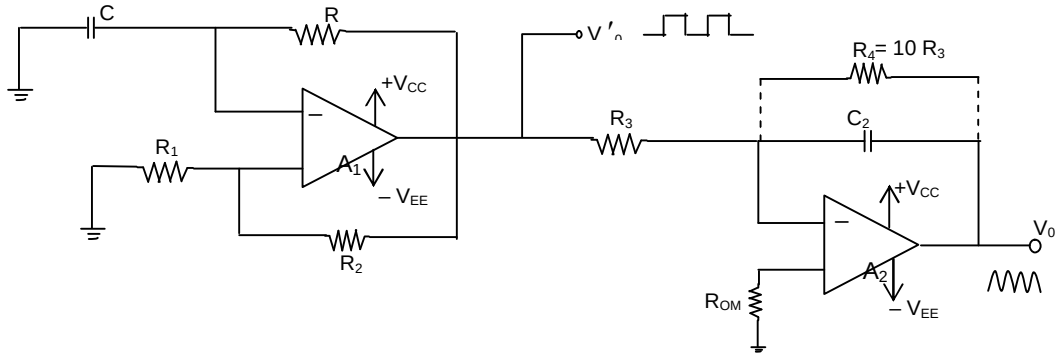


Figure 7.12 (a) Triangular Wave Generator Circuit

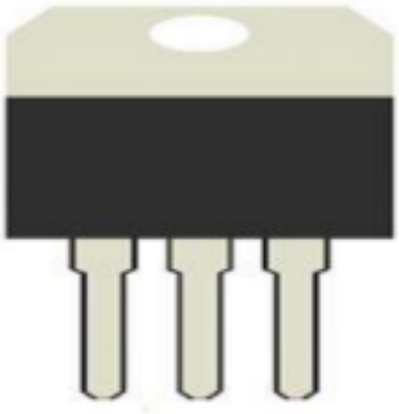


Figure 7.13 Voltage Regulator

There are mainly two types of voltage regulators: Linear voltage regulators and switching voltage regulators; these are used in wider applications. The linear voltage regulator is the easiest type of voltage regulator. It is available in two types, which are compact and used in low power, low voltage systems. Let us discuss different types of voltage regulators.

The main components used in the voltage regulator are

- Feedback Circuit
- Stable Reference Voltage
- Pass Element Control Circuit

The voltage regulation process is very easy by using the above three components. The first component of the voltage regulator, like a feedback circuit, is used to detect the changes within the DC voltage output. Based on the reference voltage as well as feedback, a control signal can be generated and drives the Pass Element to pay off the changes. Here, passes element is one kind of solid-state semiconductor device like a BJT transistor, PN-Junction Diode otherwise a MOSFET. Now, the DC output voltage can be maintained approximately stable.

7.10.1 Working of Voltage Regulator:

A voltage regulator circuit is used to make as well as maintain a permanent output voltage even when the input voltage otherwise load conditions are changed. The voltage regulator gets the voltage from a power supply, and it can be maintained in a range that is well-suited with the remaining electrical components. Most commonly these regulators are used for converting DC/DC power, AC/AC otherwise AC/DC.

7.10.2 Types of Voltage Regulators and their Working:

These regulators can be implemented through integrated circuits or discrete component circuits. Voltage regulators are classified into two type's namely linear voltage regulator & switching voltage regulator. These regulators are mainly used to regulate the voltage of a system; however, linear regulators work with low efficiency as well as switching regulators which work through high efficiency. In switching regulators with high efficiency, most of the i/p power can be transmitted to the o/p without dissipation shown in figure 7.14.

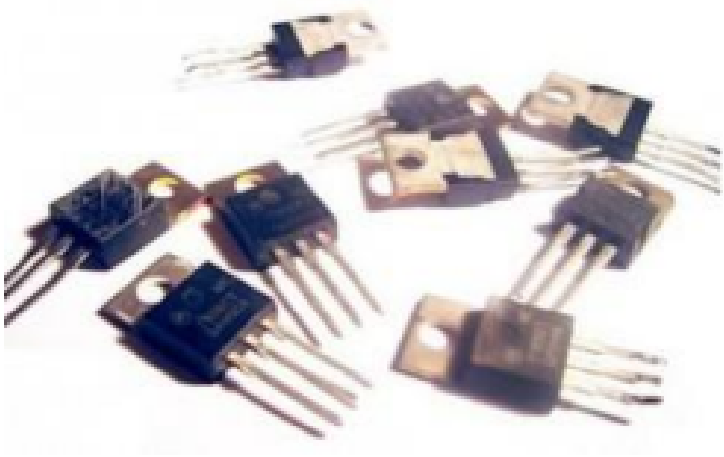


Figure 7.14 Types of Voltage Regulator

Basically, there are two types of Voltage regulators: Linear voltage regulator and switching voltage regulator.

- There are two types of Linear voltage regulators: Series and Shunt.
- There are three types of Switching voltage regulators: Step up, Step down, and Inverter voltage regulators.

7.11 SUMMARY:

This study explores oscillators, their principles, types, and applications in electronics. Oscillators are feedback amplifiers that generate continuous waveforms, such as sinusoidal, square, and triangular signals, based on circuit components and feedback mechanisms. The study examines various oscillator types, including RC, LC, crystal, phase shift, and Wien bridge oscillators, emphasizing their frequency stability and performance factors. It highlights the significance of phase shift, loop gain conditions, and feedback in sustaining oscillations. Additionally, it covers multivibrators astable, monostable, and bistable - used in waveform generation and pulse circuits. The study also delves into square and triangular wave generators and voltage regulators, which maintain stable output voltages. Key factors like temperature variations, power supply fluctuations, and circuit design influence oscillator efficiency. Understanding these concepts helps in designing stable and efficient circuits for communication, signal processing, and electronic measurement applications. This study provides a comprehensive insight into oscillator functionality in modern electronics.

7.12 TECHNICAL TERMS:

Phase Shift, Loop Gain, Harmonic Distortion, Voltage Regulation, Resonant Frequency, Multivibrator.

7.13 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are the key components required to design a phase shift oscillator?
2) How does feedback influence the stability and performance of an oscillator circuit?
3) What is the significance of the Barkhausen criterion in oscillator design?

Short Answer Questions:

1) How can harmonic distortion be minimized in practical oscillator circuits?
2) What are the differences between LC, RC, and crystal oscillators in terms of performance and applications?
3) How does temperature variation affect the frequency stability of an oscillator?

7.14 SUGGESTED READINGS:

1) Sedra, A.S., & Smith, K.C. (2018). *Microelectronic Circuits* (8^{th} ed.). Oxford University Press.

2) Millman, J., & Halkias, C.C. (2010). *Integrated Electronics: Analog and Digital Circuits and Systems* (2nd ed.). McGraw-Hill.

3) Boylestad, R.L., & Nashelsky, L. (2020). *Electronic Devices and Circuit Theory* (12th ed.). Pearson.

4) Malvino, A.P., & Bates, D.J. (2016). *Electronic Principles* (8th ed.). McGraw-Hill.

5) Schilling, D.L., & Belove, C. (2002). *Electronic Circuits: Discrete and Integrated* (3rd ed.). McGraw-Hill.

6) Razavi, B. (2001). *Design of Analog CMOS Integrated Circuits.* McGraw-Hill.

7) Floyd, T.L. (2017). *Electronic Devices* (10th ed.). Pearson.

Prof. R.V.S.S.N. Ravi Kumar

LESSON-8

MODULATION AND DEMODULATION

8.0 AIM AND OBJECTIVES:

The aim of this lesson is to provide a comprehensive understanding of modulation and demodulation, essential concepts in communication systems. Modulation is the process of varying a carrier signal to encode information, while demodulation retrieves the original message from the modulated wave. The objective is to explain amplitude modulation (AM), its principles, mathematical representation, and significance in signal transmission. It also covers different forms of AM, including Double Sideband Suppressed Carrier (DSB-SC) and Single Sideband Suppressed Carrier (SSB-SC), along with their advantages. Additionally, the lesson explores various methods for generating AM waves using nonlinear devices such as transistors and the significance of sidebands in transmission efficiency. The demodulation process, including the use of diode detectors for envelope detection, is also discussed. By the end, learners will grasp the importance of modulation in efficient signal transmission and its role in modern communication technologies.

STRUCTURE:

8.1 AMPLITUDE MODULATION:

Definition: When the amplitude of high frequency carrier wave is changed in accordance with the instantaneous value of the amplitude of the signal, it is called amplitude modulation. However, the frequency of the modulated wave remains the same i.e. at carrier frequency.

The following fig (1) shows the principle of amplitude modulation. Fig 8.1 shows the audio electrical signal whereas fig 8.2 shows a carrier wave.

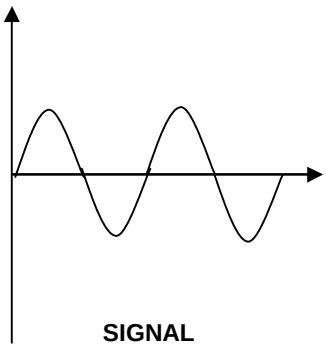


Figure 8.1 Shows the Amplitude Modulated (AM) Wave

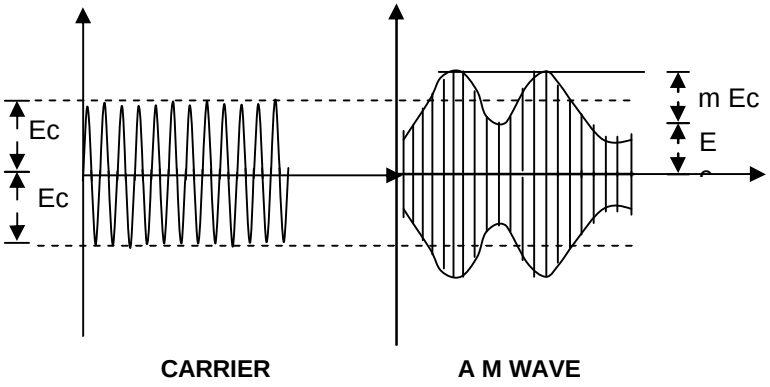


Figure 8.2 Amplitude Modulation

ote that the amplitudes of both positive and negative half cycles of carrier wave are changed in accordance with the signal. Amplitude modulation is done by, an electronic circuit called modulator.

8.1.1 Modulation Factor:

An important consideration in A.M. is to describe the depth of modulation, i.e. the extent to which the Amplitude of carrier wave is changed by the signal. This is described by a factor called modulation index, which may be defined as under.

8.1.1.1 Definition of Modulation index / factor:

The ratio of change of carrier wave amplitude due to modulation to the amplitude of un-modulated carrier wave is called the modulation factor m.

8.1.2 Mathematicalanalysis of Wave:

A carrier wave may be represented by $e_c = E_c \cos \omega_c t$, where

e_c = instantaneous voltage of carrier

E_c = amplitude of carrier

$\omega_c = 2\Pi f_c$ = angular velocity of carrier of frequency 'f_c'.

The modulating signal can be represented by

$e_s = Es \cos \omega_s t$

where is the instantaneous value of modulating signal. Es is its amplitude.

$\omega_s = 2\Pi f_s$ = angular frequency 'f_s'.

In amplitude modulation the amplitude E_c of the carrier wave is varied in accordance with the instantaneous value of the signal as shown in fig (8.1).

the amplitude of modulated signals can be represented by

$E = Ec + k_a E_m \cos\omega_s t$

Where k_a is called coefficient of modulation

The carrier wave amplitude is varied at signal frequency f_s. Therefore

$$e = E_c. (1+ m \cos \omega_s t) \cos \omega_c t. \qquad \rightarrow (8.1)$$

Where $m = k_a Es/Ec$ is called modulation index or modulation factor

The expression for A.M. wave can be expanded as

$= E_c . \cos \omega_c t + m . E_c . \cos \omega_s t . \cos \omega_c t.$

$$= E_c . \cos \omega_c t + m . \frac{E_c}{2} . \{2 \cos \omega_s t . \cos \omega_c t\} \qquad \rightarrow (8.2)$$

$$= E_c . \cos\omega_c t + m . \frac{E_c}{2} \{\cos (\omega_c+\omega_s)t + \cos (\omega_c-\omega_s)t\} \qquad \rightarrow (8.3)$$

$$= Ec . \cos \omega_c t + . \frac{mE_c}{2} . \cos (\omega_c+\omega_s).t + \frac{mE_c}{2} . \cos (\omega_c -\omega_s) t. \qquad \rightarrow (8.4)$$

The following points may be noted from the equation of Amplitude modulated wave.

The A.M. wave is equivalent to the summation of three sinusoidal waves: one having amplitude E_c and frequency 'f_c', the second having amplitude $\frac{mE_c}{2}$ and frequency $(f_c + f_s)$ and the third having amplitude $\frac{mE_c}{2}$ and frequency $(f_c - f_s)$.

The A.M. wave contains three frequencies. They are f_c, $f_c + f_s$ and $(f_c - f_s)$. Thus, the process of amplitude modulation does not change the original carrier frequency but produces two new frequencies $(f_c + f_s)$ and $(f_c - f_s)$, which are called side band frequencies. The sum of carrier frequency and signal frequency i.e., $(f_c + f_s)$ is called upper side band frequency. The lower side band frequency is $(f_c - f_s)$ i.e., the difference between carrier and signal frequencies. The amplitudes of the side bands are equal and proportional to depth of modulation. It can be shown that the maximum power in each side band occurs when m = 1 and is equal to one fourth of carrier power.

8.1.3 Upper and Lower Sidebands:

In the above discussion, it was assumed that the modulating signal was composed of one frequency component only. However, in a broadcasting station, the modulating signal is the human voice (or music), which contains waves with a frequency range of 20-4000 Hz. Each of these waves have its own LSF and USF. When combined, they give rise to an Upper-side band (USB) and a Lower-sideband (LSB) as Shown in figure 8.3.

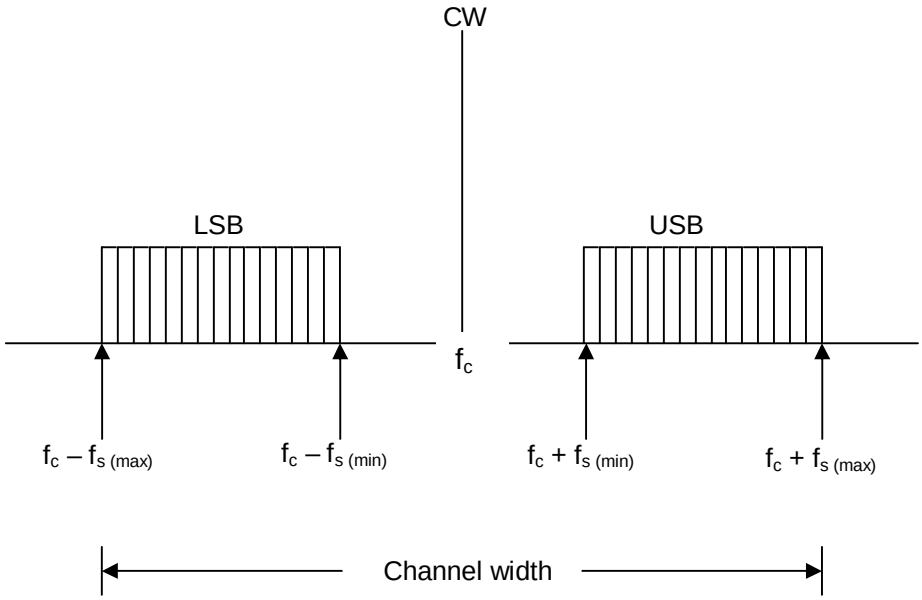


Figure 8.3 Lower Side Band and Upper Side Band

The USB, in fact, contains all sum components of the signal and carrier frequency whereas LSB contains their different components. The channel width (or bandwidth) is given by the difference between extreme frequencies i.e., between maximum frequency of USB and minimum frequency of LSB. As seen

Channel width = 2 x maximum frequency of modulating signal = 2 x f_s(max).

Example: An audio signal given by 15 sin 2Π (2000 t) amplitude-modulates a sinusoidal carrier wave 60 sin 2Π (100,000) t. Assuming ka=1 determines a) modulation index, (b)

percent modulation, (c) frequencies of signal and carrier, (d) frequency spectrum of the modulated wave.

Solution: Here carrier amplitude, A = 60 and modulating signal amplitude B = 15. Therefore

a) Modulation index, $m = \frac{B}{A} = \frac{15}{60} = 0.25$

b) b) Percent modulation, M = m x 100 = 0.25 x 100 = 25%

c) f_s = 2000 Hz --------by inspection of the given equation

f_c = 100,000 Hz --------by inspection of the given equation

d) The three frequencies present in the modulated carrier wave are

(i) 100,000 Hz = 100 kHz,(ii) 120,000 or 120 kHz (iii) 80,000 or 80kHz

Experimentally by measuring the maximum and minimum value of the modulated carrier amplitude one can determine the depth of modulation from the relation

$m = (E_{max} - E_{min}) / (E_{max} + E_{min})$

From this relation we see that, when $E_{max} = E_{min,}$ m = 0 or there is no modulation and when E_{min} = 0 there is 100 % modulation. In commercial radio broadcasting the depth of modulation is maintained around 40%.

8.2 GENERATION OF AM WAVES:

Amplitude modulation is produced by combining the carrier and the signal frequencies using a non-linear device. Diodes are non-linear devices, but they are not used as they do not offer any gain. Transistors behave as non-linear elements and offer gain as such they are suitable for these applications. Fig 8.4 shows a simple amplitude – modulated amplifier.

The supply V_{CC} in combination with the resistors R_1, R_2, R_C & R_E sets the quiescent point for the transistor. The carrier e_C is the input to the CE amplifier. The circuit amplifies the carrier by a factor A_V where A_V is the voltage gain, so that the amplifier output is $A_V e_C$. The modulating signal e_m is applied in emitter circuit and hence forms a part of the biasing. It produces variations in emitter current at modulating frequency, which in turn produces variations in base emitter resistance and gain A_V. Thus, the amplitude of the carrier varies in accordance with the strength of the signal there by producing Amplitude modulated output across 'R_L'.

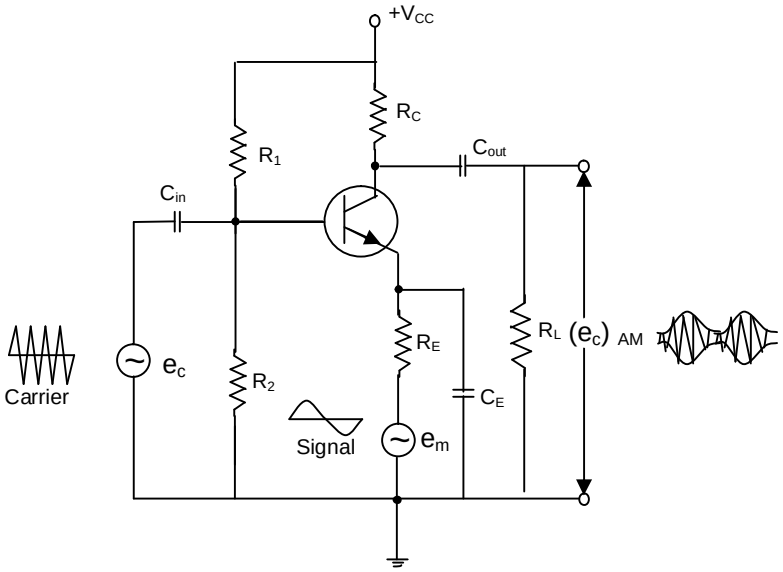


Figure 8.4 An Amplitude-Modulated Amplifier

8.2.1 Forms of Amplitude Modulation:

We know now that one carrier and two side bands are produced in A.M. generation. It is not necessary to transmit all these signals to enable the receiver to reconstruct the original signal. Accordingly, we may attenuate or altogether remove the carrier or any one of the side bands without affecting the communication process. The advantages are,

1) Less transmitted power and
2) Small bandwidth required.

The different suppressed component systems are: (a) DSB-SC (b) SSB-TC (C) SSB-SC.

a) DSB-SC: It stands for double side band suppressed carrier system. Here carrier components are suppressed thereby saving enormous amounts of power.

b) SSB-TC: It stands for single side band transmitted carrier system. In this case one sideband is suppressed but the other sideband and carrier are transmitted.

c) SSB-SC: It stands for single side band suppressed carrier system. It suppresses one side band and the carrier and transmits only the remaining side bond.

Out of the above three systems, the SSB-SC (or simply SSB) is preferable because of its advantage over other.

8.2.2 Generation of SSB-SC or SSB Systems:

The SSB-SC or simply SSB is generated mainly through the following three methods:

1) Filter method.
2) Phase cancellation (shift) method.
3) The third method.

8.2.2.1. Filter Method: In this method, DSB-SC signal is generated by, the balanced modulator. This signal is allowed to pass through side band filters, which are narrow band pass filters that only allow to pass the desired side band of frequencies. The filter may be a LC, mechanical or a crystal filter. All these filters have the disadvantage that their operating frequency is below the usual transmitting frequencies. Thus, a balanced mixer and crystal oscillator are used to provide up conversion to the final transmitter frequency.

8.2.2.2. Phase Shift Method: In this method, two balanced modulators and two-phase shifting networks are used. One of the balanced modulators receives the carrier voltage shifted by 90^0 and the modulating voltage signal while the other balanced modulator receives the carrier voltage and the modulating voltage shifted by 90^0. The output of both modulators consists of sidebands only. The signal is applied directly to modulatorM_1; therefore the modulator puts out two side bands, each one is shifted in phase by 90^0. The signal is shifted 90^0 before it is applied to the modulator, therefore, the modulator also puts two side bands, but in this case the upper and lower side bands and shifted by $+90^0$ and -90^0 respectively. The two lower side bands, which are out of phase, when combined in the adder cancel each other. The upper side hands, which are in phase add directly into the adder. Thus, SSB in which lower sideband is removed is produced. The block diagram of this method is shown in the fig.

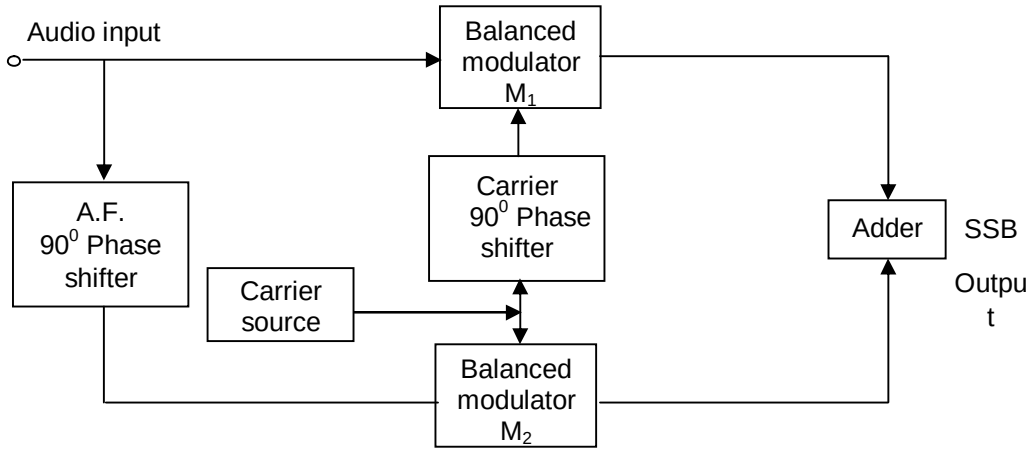


Figure 8.5: SSB-SC Transmitter Using Phase Shift.

The output of first balanced modulator M_1

$$v_1 = 2\, v_c .\, m_a \cos \omega_c t \,.\, \sin \omega_m .t. \qquad \rightarrow (8.5)$$

The output from the balanced modulator M_2

$$v_2 = 2\, v_c\, m_a \text{ co sin } \omega_c t \,.\, \cos \omega_m .\, t \qquad \rightarrow (8.6)$$

The summing amplifier, the resultant output

$$v = v_1 + v_2 = 2\, v_c \,.\, m_a \sin (\omega_c + \omega_{m)\, t} \qquad \rightarrow (8.7)$$

(By adding (1) & (2))

Thus, the output contains only an upper sideband, similarly, the output with lower side band can be obtained by passing the signal carrier directly to balanced modulator M_1 and each through 90^0 phase shifting networks to the modulator M_2.

8.2.2.3 The Third Method:

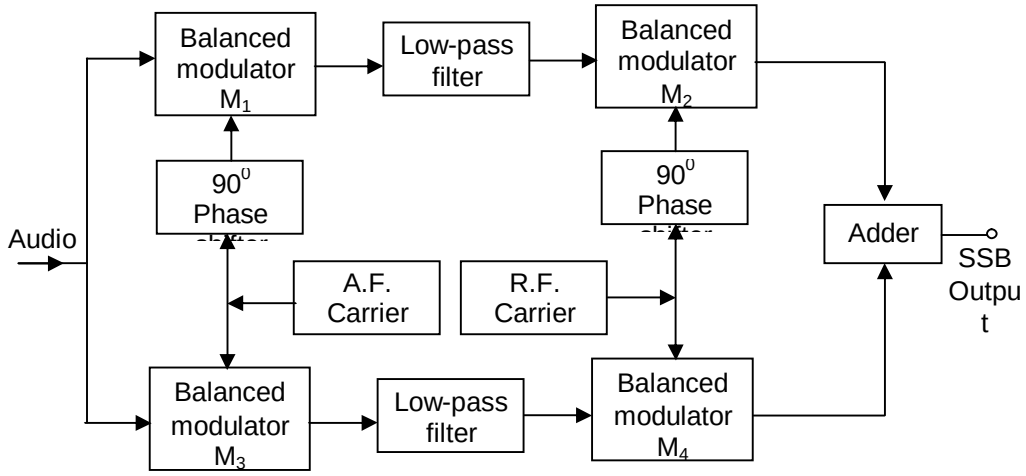


Figure 8.6 Third Method for SSB Generation.

This method can generate SSB at any frequency and low audio frequencies. It neither refuses a filter circuit nor a wide band, audio phase shift network. The circuit is identical to that of the phase shift method, but the way in which voltages are fed to the two balanced modulators has been changed. Due to this reason, this method is called modified phase shift method. The block diagram of the method is shown in the following fig.8.6.

8.3 DEMODULATION OF AM WAVES:

8.3.1 Definition of Demodulation:

The process of recovering the audio signal from the modulated wave is known as demodulation of detection.

At the broadcasting station, modulation is done to transmit the audio signal over larger distances to a receiver. When the modulated wave is picked up by the radio receiver, it is necessary to recover the audio signal from it. This process is accomplished in the radio receiver and is called demodulation.

8.3.2 Necessity of Demodulation:

We know that an A.M. wave consists of carrier and side band frequencies. The audio signal is inside band frequencies, which are radio frequencies. If the modulated wave after amplification is directly fed to the speaker, no sound will be heard because of the inertia of the diaphragm as it is not able to respond to such high frequencies.

8.3.3 Essentials of Demodulation:

For a modulated wave to be audible, it is necessary to change the nature of modulated waves. This is done by a circuit called detector. A detector circuit performs the following two functions

It rectifies the modulated wave.

It separates the audio signal from the carrier.

The demodulation can be accomplished with the help of a diode detector.

8.3.4 A.M. Diode detector:

Diode detection is also known as envelope detection or linear detection.

In appearance it looks like an ordinary half wave rectifier circuit with capacitor input. As shown in the fig below--

It is called envelope detection because it recovers the A.F. signal envelope from the composite signal. Similarly, diode detectors are called linear detector because its output is proportional to the voltage of the input signal.

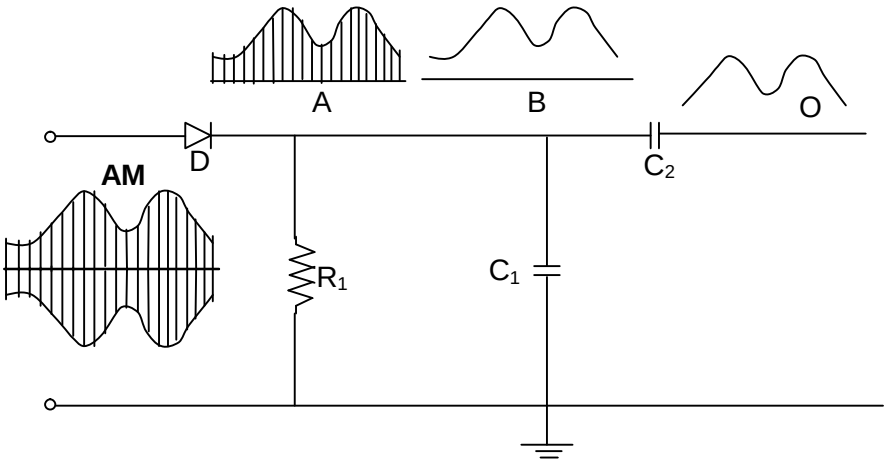


Figure 8.7: Simple Detector AM-Amplitude Modulated Wave

8.3.5 Circuit Action:

This circuit involves the rectification of the signal, and filtering the r.f. and dc components. The signal is rectified by diode 'D'. The rectified wave is shown at A in the fig. The combination $C_1 – R_1$ removes the high frequency component and provide the signal super posed over a dccomponent. The dc component is eliminated by capacitor 'C_2' to give the output shown in the figure 8.7.

8.4 DSBSC MODULATION:

Double-sideband suppressed-carrier transmission (DSB-SC) is transmission in which frequencies produced by amplitude modulation (AM) are symmetrically spaced above and below the carrier frequency and the carrier level is reduced to the lowest practical level, ideally being completely suppressed.

8.4.1 DSB-SC Spectrum:

DSB-SC is basically an amplitude modulation wave without the carrier, therefore reducing power waste, giving it a 50% efficiency. This is an increase compared to normal AM transmission (DSB), which has a maximum efficiency of 33.333%, since 2/3 of the power is in the carrier which carries no intelligence, and each sideband carries the same information. Single Side Band (SSB) Suppressed Carrier is 100% efficient is shown in Figure 8.8.

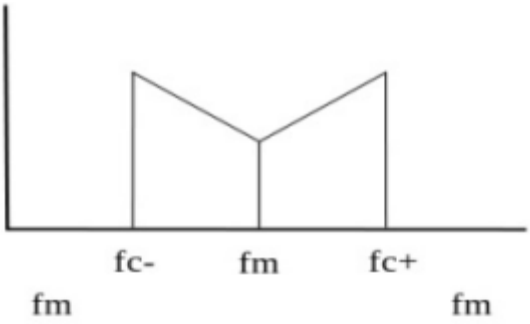


Figure 8.8 Spectrum Plot of an DSB-SC Signal

8.5 GENERATION OF DSBSC MODULATION:

DSB-SC is generated by a mixer. This consists of a message signal multiplied by a carrier signal. The mathematical representation of this process is shown below, where the product-to-sum trigonometric identity is used. Figure shows the 8.9 Generation of DSB Signals.

$$\underbrace{V_m \cos(\omega_m t)}_{\text{Message}} \times \underbrace{V_c \cos(\omega_c t)}_{\text{Carrier}} = \underbrace{\frac{V_m V_c}{2}\left[\cos((\omega_m + \omega_c)t) + \cos((\omega_m - \omega_c)t)\right]}_{\text{Modulated Signal}}$$

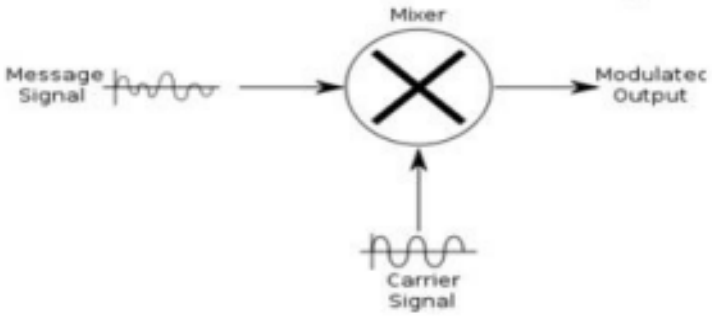


Figure 8.9 Generation of DSB Signals

8.5.1 Multiplier Modulator:

Here the modulation is directly achieved by multiplying m(t) with A f t c cos2π c, using an analog multiplier whose output is proportional to the product of two input signals. Typically, such a multiplier may be obtained from a variable gain amplifier in which the gain parameter (such as βof a transistor) is controlled by one of he signals, say m(t). When the signal A f t c cos2π c is applied at the input of this amplifier, the output is proportional to A_cm(t) f t cos2π c .

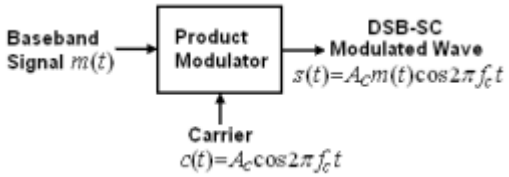


Figure 8.10 Block Diagram of Multiplier Modulator

8.5.2 Balanced Modulator:

The block diagram of balanced modulator is shown in Fig 8.11. A balanced modulator consists of two standard AM modulators arranged in a balanced configuration to suppress the carrier wave. We assume that the two modulators are identical except for the signal reversal of the modulating wave applied to the input of one of them.

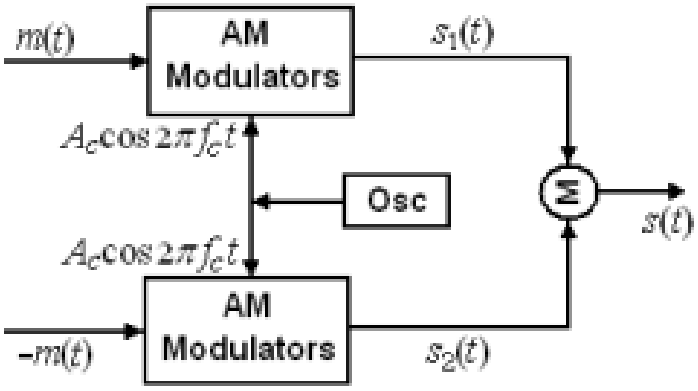


Figure 8.11 Block Diagram of Balanced Modulator

Thus the output of the two modulators are expressed as follows:

$S_1(t) = A_c[1+k_a m(t)] \cos 2\pi f_c t$ and $S_2(t) = A_c[1-k_a m(t)] \cos 2\pi f_c t$ → (8.8)

By subtracting $S_2(t)$ from $S_1(t)$, we obtain $s(t) = S_1(t) - S_2(t) = [2k_a m(t)\ A_c \cos 2\pi f_c t]m(t)$

Hence except for the scaling factor $2k_a$ the balanced modulator output is equal to the product of the modulating wave and carrier as required.

8.5.3 Switching Modulator:

The multiplication operation required for modulation can be replaced by a simple switching operation, if we realize that a modulated signal can be obtained by multiplying m(t) not only by a pure sinusoid by any periodic signal of the fundamental radian frequency fc. Fig 8.12(a) shows one such electronic switch driven by a sinusoid A f t c cos2π c. D1, D2, D3 and D4 are matched pairs. When the signal A f t c cos2π c is of a polarity that will make terminal 'c' positive with respect to 'd', all the diodes conduct. Because diodes D1 and D2 are matched, terminals 'a'and 'b' have the same potential and are effectively short circuited. During the next half cycle, terminal 'd' is positive with respect to 'c', all four diodes is reversed biased and terminals 'a'and 'b' are open circuited. Fig 8.12(b) shows series diode-bridge modulators and Fig 8.12(c) shows parallel diode-bridge modulators driven by this switch. The diode bridge in Fig 8.12(a) therefore serves as a desired electronic switch. Where terminals 'a' and 'b' open and close periodically with carrier frequency fc when sinusoid cos2π f tc is applied across terminals 'c' and 'd'. The switching operation can be represented by a square wave w(t). as shown in Figure. To obtain the signal m(t)w(t), we may place this electronic switch (terminals 'a' and 'b') in series (Fig 8.12(b) or across (in parallel) m(t) as shown in Fig 8.12(c). These modulators are known as Series-bridge diode modulator and the shunt-bridge diode modulator respectively. This switching on and off of m(t) repeats for each cycle of the carrier, resulting in the switched signal m(t)w(t), which when bandpass filtered, yields the desired modulated signal (2/π) m(t) cos2π f tc. The spectral analysis of this modulation is illustrated in Fig 8.12.

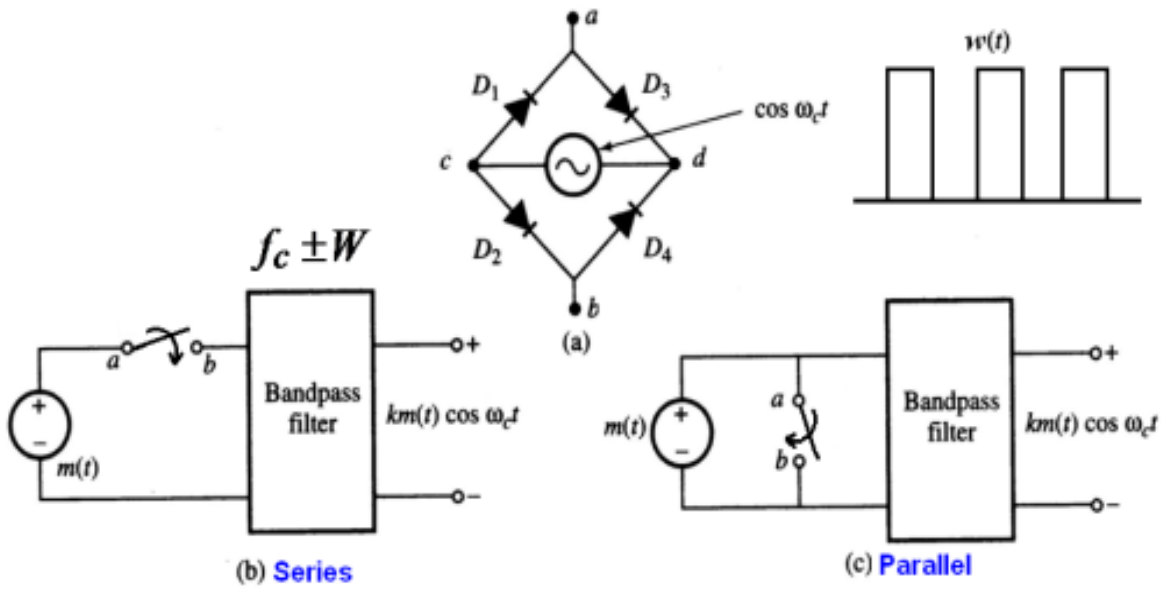


Figure 8.12 (a) Diode Bridge Electronic Switch (b) Series Bridge Modulator (c) Shunt Bridge Modulator

8.6 COHERENT DETECTION OF DSBSC WAVES:

Coherent / Synchronous Detection: Here the signal recovery is done by a reverse frequency translation by multiplying the translated signal with cos2π f t c. It is assumed that the local oscillator output is coherent or synchronized in both frequency and phase with the carrier wave used to generate modulated signal (transmitted) s(t). This method of demodulation is known as coherent detection or synchronous detection. The block diagram of synchronous detection is shown in Figure 8.13.

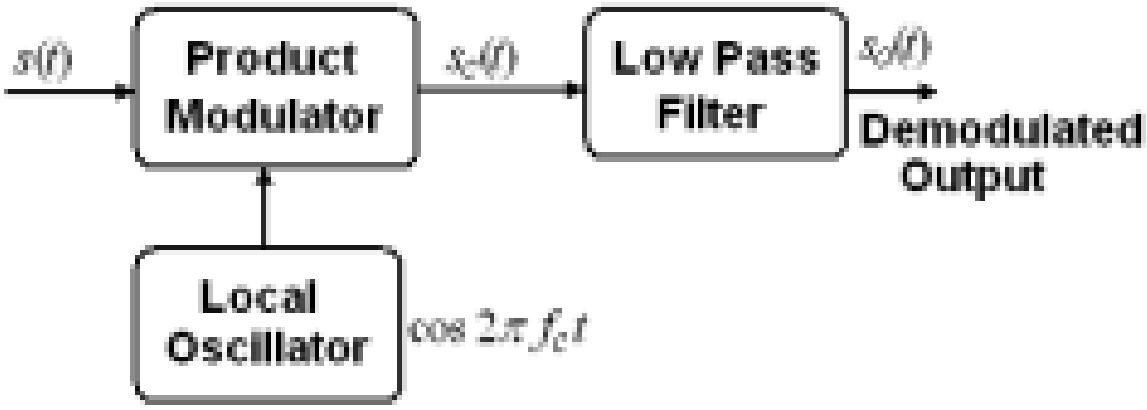


Figure 8.13 Block Diagram of Coherent Detection

Let DSB-SC signal is $s(t) = A_c m(t) \cos 2\pi f_c t$, Then the product modulator output $S_c(t)$is given by

$$S_c(t) = S(t) \cos 2\pi f_c t = [A_c m(t) \cos 2\pi f_c t] \cos 2\pi f_c t \quad \rightarrow (8.9)$$

$$= A_c m(t) \cos^2 2\pi f_c t = A_c m(t) \tfrac{1}{2} [1+\cos 4\pi f_c t]$$

$$= \tfrac{1}{2} A_c m(t) + \tfrac{1}{2} \cos 4\pi f_c t \quad \rightarrow (8.10)$$

This product modulator output s t c () consists of required message signal with magnitude Ac/2 and unwanted component having twice the frequency of carrier wave. This unwanted component is removed by a low pass filter with cut off frequency of message signal. To demodulate DSB-SC signal by synchronous detection, one must generate a local carrier of same frequency and phase angle at the receiver side. Any discrepancy in the frequency and phase of local carrier gives rise to a distortion in the detector output. It is thus instructive to examine the detected output when the local carrier frequency and phase are different from the carrier of the incoming DSBSC signal. We consider the following two situations.

1) The local oscillator has an ideal frequency, but arbitrary phase difference measured with respect to the carrier is referred to as 'Phase Error'.

2) The local oscillator has an identical phase but a difference frequency with respect to carrier is referred to as 'Frequency error'.

(a) Phase Error ($\phi \neq 0$; $\Delta f = 0$); Let the carrier is $\cos(2\pi f_c t + \phi)$, where ϕ being the phase difference between the local oscillator signal and the carrier at the transmitter, we get

$$s_c(t) = s(t)\cos(2\pi f_c t + \phi) = [A_c m(t)\cos(2\pi f_c t]\cos(2\pi f_c t + \phi) \quad \rightarrow (8.11)$$

$$= A_c \tfrac{1}{2} m(t)[\cos\phi + \cos(4\pi f_c t + \phi)]$$

$$= \tfrac{1}{2} A_c m(t)\cos\phi + \tfrac{1}{2} m(t)\cos(4\pi f_c t + \phi) \quad \rightarrow (8.12)$$

The first term represents the message term with phase error $\cos\phi$. The second term represents a DSB-SC wave with a carrier $2f_c$, which can be removed by low pass filter. The cut off frequency of this filter is greater than the message signal frequency 'W', but less than $2f_c$-W, thus the output of the filter is

$$s_o(t) = \tfrac{1}{2} A_c m(t)\cos\phi \quad \rightarrow (8.13)$$

Thus the demodulator output is proportional to mt(t) when the phase error $\cos\phi$ is constant. The output of the demodulated signal is maximum when $\phi = 0$ and minimum (zero) when $2\pi\phi = \pm$. The zero demodulated signal which occurs for $2\pi\phi = \pm$ represents quadrature null effect of the coherent detection. It is thus seen that the detector output is attenuated by a factor $\cos\phi$, So long as the phase error is constant, it provides an undistorted version of original message signal mt (Δ).Unfortunately, due to random variation in the communication channel, the multiplication factor $\cos\phi$ varies randomly with time. This results in distortion of the signal.

(b) Frequency Error ($\phi = 0$; $\Delta f \neq 0$): Suppose that the local oscillator signal $\cos 2\pi(f_c + \Delta f)t$ has phase is zero and Δf is frequency error. Then the output of the product modulator is

$$s_c(t) = s(t)\cos\{2\pi(f_c t + \Delta f)t\} = [A_c m(t)\cos(2\pi f_c t]\cos\{2\pi(f_c t + \Delta f)t\} \quad \rightarrow (8.14)$$

$$= A_c \tfrac{1}{2} m(t)[\cos 2\pi\Delta f t + \cos\{2\pi(f_c t + \Delta f)t\}]$$

$$= \tfrac{1}{2} A_c m(t)\cos 2\pi\Delta f t + \tfrac{1}{2} m(t)\cos\{2\pi(f_c t + \Delta f)t\} \quad \rightarrow (8.15)$$

Thus when this signal is passed through the LPF, we get $s_o(t)= ½ Ac\ m(t) \cos 2\pi \Delta f t$. The resulting signal will be unacceptable if Δf is comparable to the baseband signal frequency.

The spectrum of the output of LPF is $S_o(f) = ½ A_c[M(f-\Delta f) + M(f+\Delta f)]$. → (8.16)

It can be seen that each frequency component of the baseband signal is affected by an amount of Δf. Therefore the frequencies which were originally harmonically related in the baseband signal will no longer remain so after demodulation. This results in undersired distortion in the recovered signal.

8.7 SUMMARY:

Modulation and demodulation are fundamental processes in communication systems that enable efficient signal transmission. Modulation involves altering a carrier signal's amplitude, frequency, or phase to encode information, while demodulation extracts the original message. Amplitude Modulation (AM) is widely used in radio broadcasting and wireless communication, with key types including Double Sideband Suppressed Carrier (DSB-SC) and Single Sideband Suppressed Carrier (SSB-SC), each offering distinct advantages in bandwidth efficiency. The generation of AM waves utilizes nonlinear devices like transistors, ensuring reliable signal transmission. Sidebands play a crucial role in determining transmission efficiency, with SSB-SC being particularly effective in reducing power consumption and bandwidth usage. Demodulation techniques, such as envelope detection using diode detectors, help recover transmitted signals accurately. Understanding these processes is essential for optimizing communication systems, improving signal clarity, and ensuring effective data transmission across various modern applications, including broadcasting, telecommunication, and satellite communication.

8.8 TECHNICAL TERMS:

Carrier Signal, Sidebands, Envelope Detector, Bandwidth Efficiency, Nonlinear Device, Suppressed Carrier Modulation

8.9 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What is the role of a carrier signal in amplitude modulation?
2) How do sidebands contribute to the transmission of information in AM?
3) What is the function of an envelope detector in demodulation?

Short Answer Questions:

1) How does bandwidth efficiency affect communication system performance?
2) Why are nonlinear devices used in AM signal generation?
3) What are the advantages of suppressed carrier modulation techniques?

8.10 SUGGESTED READINGS:

1) Proakis, J.G., & Salehi, M. (2007). Digital Communications (5^{th} ed.). McGraw-Hill.
2) Haykin, S. (2010). Communication Systems (5th ed.). Wiley.
3) Lathi, B.P., & Ding, Z. (2018). Modern Digital and Analog Communication Systems (5^{th} ed.). Oxford University Press.
4) Taub, H., Schilling, D.L., & Saha, G. (2013). Principles of Communication Systems (3^{rd} ed.). McGraw-Hill.
5) Couch, L.W. (2013). Digital and Analog Communication Systems (8^{th} ed.). Pearson.

Prof. R.V.S.S.N. Ravi Kumar

LESSON-9

SSB MODULATION, VSB MODULATION AND FDM

9.0 AIM AND OBJECTIVES:

The aim of this study is to explore the principles, generation, and detection of Single Sideband (SSB) modulation, Vestigial Sideband (VSB) modulation, and Frequency Division Multiplexing (FDM), highlighting their significance in communication systems. The objective is to understand the efficiency of SSB in reducing bandwidth usage by transmitting only one sideband, analyze the application of VSB in television broadcasting where low-frequency components are crucial, and examine how FDM enables multiple signals to share a single communication channel without interference. The study further aims to investigate the role of filtering techniques in SSB generation, the importance of coherent demodulation in signal recovery, and the necessity of guard bands in FDM to prevent inter-channel crosstalk. By comprehending these modulation techniques, we can enhance bandwidth utilization, improve signal clarity, and optimize transmission efficiency in various communication applications, including radio, television, and telephony.

STRUCTURE:

9.1 SSB MODULATION:

As discussed earlier conventional AM and DSB-SC is wasteful from the band width point of view and requires a bandwidth equal to twice the message signal bandwidth. In transmitted signal one half is occupied by USB, while the other half is occupied by LSB. However, the USB and LSB are unequally related to each other by virtue of their symmetry about the carrier. Moreover, the message signal information is available in both sidebands. As far as the

transmission is concerned only one sideband is necessary to reproduce the baseband signal unequally at the receiver end. Thus, in a conventional AM, if the carrier and one of the sidebands is suppressed, no information is lost. The advantage of such suppression is that the transmission bandwidth required for this case is equal to the message signal bandwidth only. A modulation scheme in which one sideband is transmitted is known as SSB-SC or simply SSB modulation.

9.1.1 Baseband Signal SSB Modulation:

9.1.1.1 Frequency Domain Description:

Let the message signal m(t) is band-limited to W Hz and its Hilbert Transform is $m_h(t)$.

$M_+(f)=M(f)u(f)=M(f)1/2\ [1+sgn(f)]=1/2\ [M(f)+jM_h(f)]$ → (9.1)

and

$M\text{-}(f)=M(f)u(-f)=M(f)1/2\ [1-sgn(f)]=1/2\ [M(f)-jM_h(f)]$ → (9.2)

where $M_h(f)$ is the Fourier Transform of $m_h(t)$ (which is the Hilbert transform of m(t)), u(f) and u(−f) are unit step functions in the directions of f and −f. Now we can express the SSB signal in terms of m(t) and $m_h(t)$.

$\Phi_{USB}(f)=M_+(f-fc)+M_+(f+fc)$ → (9.3)

$=1/2\ [M(f-fc)+M(f+fc)]-\ 1/2j[M_h(f-fc)-M_h(f+fc)]$ → (9.4)

Similarly,

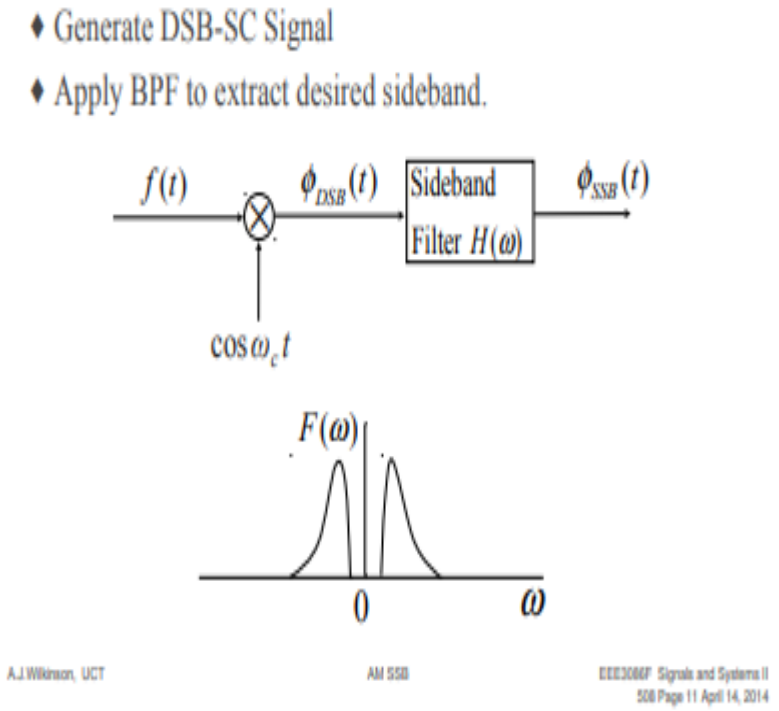
$\Phi_{USB}(f)=M_+(f+fc)+M_-(f+fc)$

$=1/2\ [M(f-fc)+M(f+fc)]+\ 1/2j\ [M_h(f-fc)-M_h(f+fc)]$ → (9.5)

9.2 GENERATION AND DETECTION OF SSB WAVES:

9.2.1 SSB Generation (Modulators):

From the foregoing discussion it is seen that SSB-SC wave can be described in frequency domain as well as time-domain for arbitrary baseband signal. A closer examination of the two descriptions reveals that an SSB can be generated by taking help of either representation. Thus, SSB waves can be generated by frequency discrimination method and by the phase discrimination method based on frequency domain and time domain descriptions of SSB respectively.



SSB Generation Via Filtering

Fig. 9.1 SSB Generation via Filtering

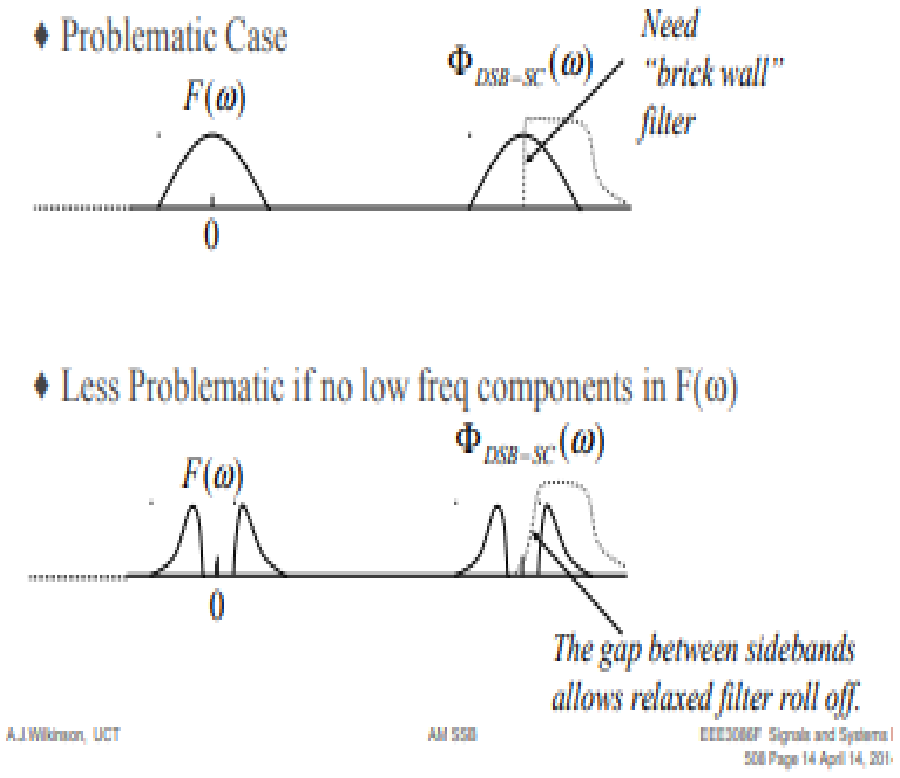


Figure 9.2 SSB Generation: Filter Rolls Off Problem

9.2.2 Demodulation or Detection of SSB Signals: Like DSB-SC modulation signal can be demodulated by synchronous detection. The incoming signal is multiplied with locally generated sinusoidal signal and then filtered by low pass filter. The filter is chosen to have the same bandwidth as the message signal bandwidth 'W' or somewhat larger. It is necessary that local oscillators are synchronized with the carrier in phase and frequency.

9.2.3 Coherent Demodulation:

Observe that $^{s}SSB(t)\cos\omega_c t = m(t)\cos^2(\omega_c t) \mp m_h(t)\sin(\omega_c t)\cos(\omega_c t)s \rightarrow (9.6)$

$$=1/2\ m(t) + 1/2\ m(t)\cos 2\omega_c t \pm \tfrac{1}{2}\ m_h(t)\sin 2\omega_c t \rightarrow (9.7)$$

If we filter $\phi_{SSB}\cos\omega_c t$ with a **LPF** (Low-Pass Filter), we can eliminate the components centered at $2f_c$ and the filter output will be approximately m(t).

Hence, any of the **coherent demodulation techniques** applicable for **DSB-SC signals** can be used.

9.3 VESTIGIAL SIDE BAND MODULATION:

Single sideband modulation is well suited for the transmission of speech because of the energy gap that exists in the spectrum of speech signals between zero and a few hundred hertz. When the message signal contains significant components at extremely low frequencies(as in the case of television signals), the upper and lower sidebands meet at the carrier frequency.

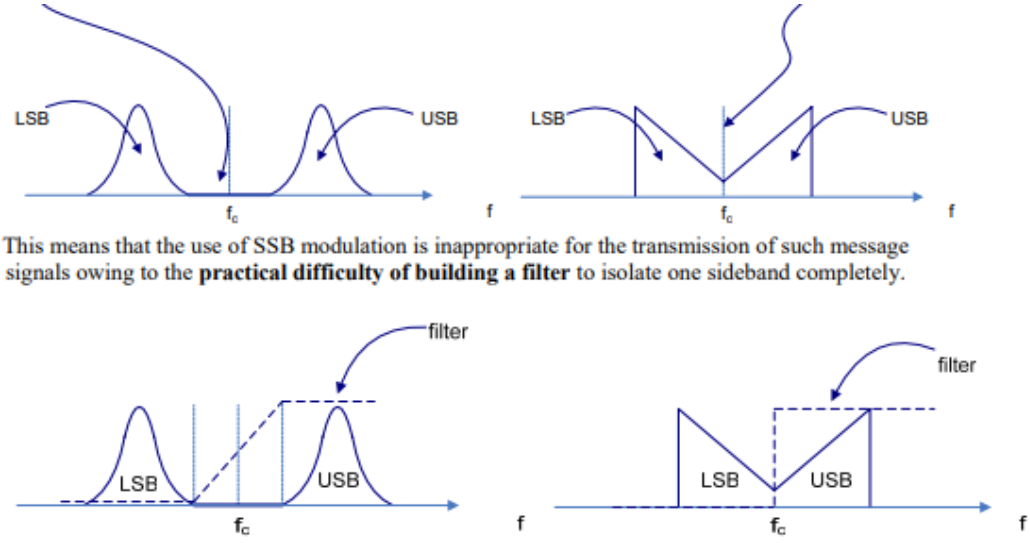


Fig 9.3 Vestigial Side Band Modulation

This difficulty suggests another scheme known as vestigial sideband modulation (VSB), which is a compromise between SSB and DSB-SC forms of modulation. In VSB modulation, one sideband is passed almost completely whereas just a trace or vestige, of the other sideband is retained. Figure 17 illustrates the spectrum of a VSB modulated wave s(t) in relation to that of the message signal m(t) assuming that the lower sideband is modified into the vestigial sideband.

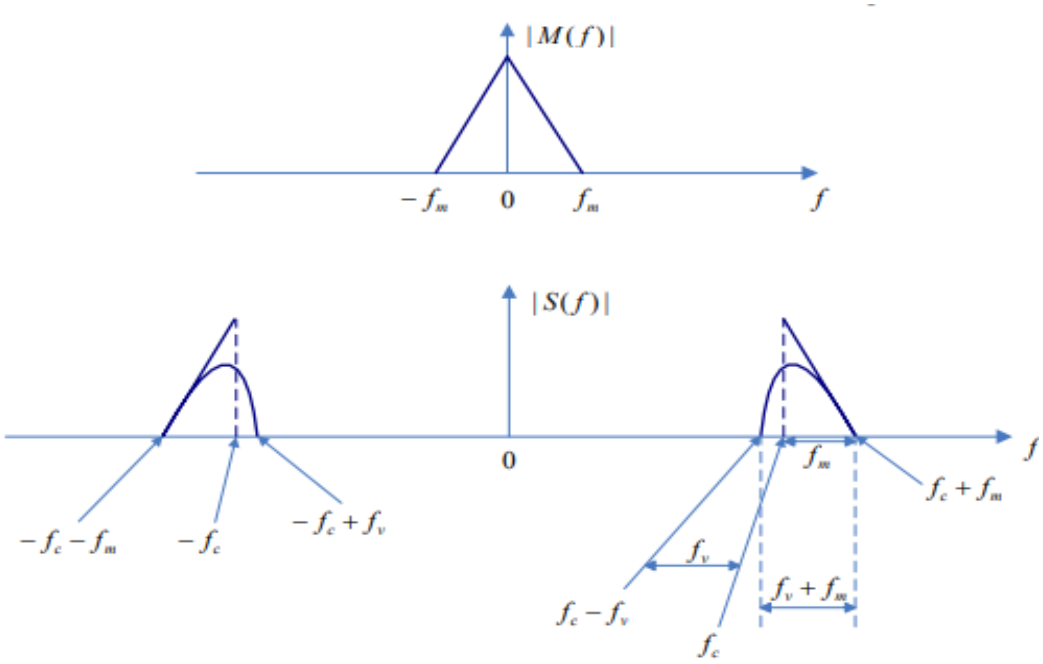


Fig. 9.4: DSB-SC Modulated Wave

The key to VSB is the sideband filter, a typical transfer function being shown below:

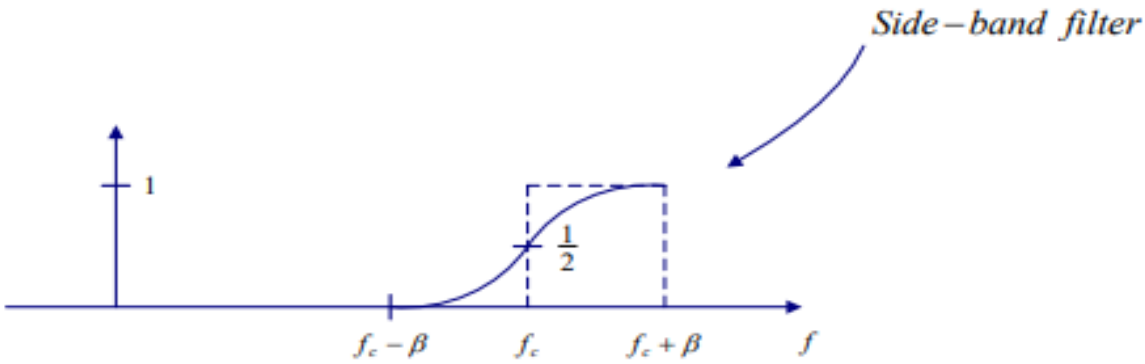


Fig. 9.5 Side Band Filter

While the exact shape of the response is not crucial, it must have odd symmetry above the carrier frequency and a relative response of 2 1 at c f. The filter response is designed so that the original message spectrum M (ω) is reproduced on demodulation as a result of the superposition of two spectra: On demodulation.

- The positive frequency part of S(ω){i.e. spectrum of transmitted signal s(t)} is shifted downward in frequency by c f .
- The negative frequency part of S(ω) is shifted upward in frequency by c f .

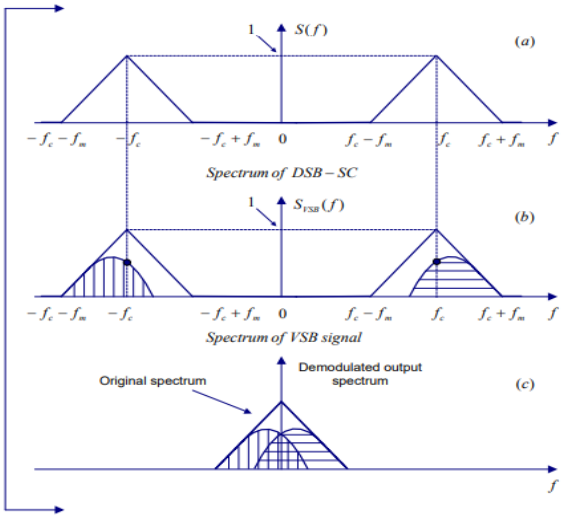


Fig. 9.6 The Negative Frequency Part of S(ω) is Shifted Upward In Frequency by c f

The magnitudes of these two spectral contributions are shown in (b) above. In effect, a reflection of the vestige of the lower sideband makes up for the missing part of the upper sideband. Vestigial sideband modulation has the virtue of conserving bandwidth almost as efficiently as single sideband modulation, while retaining the excellent low-frequency characteristics of double sideband modulation. The VSB modulation has become standard for the analogue transmission of television and similar signals, where transmission of low-frequency components is important, but the bandwidth required for double sideband transmission is unavailable.

In the transmission of television signals in practice, a controlled number of carriers is added to the VSB modulated signal. This is being done to permit the use of an envelope detector for demodulation. The design of the receiver is thereby considerably simplified.

9.4 FREQUENCY DIVISION MULTIPLEXING (FDM):

In frequency division multiplexing, the available bandwidth of a single physical medium is subdivided into several independent frequency channels. Independent message signals are translated into different frequency bands using modulation techniques, which are combined by a linear summing circuit in the multiplexer, to a composite signal. The resulting signal is then transmitted along the single channel by electromagnetic means as shown in Fig. 9.7Basic approach is to divide the available bandwidth of a single physical medium into a number of smaller, independent frequency channels. Using modulation, independent message

signals are translated into different frequency bands. All the modulated signals are combined in a linear summing circuit to form a composite signal for transmission. The carriers used to modulate the individual message signals are called sub-carriers, shown as f1, f2, …, fn in Fig. 9.8(a).

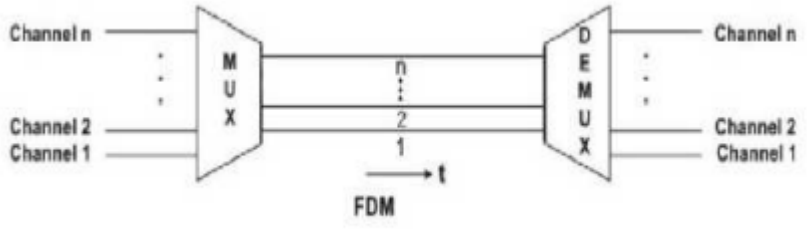


Fig. 9.7: Basic Concept of FDM

At the receiving end the signal is applied to a bank of band-pass filters, which separates individual frequency channels. The band pass filter outputs are then demodulated and distributed to different output channels as shown in Fig. 9.8(b).

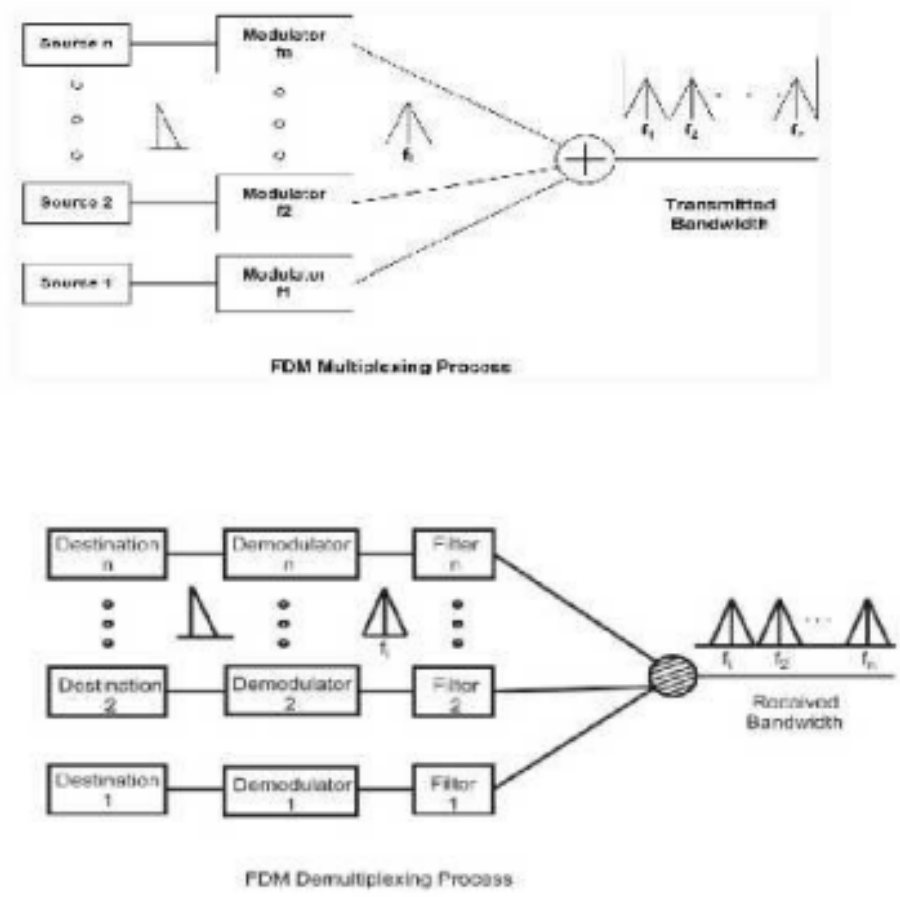


Fig 9.8 (a) FDM Multiplexing Process, (b) FDM De Multiplexing Process

If the channels are very close to one another, it leads to inter-channel cross talk. Channels must be separated by strips of unused bandwidth to prevent inter-channel cross talk. These unused channels between each successive channel are known as guard band s as shown in Figure 3.7.

FDM is commonly used in radio broadcasts and TV networks. Since, the frequency band used for voice transmission in a telephone network is 4000 Hz, for a particular cable of 48 KHz bandwidth, in the 70 to 108 KHz range, twelve separate 4 KHz sub channels.

Could be used for transmitting twelve different messages simultaneously. Each radio and TV station, in a certain broadcast area, is allotted a specific broadcast frequency, so that independent channels can be sent simultaneously in different broadcast areas. For example, the AM radio uses 540 to 1600 KHz frequency bands while the FM radio uses 88 to 108 MHz frequency bands.

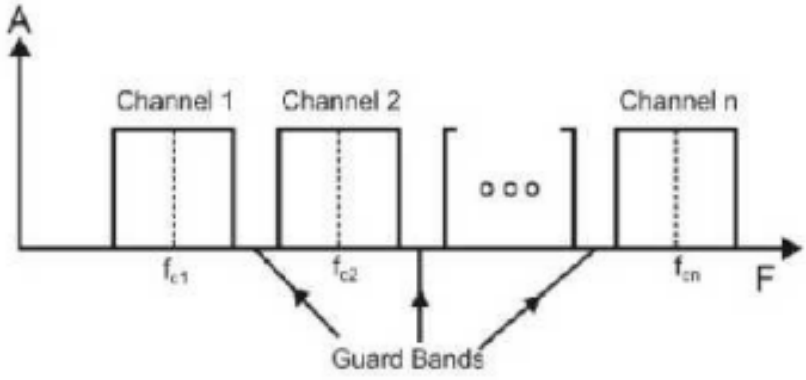


Fig. 9.9: Use of Guard Bands in FDM

9.5 SUMMARY:

This document explores key modulation techniques used in communication systems, focusing on Single Sideband (SSB), Vestigial Sideband (VSB), and Frequency Division Multiplexing (FDM). SSB modulation is an efficient alternative to conventional Amplitude Modulation (AM) and Double Sideband Suppressed Carrier (DSB-SC), as it reduces bandwidth usage by transmitting only one sideband. The generation of SSB signals can be achieved through filtering or phase discrimination methods, while coherent demodulation ensures accurate signal recovery. VSB modulation, a compromise between SSB and DSB-SC, is widely used in television broadcasting, where a partial sideband is retained to preserve low-frequency components. FDM allows multiple signals to share a single transmission medium by allocating different frequency bands to each signal, with guard bands preventing inter-channel interference. These techniques enhance spectral efficiency, reduce bandwidth wastage, and improve communication quality in applications such as radio, television, and telephony, making them essential in modern telecommunications.

9.6 TECHNICAL TERMS:

Single Sideband Suppressed Carrier (SSB-SC), Vestigial Sideband Modulation (VSB), Coherent Demodulation, Frequency Division Multiplexing (FDM).

9.7 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What is Single Sideband (SSB) modulation, and how does it improve bandwidth efficiency compared to conventional AM and DSB-SC?
2) How does Hilbert Transform help in generating an SSB signal?
3) What are the key differences between Single Sideband (SSB) and Vestigial Sideband (VSB) modulation?

Short Answer Questions:

1) Why is coherent demodulation necessary for detecting SSB signals, and how does it work?
2) How does Frequency Division Multiplexing (FDM) enable multiple signals to be transmitted simultaneously without interference?
3) What is the purpose of guard bands in FDM, and how do they prevent inter-channel crosstalk?

9.8 SUGGESTED READINGS:

1) **Lathi, B.P., & Ding, Z.** (2010). Modern Digital and Analog Communication Systems (4^{th} ed.). Oxford University Press.
2) **Haykin, S.** (2001). Communication Systems (4^{th} ed.). Wiley.
3) **Proakis, J.G., & Salehi, M.** (2007). Fundamentals of Communication Systems. Pearson Education.
4) **Taub, H., & Schilling, D.L.** (1991). Principles of Communication Systems (2^{nd} ed.) McGraw-Hill.
5) **Tomasi, W.** (2004). Electronic Communications Systems: Fundamentals through Advanced (5^{th} ed.) Pearson.

Prof. G. Naga Raju

LESSON-10

BOOLEAN ALGEBRA, LOGIC GATES AND KARNAUGH MAPS

10.0 AIM AND OBJECTIVES:

The aim of this study is to explore the principles of Boolean algebra and logic gates, which form the foundation of digital circuits. It focuses on simplifying Boolean expressions using algebraic methods and Karnaugh Maps (K-Maps) to optimize logic circuits for efficiency and reduced complexity. The study also examines fundamental logic gates, including AND, OR, NOT, XOR, XNOR, and their applications in combinational logic. Additionally, it aims to demonstrate the practical implementation of Boolean simplifications in circuit design, ensuring minimal hardware usage and improved performance. The objective is to provide a comprehensive understanding of logic gate operations, their truth tables, and their role in digital systems. By understanding Boolean laws, simplification techniques, and truth tables, the study facilitates designing efficient logic circuits, reducing power consumption and improving computational speed. This knowledge is crucial for fields such as computer engineering, embedded systems, and automation, where logical decision-making is essential.

STRUCTURE:

10.1 SIMPLIFICATION OF BOOLEAN EXPRESSIONS:

The fundamental operations in Boolean algebra are OR, AND and NOT, with symbols + (plus),. (dot) and bubble or bar over Boolean variable respectively.

In real life applications, one combined several logic gates, and the combined effect determines a system's behavior depending on the states of individual logic variables. This is usually expressed in terms of a truth table. In truth table, various inputs are listed, and various combinations are worked out to determine the system behavior. As a simple example, see the OR gate truth table given. Here A and B are the input variables. These two variables can have 2^2=4 combinations of 1s and 0s. As per the Boolean equation y = A + B (A+B to be red as A or B), output is true for 3 input combinations and false for one combination.

The following are the laws of Boolean algebra and complicated laws can be proved using simple Boolean laws. These laws can be verified by writing truth tables for L.H.S and R.H.S expressions.

1. $A + 0 = A$
2. $A + 1 = 1$
3. $A + A = A$
4. $A + \overline{A} = 1$

} Laws of 'OR'

5. $A . 0 = 0$
6. $A . 1 = A$
7. $A . A = A$
8. $A . \overline{A} = 0$

} Laws of 'AND'

9. $\overline{0} = 1$
10. $\overline{1} = 0$
11. of $A = 0$ then $\overline{A} = 1$
12. of $A = 1$ then $\overline{A} = 0$

} Laws of complementation (NOT Laws)

13. $\overline{\overline{A}} = A$
14. $A + B = B +$Commutative Laws
15. $A . B = B . A$

} Commutative Laws

16. $A + (B + C) = (A + B) + C$
17. $A .(B . C) = (A . B) . C$

} Associative Laws

18. $A . (B + C) = (A + B) . C$
19. $A + B . C = (A + B) . (A + C)$

} Distributive Laws

20. $A + \overline{A}B = A + B$

21. $A + AB = A$

22. $A(A + B) = A$

23. $A(\overline{A} + B) = AB$

24. $AB + \overline{A}B = A$

25. $(A + B)(A + \overline{B}) = A$

26. $AB + \overline{A}C = (A + C)(\overline{A} + B)$

27. $(A + B)(\overline{A} + C) = AC + \overline{A}B$

28. $AB + \overline{A}C + BC = AB + \overline{A}C$

29. $(A + B)(\overline{A} + C).(B + C) = (A + B)(\overline{A} + C)$

30. $\overline{A + B + C + \text{--}} = \overline{A}.\overline{B}.\overline{C}.\text{----}$ } De Morgan's Laws

31. $\overline{A.B.C.\text{--}} = \overline{A} + \overline{B} + \overline{C} + \text{----}$ }

10.2 ALGEBRAIC METHOD:

The Quantum Action Principle (QAP) [16] is the fundamental theorem of renormalization theory. It guarantees the locality of the counter terms and the polynomial character of the renormalization procedure. The QAP also implies that all breaking terms of the STIs and WTIs are local and that they can be fully characterized in terms of classical fields, their quantum numbers and symmetrical properties. In the case of STIs, the QAP implies that, if the green functions IΓ(n−1) satisfy all the symmetry constraints at lower orders, the (subtracted) nth-order green functions Γ(n) fulfill them up to local insertions Δ(n) S in the one-particle-irreducible (1PI) Green functions:

$$\mathbf{[S(\Gamma)]^{(n)} = \Delta_S^{(n)}} \qquad \rightarrow \mathbf{(10.1)}$$

Here Δ(n) S is an integrated, Lorentz-invariant polynomial of the fields, of the anti-fields and of their derivatives, Δ(n) S = P I γ(n) S, I R d4x ΔS, i(x), with ultra-violet (UV) degree ≤ 4 and infrared (IR) degree ≥ 3 (assuming four space-time dimensions). In the same way, the WTIs are spoiled by breaking terms of the form

$$\mathbf{W_{(\lambda)}\,\Gamma^{(n)} = \Delta_W^{(n)}\ (\lambda)} \qquad \rightarrow \mathbf{(10.2)}$$

where Δ(n) W (λ) = P I γ(n) W, I R d4x ΔW, i(x, λ) is again an integrated, Lorentz-invariant polynomial, depending linearly on the infinitesimal parameter λ, with UV degree ≤ 4 and IR degree ≥ 3.

Although Eqs. (10.1) and (10.2) apply to any renormalization scheme, the coefficients γ(n) S,I and γ(n) W,I of the various Δ(n) 's depend on the particular scheme adopted and on the order of the computation. In fact, the definitions of Δ(n) S and Δ(n) W (λ) rely on specific conventions for composite operators. Thus, a renormalization description for the composite operators is necessary. Usually, one uses the concept of Normal Product Operators introduced by Zimmermann or the conventional counter term technique, which is preferable from the practical point of view.

Once the breaking terms Δ(n) S and Δ(n) W (λ) are given, the main objective of the algebraic method can be discussed. This essentially entails a prescription to restore the identities by suitable local counterterms2, ΓCT, (n) = P I ξ (n) I R d4xLCT I (x), such that one has at nth order:

$$[S(\Gamma)]^{(n)} \equiv S_0(\Gamma^{(n)}) + \sum^{n-1}(\Gamma^{(j)},\Gamma^{(n-j)}) - S_0(\Gamma^{CT,(n)}) = 0, \qquad \rightarrow (10.3)$$

$$W_{(\lambda)}(\Gamma^{(n)}) \equiv W_{(\lambda)}(\Gamma^{(n)}) - W_{(\lambda)}(\Gamma^{CT,(n)}) = 0, \qquad \rightarrow (10.4)$$

where the decomposition given in Eqs. (10.1) and (10.2) have been used. Notice that, since the green functions IΓ(j) with j

$$[S(\Gamma)]^{(n)} = \Delta_S^{(n)}, \qquad W_{(\lambda)}\,\Gamma^{(n)} = \Delta_W^{(n)}\ (\lambda) \qquad \rightarrow (10.5)$$

Where S0 is defined below Eq. (10.2). The solution fixes a subset of the coefficients ξ (n) I of the counter terms in terms of γ(n) S, I and γ(n) Wi. These equations turn out to be solvable in the absence of anomalies, where only the consistency conditions must be used. Moreover, because of a non-trivial kernel of the operators S0 and W(λ) (i.e. the space of invariant counter terms), one is allowed to impose renormalization conditions by tuning the free parameters of the model (namely the remaining coefficients ξ (n) I of ΓCT, (n))

$$N_I(\Pi^{(n)}) = N_I(\Gamma^{(n)} + \Gamma^{CT,(n)}) = 0, \qquad \rightarrow (10.6)$$

Where the index I runs over all independent parameters of the SM and NI denotes the normalization condition operators. The existence and uniqueness of ΓCT, (n) as a solution of the system has been proven in for gauge theories, in for supersymmetric models, and in for non-semi-simple models coupled to fermion and scalar fields. Indeed, the main purpose of algebraic renormalization is to demonstrate the existence of a unique solution (up to normalization conditions) to the algebraic problem in the absence of anomalies.

Unfortunately, this does not necessarily provide a practical technique to compute the breaking terms Δ(n) and to determine the corresponding counter terms ΓCT, (n).

10.3 KARNAUGH METHOD:

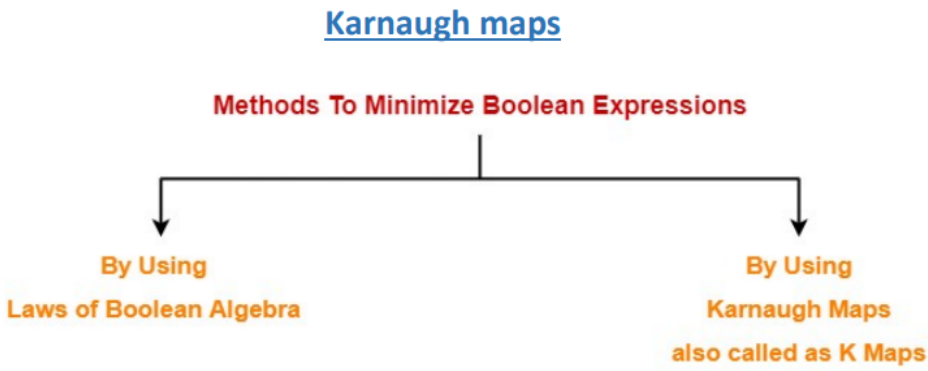


Fig. 10.1: Karnaugh Maps

1) By using laws of Boolean Algebra
2) By using Karnaugh Maps, also called K Maps

10.3.1 Karnaugh Map:

The Karnaugh Map also called as K Map is a graphical representation that provides a systematic method for simplifying the Boolean expressions.

For a Boolean expression consisting of n-variables, number of cells required in K Map = 2n cells.

10.3.1.1 Two Variable K Map:

- Two variable K Map is drawn for a Boolean expression consisting of two variables.
- The number of cells present in two variable K Map = 22 = 4 cells.
- So, for a Boolean function consisting of two variables, we draw a 2 x 2 K Map

Two variable K Map may be represented as

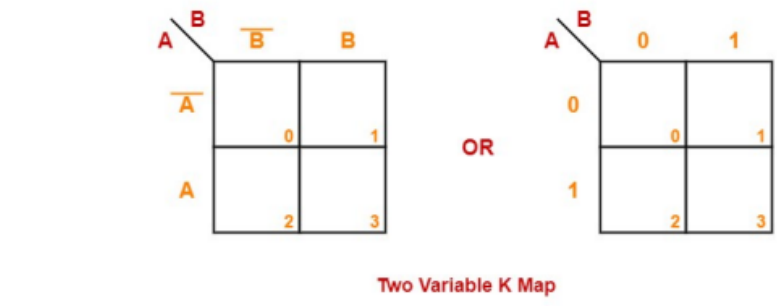


Fig 10.2: Two Variable K Map

10.3.1.2 Three Variable K Map:

- Three variable K Map is drawn for a Boolean expression consisting of three variables.
- The number of cells present in three variable K Map = 23 = 8 cells.
- So, for a Boolean function consisting of three variables, we draw a 2 x 4 K Map.

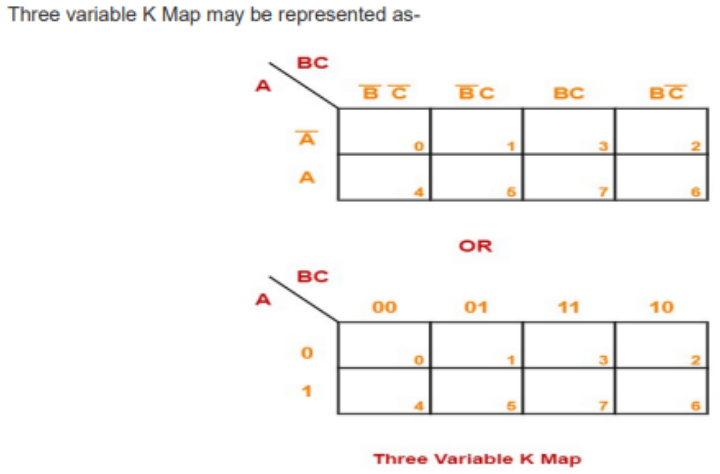


Fig. 10.3: Three Variable K Map

10.3.1.3 Four Variable K Map:

- Four variable K Map is drawn for a Boolean expression consisting of four variables.
- The number of cells present in four variable K Map = 24 = 16 cells.
- So, for a Boolean function consisting of four variables, we draw a 4 x 4 K Map.

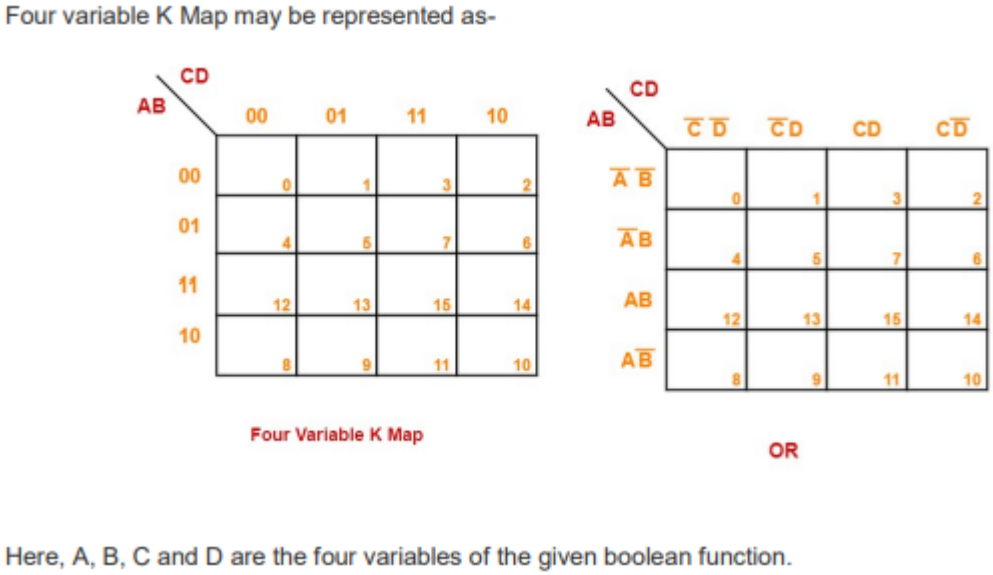
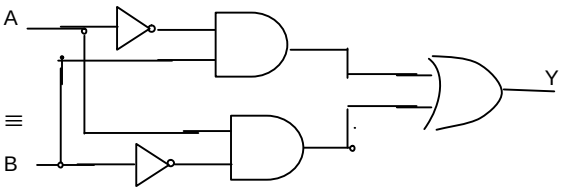


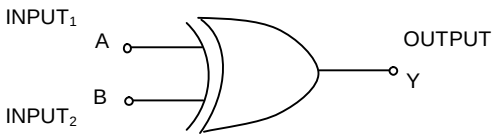
Fig. 10.4: Three Variable K Map

10.4 EX-OR:

An OR gate recognizes words with one or more 1s, whereas the exclusive – OR gate recognizes only words that have an odd number of 1s. Boolean expression for two – input

EX-OR gate is $Y = A\overline{B} + \overline{A}B = A \oplus B$. (read as Y equals A EX-OR B). The logic diagram, symbol and its truth table is shown in Fig.





Logic Symbol of EX-OR gate

TRUTH TABLE

Input		Output
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

Fig. 10.5: The Logic Diagram, Symbol and its Truth Table

When both inputs A & B in the above circuit are high or low both AND gates have low outputs to have the final output of zero. But when any one of the inputs (A or B) is high, the corresponding AND gate output is high. So, the final output is one as shown in the truth table. IC 7486 is the IC version of EX – OR gate.

10.5 X-NOR GATES:

XNOR Gate: The exclusive-NOR (XNOR), operator uses the symbol , and it performs the following logic operation X Y = X Y + X' Y' = (X ⊕ Y)' The graphic symbol and truth table of XNOR (Equivalence) gate is shown in the figure.

X	Y	XNOR
0	0	1
0	1	0
1	0	0
1	1	1

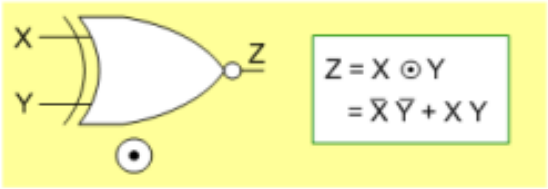


Fig. 10.6: The Logic Diagram, Symbol and its Truth Table

The result is 1 when both X and Y are 0's or when both are 1's. That is why this gate is often referred to as the Equivalence gate. The truth tables clearly show that the exclusive-NOR operation is the complement of the exclusive-OR. This can also be shown by algebraic manipulation as follows: $(X \oplus Y)' = (X Y' + X' Y)'$

$= (X Y')' (X' Y)' = (X' + Y) (X + Y')$

$= (XY + X'Y') = X \quad Y$

10.6 COMBINATIONAL LOGIC GATES:

Introduction:

- We will introduce Boolean algebra and logic gates
- Logic gates are the building blocks of digital circuits

Logic Variables:

- Different names for the same thing
    - Logic variables
    - Binary variables
    - Boolean variables
- Can only take on 2 values, e.g.,
    - TRUE or False
    - ON or OFF
    - 1 or 0

- In electronic circuits the two values can be represented by e.g.,
  - High voltage for a 1
  - Low voltage for a 0
- Note that since only 2 voltage levels are used, the circuits have greater immunity to electrical noise.

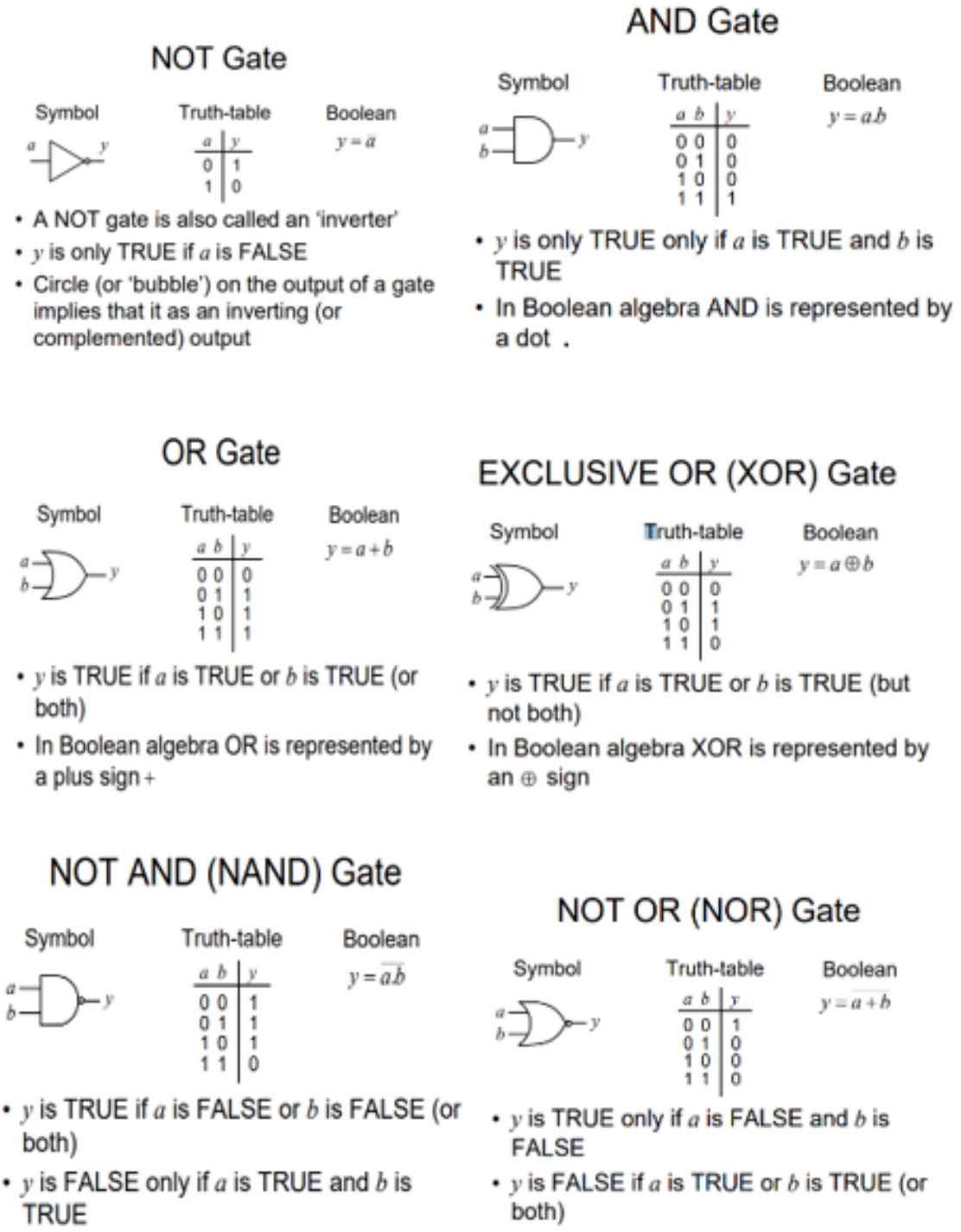


Fig 10.7: All Logic Gates Schematic Symbols with Truth Table

10.7 SUMMARY:

This study explores Boolean algebra and logic gates, which are fundamental to digital circuit design. It discusses Boolean operations, including AND, OR, NOT, XOR, and XNOR, and their applications in combinational logic. The study highlights methods for simplifying Boolean expressions using algebraic techniques and Karnaugh Maps (K-Maps) to optimize circuit efficiency. It explains truth tables, logic diagrams, and the practical implementation of logic simplifications to minimize hardware complexity. Additionally, it examines exclusive-OR and exclusive-NOR gates, emphasizing their role in digital computations. By understanding Boolean laws, simplification techniques, and truth tables, the study enables the design of efficient digital circuits with reduced power consumption and improved performance. These concepts are crucial in various applications, including computer engineering, automation, and embedded systems, where logical decision-making and circuit optimization play a significant role in enhancing overall system functionality.

10.8 TECHNICAL TERMS:

Boolean Algebra, Karnaugh Map (K-Map), Combinational Logic, Truth Table, Logic Gates, De Morgan's Theorems

10.9 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are the fundamental operations in Boolean algebra, and what symbols represent them?
2) How does the Karnaugh Map (K-Map) help in simplifying Boolean expressions?
3) What is the difference between an XOR gate and an XNOR gate in terms of functionality and truth table?

Short Answer Questions:

1) Explain De Morgan's Theorems and their significance in Boolean algebra simplification.
2) How do combinational logic circuits differ from sequential logic circuits in terms of operation and design?
3) Why is Boolean algebra important in digital circuit design, and how does it contribute to system optimization?

10.10 SUGGESTED READINGS:

1) M. Morris Mano, Digital Design, Pearson Education, 5th Edition, 2016.

2) R.P. Jain, Modern Digital Electronics, McGraw Hill Education, 4th Edition, 2009.

3) Thomas L. Floyd, Digital Fundamentals, Pearson Education, 11th Edition, 2014.

4) Zvi Kohavi and Niraj K. Jha, Switching and Finite Automata Theory, Cambridge University Press, 3rd Edition, 2009.

5) Charles H. Roth, Jr. and Larry L. Kinney, Fundamentals of Logic Design, Cengage Learning, 7th Edition, 2013.

Prof. G. Naga Raju

LESSON-11

DECODERS, ENCODERS AND MULTIPLEXERS

11.0 AIM AND OBJECTIVES:

The aim of this lesson is to provide a comprehensive understanding of decoders, encoders, multiplexers, and demultiplexers, which are essential combinational circuits in digital electronics. The objective is to explain the working principles, design, and applications of these circuits in various digital systems. Learners will understand how decoders convert binary inputs into unique outputs, how encoders perform the reverse function, and how multiplexers efficiently select data from multiple sources. The lesson also covers demultiplexers, which distribute a single input signal to multiple outputs. Additionally, the concept of sequential logic circuits, their classification, and their significance in memory and control applications will be explored. By the end of this lesson, students will be able to apply these concepts in designing digital circuits for data processing, communication, and control systems, enhancing their practical knowledge of modern digital electronics.

STRUCTURE:

11.1 DECODERS:

A decoder is a combinational circuit that converts binary information from ‘n’ input lines to a maximum of 2^n unique output lines. Decoders are available with several output configurations: active low voltage, high-sink current for direct driving of indicator lamps. Output voltage ratings range from +5V to over +100V. The decoder is used in conjunction with some code converters such as a BCD - to - 7 segment decoders, BCD - to - Decimal

decoder. The decoder presented here is called n - to 2^n line decoders decoder. Its purpose is to generate the 2^n min terms of n input variables. These decoders form a combinational circuit with n input variables and 2^n output variables. For each binary input combination of 1s and 0s, there is one and only one output line that assumes the value of 1. Figure 1.1(a) shows a 2 by 4 decoder. It has four AND gates and two inverters.

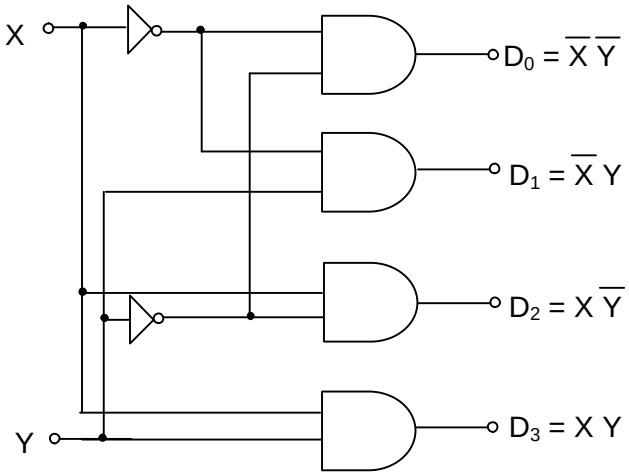


Fig 11.1(a) - Bit Decoder

X	Y	D_0	D_1	D_2	D_3
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0

Fig 11.1(b) - Bit Decoder Truth Table

Sometimes an enable input may be included with a decoder to control the circuit operation. In this, all outputs will be equal to 0, if the enable input is zero. When the enable input is 1, the circuit operates as a conventional decoder. Block diagram of 3 to 8 decoder with enable signal is shown in Fig 11.1.

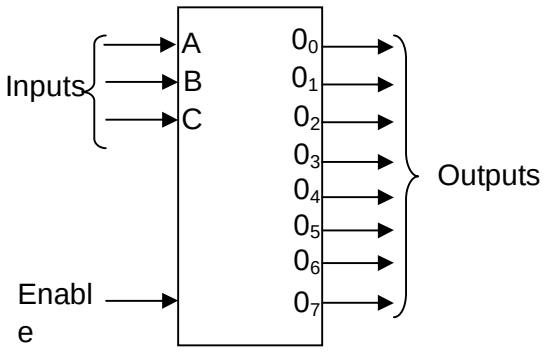


Fig 11.2 Block Diagram of 3-to-8-Bit Decoder

11.2 ENCODERS:

A decoder is a combinational circuit that converts binary information from 'n' input lines to a maximum of 2^n unique output lines. Decoders are available with several output configurations: active low voltage, high-sink current for direct driving of indicator lamps. Output voltage ratings range from +5V to over +100V. The decoder is used in conjunction with some code converters such as a BCD - to - 7 segment decoders, BCD - to - Decimal decoder. The decoder presented here is called n - to 2^n line decoders decoder. Its purpose is to generate the 2^n min terms of n input variables. These decoders form a combinational circuit with n input variables and 2^n output variables. For each binary input combination of 1s and 0s, there is one and only one output line that assumes the value of 1. Figure 11.1(a) shows a 2 by 4 decoder. It has four AND gates and two inverters.

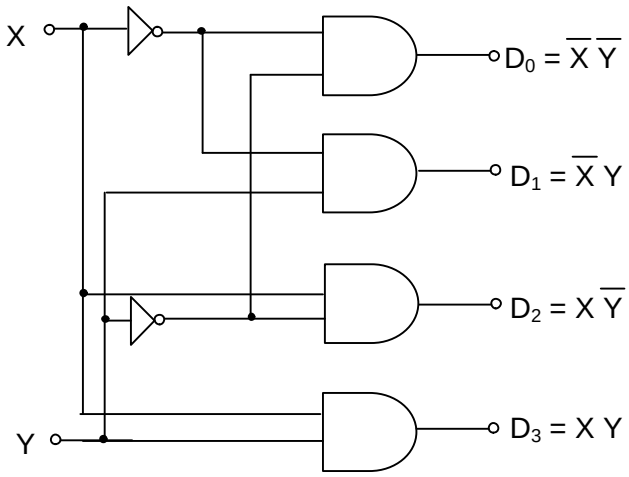


Fig11.3 -Bit Decoder

X	Y	D_0	D_1	D_2	D_3
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0

Fig 11.4-Bit Decoder Truth Table

Sometimes an enable input may be included with a decoder to control the circuit operation. In this, all outputs will be equal to 0, if the enable input is zero. When the enable input is 1, the circuit operates as a conventional decoder. Block diagram of 3 to 8 decoder with enable signal is shown in Fig. 11.4.

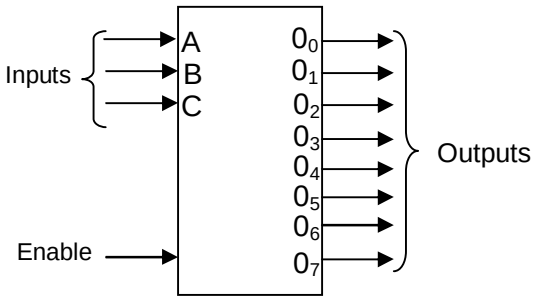


Fig 11.5 Block Diagram of 3-to-8-Bit Decoder

11.3 MULTIPLEXER (DATA SELECTORS):

It is a combinational circuit which selects binary information from one of many input lines and directs it to single outputs line. The selection of a particular input line is controlled by a set of selection lines. For 'n' bits, there are 2^n input lines and n selection lines whose bit combinations determine which inputs are selected.

This means "many into one". It is used when a complex logic circuit is shared by a few input signals. Fig 11.5 shows the logic diagram, block diagram and function table of a 4 to 1 line multiplexer.

A line to 1 line multiplexer has 4 inputs I_0 to I_3 and one output line. Each of the four input lines is applied to one input of a three input AND gate. The remaining two inputs of it are supplied by selection lines S_1 and S_0.

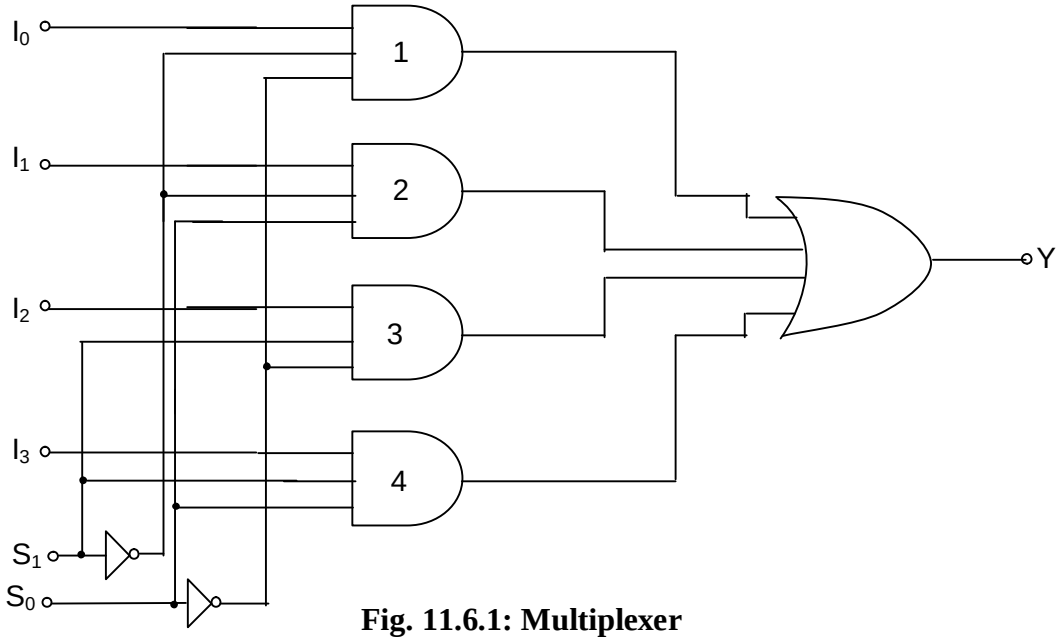


Fig. 11.6.1: Multiplexer

S_1	S_0	Y
0	0	I_0
0	1	I_1
1	0	I_2
1	1	I_3

Fig. 11.7: Truth table

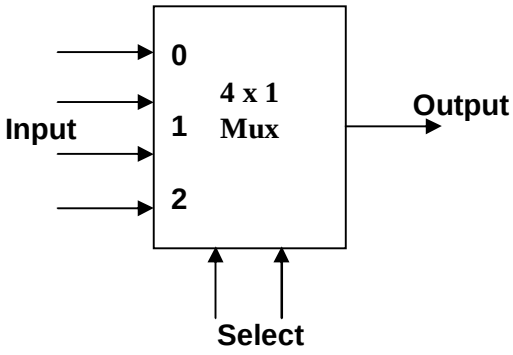


Fig 11.8: Connecting

Based on the combination of S_1 and S_0, the input associated with the selected AND gate finds out the path to reach the output through OR gate.

For example:

When $S_1S_0 = 01$, it selects the AND gate 2. So, the input I_1 is transferred to the output i.e. OR gate output is now equal to the value of I_1. Thus, providing a path from the selected input to the output. The remaining AND gates have at least one input equal to 0.

As in decoders, multiplexer IC have an enable input (or) storbe input to control the operation of the unit. It can also be used to expand into two or more multiplexers to a digital multiplexer with many inputs.

11.4 APPLICATION OF MULTIPLEXER:

Multiplexer or data selectors are combinational circuits which transfer data from many sources to output under the control of data select lines. Multiplexer has many applications right from data routing, time division multiplexing, function generator to parallel to serial converter etc. A single multiplexer can replace several logic gates ICs, saving PCB area, interconnections, design efforts and cost.

A list of popular applications is given below:

1) Data routing
2) Data bussing
3) Switch setting comparator
4) Multiplexer as a function generator
5) Parallel to serial converter
6) Cable TV signal distribution
7) Telephone network
8) Sharing printer /resources

11.5 DE MULTIPLEXER (DATA DISTRIBUTORS):

A demultiplexer is a circuit that receives information on a single line and transmits this information on one of 2^n possible output lines. The selection of a specific output line is controlled by the bit values of n selection lines. Figure 11.9 shows the block diagram of decoder and demultiplexer. The decoder with an enable input can function as a demultiplexer.

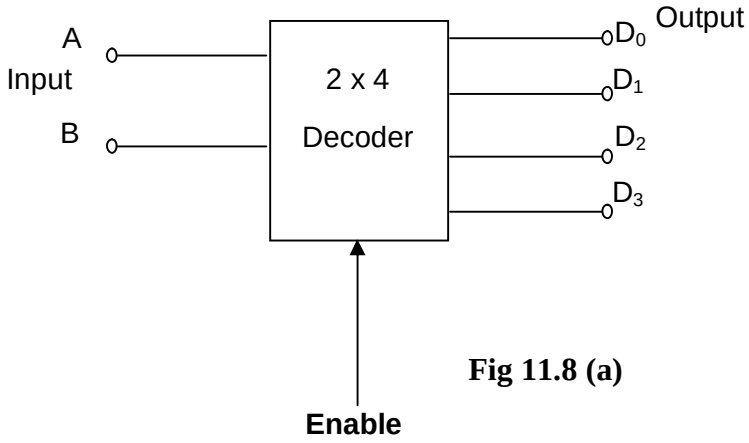


Fig 11.8 (a)

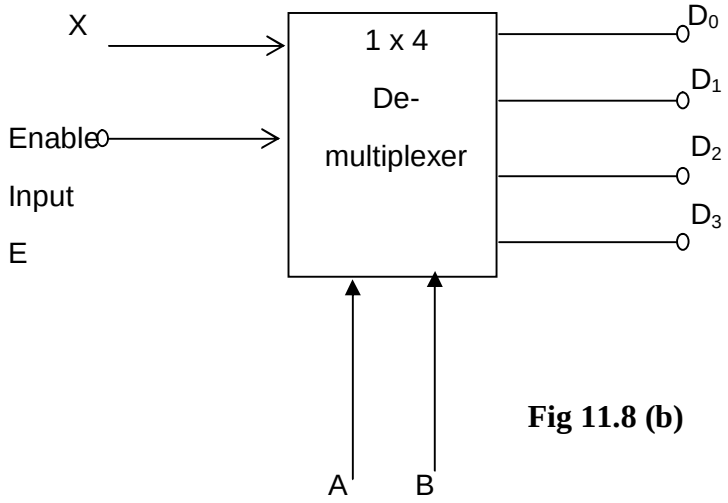
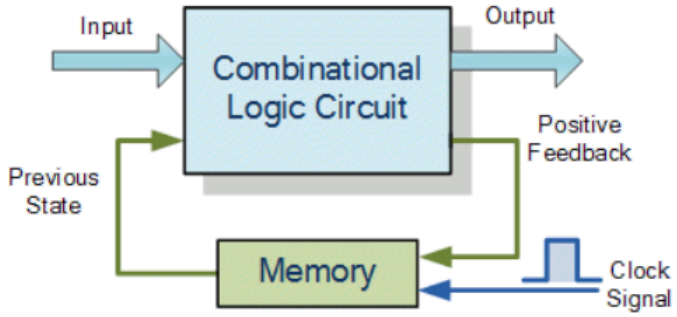


Fig 11.8 (b)

The decoder of Fig 11.9 can function as a demultiplexer if the enable line is taken as a data input line and lines A and B are taken as the selection lines as shown in Fig 11.9. Out of 4 output lines, one gets linked with input E depends upon the binary value of two selection lines A and B (i.e), if AB = 01, D_1 gets connected with the input E, so that input is available at D_1 output. While all other outputs are maintained at 1.

11.6 SEQUENTIAL LOGIC:



Sequential Logic Circuits

Sequential Logic Circuits use flip-flops as memory elements and in which their output is dependent on the input state

Fig. 11.9: Sequential Logic Circuits

Unlike circuits that change state depending upon the actual signals being applied to their inputs at that time, Sequential Logic circuits have some form of inherent “Memory” built in. This means that sequential logic circuits are able to take into account their previous input state as well as those actually present, a sort of “before” and “after” effect is involved with sequential circuits.

In other words, the output state of a “sequential logic circuit” is a function of the following three states, the “present input”, the “past input” and/or the “past output”. Sequential Logic circuits remember these conditions and stay fixed in their current state until the next clock signal changes one of the states, giving sequential logic circuits “Memory”. Sequential logic circuits are generally termed as two state or Bistable devices which can have their output or outputs set in one of two basic states, a logic level “1” or a logic level “0” and will remain “latched” (hence the name latch) indefinitely in this current state or condition until some other input trigger pulse or signal is applied which will cause the bistable to change its state once again.

11.6.1 Sequential Logic Representation:

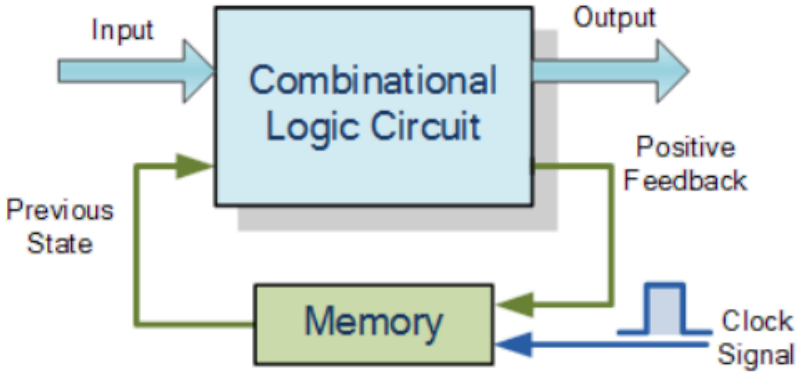


Fig. 11.10: Sequential Logic Representation

The word “Sequential” means that things happen in a “sequence”, one after another and in Sequential Logic circuits, the actual clock signal determines when things will happen next. Simple sequential logic circuits can be constructed from standard Bistable circuits such as: Flipflops, Latches and Counters and which themselves can be made by simply connecting universal NAND Gates and/or NOR Gates in a particular combinational way to produce the required sequential circuit.

Classification of Sequential Logic As standard logic gates are the building blocks of combinational circuits, bistable latches and flipflops are the basic building blocks of sequential logic circuits. Sequential logic circuits can be constructed to produce either simple edge-triggered flip-flops or more complex sequential circuits such as storage registers, shift registers, memory devices or counters. Either way sequential logic circuits can be divided into the following three main categories:

1) Event Driven - Asynchronous circuits that change state immediately when enabled.
2) Clock Driven - Aynchronous circuits that are synchronized to a specific clock signal.
3) Pulse Driven - Which is a combination of the two that responds to triggering pulses.

11.7 SUMMARY:

This lesson covered the fundamental concepts of decoders, encoders, multiplexers, and demultiplexers, which are key components in digital electronics. Decoders convert binary input into unique outputs, while encoders perform the opposite function by encoding multiple inputs into a smaller number of output lines. Multiplexers, also known as data selectors, efficiently route multiple data inputs to a single output based on selection lines, whereas demultiplexers distribute a single input to multiple outputs. The lesson also introduced sequential logic circuits, which retain memory and operate based on past and present inputs,

making them essential for flip-flops, registers, and counters. The applications of these circuits in communication, data processing, and control systems were discussed. Understanding these concepts is crucial for designing efficient digital systems, enabling students to develop solutions for modern computing and electronic applications.

11.8 TECHNICAL TERMS:

Decoder, Encoder, Multiplexer (MUX), Demultiplexer (DEMUX), Sequential Logic, Flip-Flop

11.9 SELF-ASSESSMENT QUESTIONS:

1) What is the primary function of a decoder, and how does it differ from an encoder?
2) Explain the working principle of a multiplexer (MUX). How does it differ from a demultiplexer (DEMUX)?
3) What is the role of selection lines in a multiplexer and demultiplexer?
4) Describe a practical application of multiplexers in digital systems.
5) What is sequential logic, and how does it differ from combinational logic circuits?
6) Name and briefly explain the three main categories of sequential logic circuits.

11.10 SUGGESTED READINGS:

1) **Anand Kumar**, Fundamentals of Digital Circuits, PHI Learning – Discusses digital circuit design, including logic gates and combinational circuits.
2) **William I. Fletcher**, An Engineering Approach to Digital Design, Prentice-Hall – Explains digital circuit design techniques with practical applications.
3) **Tocci, Widmer & Moss**, Digital Systems: Principles and Applications, Pearson – Covers both theory and real-world applications of digital circuits.
4) **Floyd & Jain**, Digital Fundamentals, Pearson Education – A comprehensive introduction to digital logic and combinational circuits.
5) **R.P. Jain**, Modern Digital Electronics, McGraw-Hill Education – Provides detailed explanations of digital circuits, including decoders, encoders, and multiplexers.

Prof. G. Naga Raju

LESSON-12

FLIP-FLOPS

12.0 AIM AND OBJECTIVES:

The aim of this lesson is to provide a comprehensive understanding of modulation and demodulation, focusing on flip-flops as essential components in digital circuits. The objective is to explain different types of flip-flops, including R-S, JK, Master-Slave, T, and D flip-flops, highlighting their working principles, truth tables, and applications. This lesson covers the role of clock pulses, preset and clear inputs, and the race-around condition, along with methods to eliminate it using master-slave configurations. Additionally, it explores the significance of flip-flops in memory storage, frequency division, and sequential circuit design. By the end of the lesson, learners should be able to identify various flip-flop types, analyze their behavior, and apply them in digital logic circuits for data storage and processing, laying the foundation for advanced sequential circuit design.

STRUCTURE:

12.1 FLIP FLOPS:

A flip-flop is a binary cell, which can store a bit of information. It itself is a sequential circuit. A flip-flop maintains any one of the two stable states that can be treated as 1 or 0 depending on the presence or absence of output signal. The state of flip-flops can only change when clock pulse has arrived. Let us first see the basic flip-flop or a latch.

12.2 A 1 BIT MEMORY:

- "Single-bit" Memory (Foreground) - Individual memory circuits that store a single bit of information and have at least the following I/O ports:

  1) data input and 2) data output

  a) "Clock-less" latch

  b) Transparent Level-Sensitive Latch

  c) Edge-triggered Flip-Flop

- "Array" Memory (Background) - Large memory circuit that stores many bits of information organized into multiple words accessed by an address

  1) SRAM
  2) Multi-ported SRAM
  3) DRAM
  4) Flash
  5) etc.

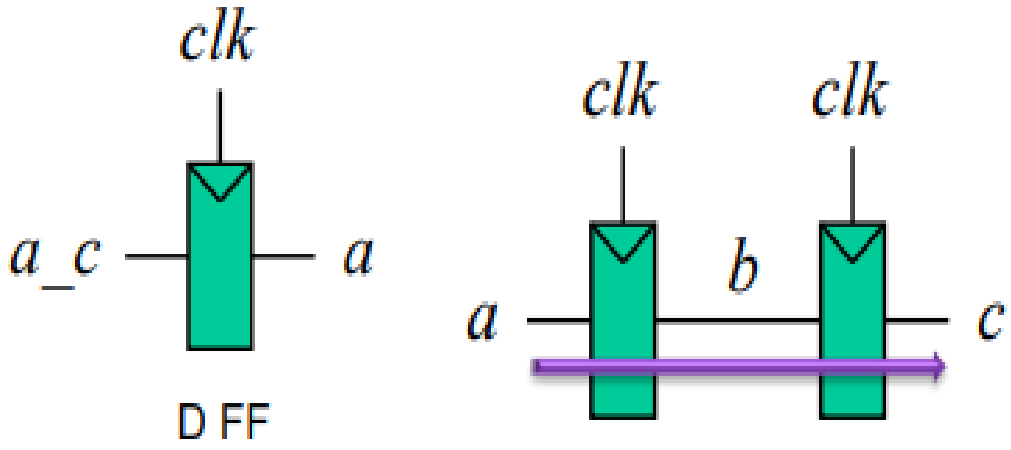


Fig. 12.1: Diagram of A 1 Bit Memory

- The correct solution is to use a "non-blocking assignment" written with "<=" which causes the simulator to evaluate the right side of the expression when the statement is encountered, but the assignment of the left side is not done until the end of that time step in "simulator time" .
- With the verilog below, the registers perform as normal FFs behave without a race regardless of their ordering reg b, c; always @(posedge clk) begin b <= a; c <= b; end.

12.3 THE R-S FLIP-FLOP:

The main feature in R-S Flip-Flops is the addition of a clock pulse input. In this flip-flop change in the value of R or S will change the state of the flip-flop only if the clock pulse at that moment is one. It is shown in fig 12.2(a).

The excitation or characteristic table basically represents the effect of S and R inputs on the state of the flip-flop, irrespective of the current state of flip-flop. The other two inputs P(present) and C(clear) are asynchronous inputs and can be used to set the flip-flop or clear the flip-flop respectively at the start of operation, independent of the clock pulse.

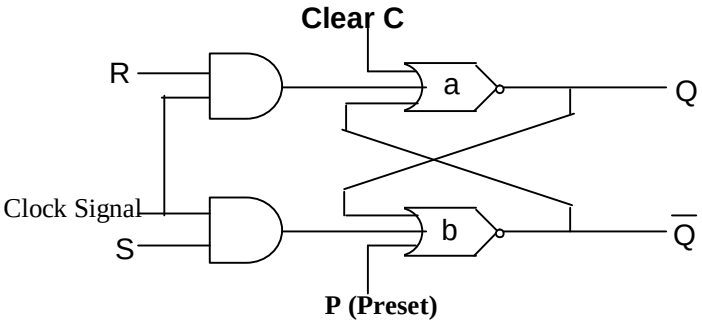


Fig. 12.2(a): Logic Diagram

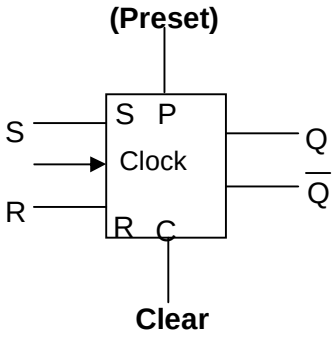


Fig. 12.2(b) Symbolic Representation

Input S	Input R	State on completion of clock cycle
0	0	No change in State
0	1	Clear the flip – flop (state 0)
1	0	Set the flip – flop (state 1)

Fig. 12.2(c): Characteristic table.

Fig. 12.2: R-S Flip-Flop

12.4 JK FLIP-FLOP:

It is not possible to achieve toggling (changing the state of the output wherever the input makes a transition from 1 to 0) with the simple flip-flops described above. By using NAND gates, one can construct a flip-flop that toggles. The symbol and truth table of one such circuit, called J-K flip-flop are shown in fig 12.3.

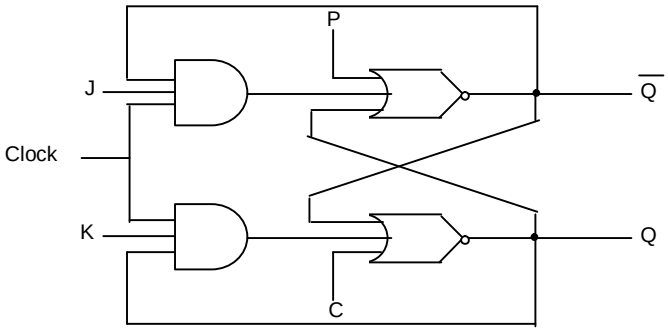
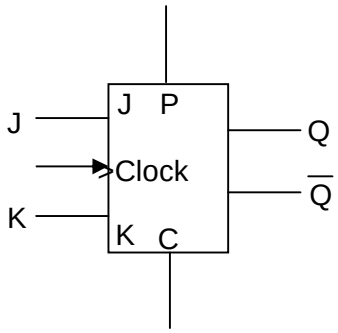


Fig 12.3(a): Logic Diagram



Input J	Input K	State on completion of clock cycle
0	0	No change in State
0	1	Clear the flip –f lop (state 0)
1	0	Set the flip – flop (state 1)

Fig. 12.3(b): Symbolic representation

Fig. 12.3(c): Truth Table

Fig 12.3: J-K Flip-Flop

A JK flip-flop is a refinement of the SR flip-flop in that the indeterminate condition of the SR type defines in the JK type. An examination of the truth table given in fig 4.1(a) reveals the following properties of the JK flip-flop.

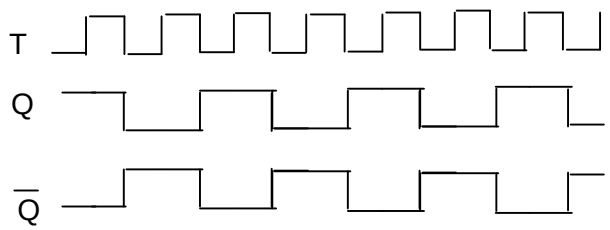
a) It retains its present state if J = 0 and K = 0

b) If J = 0, K = 1, Q = 0, it resets to zero

c) If J = 1, K = 0, Q = 1, it sets

d) If J = 1, K = 1, it Toggles.

If the first clock pulse leaves Q = 1 and $\overline{Q}$ = 0 the second clock pulsemake Q = 0 and $\overline{Q}$ = 0. The third clock pulse make Q = 1 and $\overline{Q}$ = 0 again and so on as shown



Relation between clock pulses and output JK flip-flop. When J = 1 and K = 1.

Please note Q = 1 once in every two clock pulses i.e., the output has half the frequency of the clock pulse. This property is utilized in binary counters.

12.4.1 Preset and Clear Inputs:

Usually, the J-K Flip - Flop works on the basis of the clock input. However, we can introduce two inputs called the PRESET and CLEAR in the last NAND level. As shown in fig 12.4(a) these inputs do not obey the clock pulses and override them in action. Hence, they are called the asynchronous inputs. The working of these inputs is as given below.

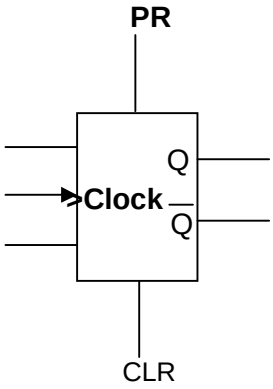


Fig 12.4(a) Symbolic Representation

PR	CLR	Q_0	Q_1
0	0	Forbidden	operations
0	1	0/1	1
		0/1	0
1	0	Normal	operations

Fig. 12.4(b): Truth Table

Fig. 12.4: J-K Flip – Flop

Let PR = 0 and CLR = 1. Then, as and when PR = 1 and CLR = 0, Q will jump to 1 from whatever state it is in. This operation presets $\overline{Q}$ to 1. When PR = 1 and CLR = 0, the lower NAND produces 1 at its output, i.e., $\overline{Q}$ = 1. This makes Q = 0. Thus whatever the state is, Q will fall back to '0'. When both PR = CLR = 1, the NANDs are enabled and the J-K flip-flop functions in the normal way. However, the state PR = CLR = 0 is forbidden, since both Q and $\overline{Q}$ will be 1 in this state. The truth table for the PR and CLR terminals is as shown in Table 12.4(a), fig. 12.4(b) shows the symbol of J-K flip-flop.

12.4.2 RACE – Around Conditions:

Consider the last two rows of the truth table of the J-K flip-flop, where both the J and K inputs are connected to logic-1 state. We find that when J = K = 1, Q_0 changes to $\overline{Q}_0$. That is, if $Q_0 = 0$, $Q_1 = 1$, and if $Q_0 = 1$, $Q_1 = 0$. These transitions gives rise to what is called the race-around problem. When J = K = 1 and CK = 1 the output will toggle from '0' to '1', and 1 to 0 continuously, until the clock pulse becomes zero. The result of this toggling action is that when the clock finally becomes zero, we well not know in what state Q would be.

This is called race-around problem, as it is generated due to the racing of this signal around the feedback path, and around the circuit.

12.5 JK MASTER SLAVE FLIP-FLOP:

The race around problem can be solved using two J-K flip-flops in the master-slave mode of operation. Here two S-R flip-flops are in the J-K mode of operation. The first flip-flop in the 'master' and the second one is the 'slave'. The master is converted into J-K mode of operation.

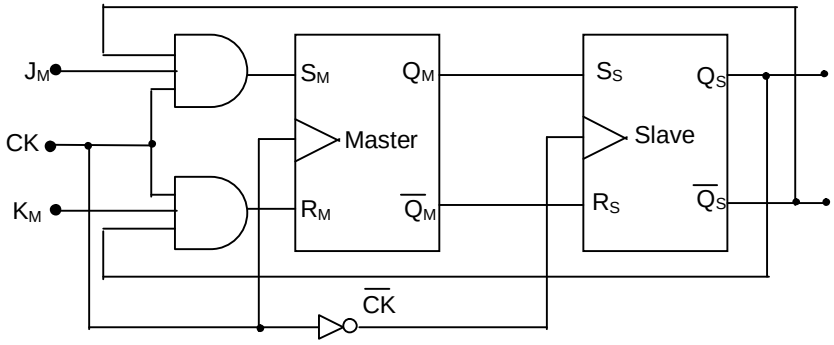


Fig. 12.5: J-K Master-Slave Flip – Flop

The slave inputs are connected to the outputs of the master. In this mode, the slave will also act as a J-K Flip – Flop. The Master is driven by clock pulse CK, while the slave is driven by $\overline{CK}$, which is obtained by inverting the CK pulses. The feedback connection can be seen to be from the Slave outputs to the master inputs.

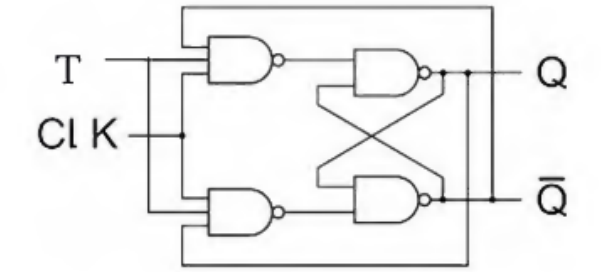
When the clock is ON, the master gets enabled, and the slave gets disabled. So, the inputs appearing at input terminals S_M and R_M of the master will appear at its outputs Q_M and $\overline{Q}_M$. respectively, which means that the external inputs now appear at the inputs S_s and R_s of the slave.

When the CK = 0, the master is disabled, and the slave is enabled because $\overline{CK} = 1$. Thus, the inputs at S_s and R_s will appear at Q_s and $\overline{Q}_s$, respectively, and hence at the inputs J_M and K_M of the master. However, as the master is now disabled, these inputs will not appear at S_s and R_s. Hence, no feedback problem occurs. Thus, the race-around problem is eliminated.

12.6 T FLIP-FLOP:

- This is a much simpler version of the J-K flip flop.
- Both the J and K inputs are connected and thus are also called a single input J-K flip flop.

- When clock pulse is given to the flip flop, the output begins to toggle.
- Here also the restriction on the pulse width can be eliminated with a master-slave or edge-triggered construction. Look at the circuit and truth table below.



Excitation Table for T Flip Flop:

Q_n	T	Q_{n+1}
0	0	0
0	1	1
1	0	1
1	1	0

K map for T Flip Flop:

Q_n \ T	0	1
0		1
1	1	

Characteristic Equation:

$$Q_{n+1} = T.\overline{Q_n} + \overline{T}.Q_n$$

Fig. 12.6: Characteristic Equations of T Filp - Flops

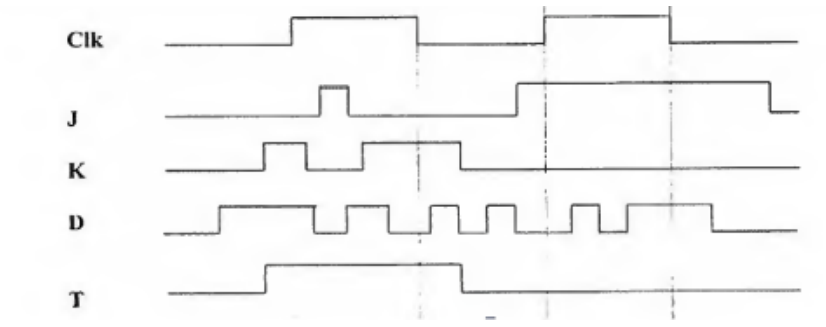


Fig. 12.7: CLK Diagram of T Flip - Flop

12.7 D FLIP – FLOP:

D flip-flops are special type offlip-flops, and they represent the currently applied input as the state of the flip-flop. It can store 1 bit of data information and sometimes refers to as Data flip-flops. The state of flip-flop changes with the applied input. It does not have a condition where the state does not change as the case in RS flip-flops, the state of R-S flip-flops does not change when S = 0 and R = 0. If we want that for a particular input state does not change, then either the clock is to be disabled during that period or a feedback of the output can be embedded with the D- flip-flop is also known as Delay Flip-Flop because it delays the 0 or 1 applied to its input by a single clock pulse.

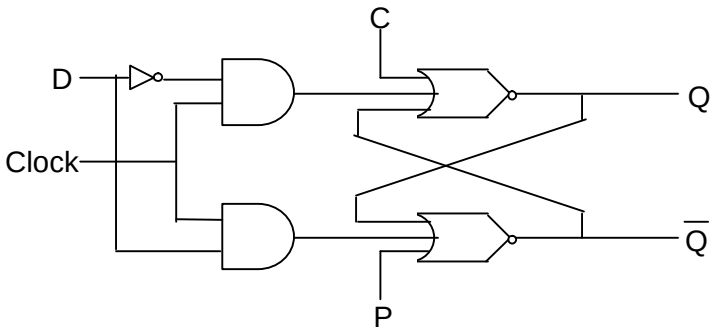


Fig. 12.8(a): Logic Diagram

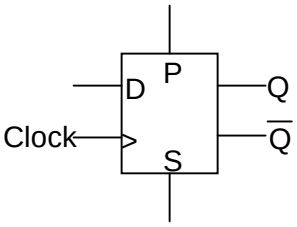


Fig. 12.8(b): Symbolic Representation

Input D	State on completion of clock cycle
0	0
1	1

Fig. 12.8(c): Truth Table of D Flip-Flop

Fig. 12.8: D Flip - Flop

12.8 SUMMARY:

This lesson explores modulation, demodulation, and the role of flip-flops in digital circuits. It introduces different types of flip-flops, including R-S, JK, Master-Slave, T, and D flip-flops, explaining their working principles, truth tables, and applications. The importance of clock pulses, preset and clear inputs, and the race-around condition is discussed, along with techniques like the master-slave configuration to overcome this issue. The lesson highlights the significance of flip-flops in memory storage, frequency division, and sequential circuit design. Additionally, it explains how D flip-flops function as data storage elements. By understanding these concepts, learners can apply flip-flops in designing reliable digital systems, ensuring proper data storage and transfer in various electronic and computing applications.

12.9 TECHNICAL TERMS:

Clock Pulse, Master-Slave Flip-Flop, Edge-Triggered Flip-Flop, Characteristic Table

12.10 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What is a flip-flop, and how does it function as a basic memory element in digital circuits?

2) Explain the race-around condition in a JK flip-flop and how it can be eliminated.

3) What are the differences between R-S, JK, T, and D flip-flops in terms of functionality and applications?

Short Answer Questions:

1) How do preset and clear inputs work in a flip-flop, and why are they considered asynchronous?
2) What is the significance of clock pulses in flip-flops, and how do edge-triggered flip-flops operate?
3) Describe the working principle of a master-slave flip-flop and its advantages over a standard JK flip-flop.

12.11 SUGGESTED READINGS:

1) **M. Morris Mano**, Digital Design, Pearson Education – A comprehensive guide to digital circuits and flip-flops.
2) **R.P. Jain**, Modern Digital Electronics, McGraw Hill – Covers sequential circuits and memory elements in depth.
3) **Floyd & Jain**, Digital Fundamentals, Pearson – Detailed explanations of flip-flops and their applications.
4) **Tocci & Widmer**, Digital Systems: Principles and Applications, Pearson – A practical approach to digital circuit design.
5) **William I. Fletcher**, An Engineering Approach to Digital Design, Prentice Hall – Discusses sequential logic circuits, including flip-flops.

Prof. G. Naga Raju

LESSON-13

SHIFT REGISTERS, COUNTERS, AND A/D AND D/A CONVERTERS

13.0 AIM AND OBJECTIVES:

The aim of this lesson is to understand the concepts and applications of shift registers, counters, and A/D to D/A converters in digital systems. Shift registers are essential for data storage and transfer, enabling efficient movement of binary data within a system. Counters, both synchronous and asynchronous, play a critical role in digital counting applications, with synchronous counters offering synchronized clock pulses and asynchronous counters operating through sequential triggering. The lesson also explores cascade counters, which enhance counting range by connecting multiple counters. Additionally, it covers the working principles of A/D and D/A converters, crucial for converting analog signals into digital format and vice versa. Various A/D conversion techniques, including simultaneous, successive approximation, and dual-slope methods, are examined, along with digital-to-analog conversion methods like weighted-resistor and R-2R ladder networks. Understanding these components is fundamental for designing efficient digital circuits used in computing, communication, and control systems.

STRUCTURE:

13.1 SHIFT REGISTERS:

Shift registers are very important in applications involving the storage and transfer of data in digital systems. A register, in general, is used solely for storing and shifting data (1s and 0s) entered from all external sources and possesses no characteristic internal sequence of states. The storage capability of a register is one of its two basic functional characteristics and makes it an important type of memory device.

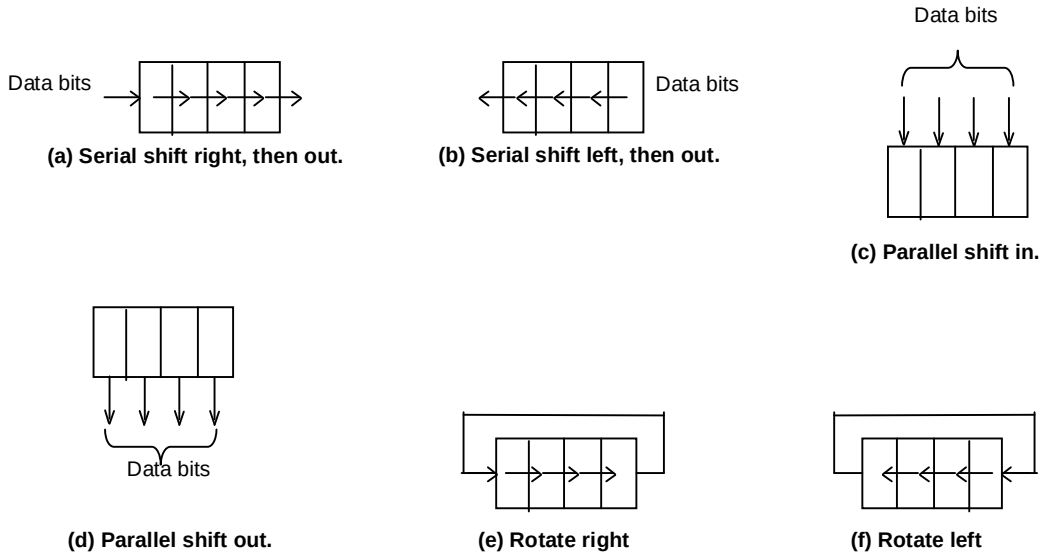


Figure 13.1: Basic Data Movement in Registers

Registers are commonly used for the temporary storage of data within a digital system. The shift capability of a register permits the movement of data from stage to stage within the register or into or out of the register upon application of clock pulses.

Fig. 13.1 shows symbolically the types of data movement in shift register operations. The block represents any arbitrary four-bit register, and the arrow indicates the direction and type of data movement.

13.2 SYNCHRONOUS AND ASYNCHRONOUS COUNTER:

13.2.1 Synchronous Counters:

Synchronous counters are distinguished from ripple counters in that clock pulses are applied to the CP inputs of all flip-flops. The common pulse triggers all the flip-flops simultaneously, rather than one at a time in succession as in a ripple counter. The decision whether a flip-flop is to be complemented or not is determined from the values of the J and K inputs at the time of the pulse. If J = K = 0, the flip-flop remains unchanged. If J = K = 1, the flip-flop complements. In this section, we present some typical MSI synchronous counters and explain their operation.

(a) Three-Bit Synchronous Binary Counter:

Three-bit synchronous binary counter is shown in fig 13.2(a) and its timing diagram in fig 13.2(b). An understanding of this counter can be achieved by a careful examination of its sequence of states.

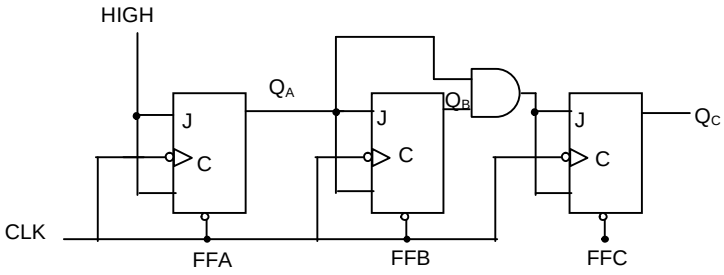


Fig. 13.2(a): A Three-Bit Synchronous Binary Counter

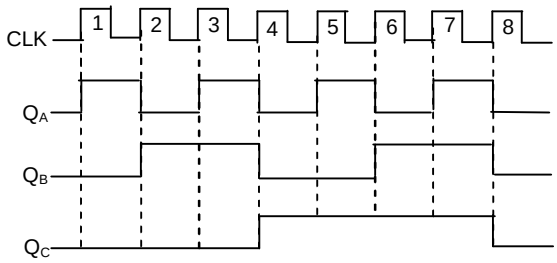


Fig. 13.2(b): Timing Diagram for the Counter of Figure

First, let us look at Q_A. Notice that Q_A changes on each clock pulse as we progress from its original state to its final state and then back to its original state. To produce this operation, FFA must be held in the toggle mode by constant HIGH on its J and K inputs. Now let us see what Q_B does. Notice that it goes to the opposite state following each time Q_A is a 1. This occurs at CLK_2, CLK_4, CLK_6 and CLK_8. CLK_8 causes the counter to recycle. To produce this operation, Q_A is connected to the J and K inputs of FFB. When Q_A is a 1 and a clock pulse occurs. FFB is the toggle mode and will change state. The other times when Q_A is a 0, FFB is in the no-change mode and remains in its present state.

Table 13.1 State sequence for a three-stage binary counter

Clock	Q_C	Q_B	Q_A
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

Next, let us see how FFC is made to change at the proper times according to the binary sequence. Notice that both times Q_C changes state, it is preceded by the unique condition of both Q_A and Q_B being HIGH. This condition is detected by the AND gate and applied to the J and K inputs of FFC. Whenever both Q_A and Q_B being HIGH, the output of the AND gate

makes the J and K inputs of FFC HIGH, and FFC toggle on the following clock pulse. At all other times, the J and K inputs of FFC are held LOW by the AND gate output, and FFC does not change state.

13.2.2 Asynchronous Counters:

In a asynchronous counter, the flip-flop output transition serves as a source of triggering other flip-flops. In other words, the CP inputs of all flip-flops are triggered not by the incoming pulses but rather by the transition that that occurs in other flip-flops. In this section, we present some asynchronous counters and explain their operation.

13.2.2.1. Four-Bit Asynchronous Binary Counter:

Four-Bit asynchronous binary counter consists of a series connection of complementing flip-flops, with the output of each flip-flop connected to the CP input of the next higher-order flip-flop. The flip-flop lading the least significant bit receives the incoming count pulses. The diagram of a 4 - bit binary ripple counter is shown in fig 13.3. All J and K inputs are equal to '1'. The small circle in the CP input indicates that the flip flop complements during a negative-going transition or when the output to which it is connected goes from 1 to 0.

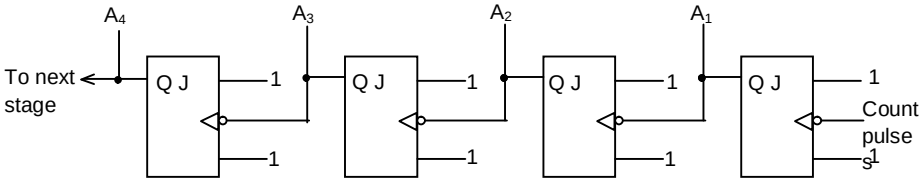


Fig. 13.3: Bit Binary Ripple Counter

Table 13.2 Count Sequence for A Binary Ripple Counter

Count sequence $A_4\ A_3\ A_2\ A_1$	Conditions for complementing flip-flops
0 0 0 0	Complement A_1
0 0 0 1	Complement A_1 A_1 will go from 1 to 0 and Complement A_2
0 0 1 0	Complement A_1
0 0 1 1	Complement A_1 A_1 will go from 1 to 0 and Complement A_2
	A_2 will go from 1 to 0 and Complement A_3
0 1 0 0	Complement A_1
0 1 0 1	Complement A_1 A_1 will go from 1 to 0 and Complement A_2

To understand the operation of the binary counter, refer to its count sequence given in Table 2.2. It is obvious that the lowest order bit A_1 must be complemented with each count pulse. Every time A_1 goes from 1 to 0, it complements A_2. Every time A_2 goes from 1 to 0, it complements A_3, and so on. For example, take the transition from count 0111 to 1000. The arrows in the table emphasize the transition in this case. A_1 is complemented with the count pulse. Since A_1 goes from 1 to 0, it triggers A_2 and complements it. As a result, A_2 goes from 1 to 0, which in turn complements A_3. A_3 now goes from 1 to 0, which complements A_4. The output transition of A_4, if connected to the next stage, will not trigger the next flip-flop, since it goes from 0 to 1. The flip-flops change is at a time in rapid success on, and the signal propagates through the counter in a ripple fashion. Hence asynchronous counters are sometimes called asynchronous counters.

13.2.2.2. The 7493A Four-Bit binary Counter:

The 7493A is presented as an example of a specific integrated circuit asynchronous counter. As the logic diagram in fig 2.1 shows, this device consists of a single flip-flop and a three -bit asynchronous counter. This arrangement is for flexibility. It can be used as a divide - by-2 device using only the single flip-flop, or it can be used as a modulus - 8 counters using only the three-bit counter position. This device also provides gated reset inputs, RO (1) and RO (2). When both of these inputs are HIGH, the counter is RESET to the 0000 state by $\overline{CLR}$.

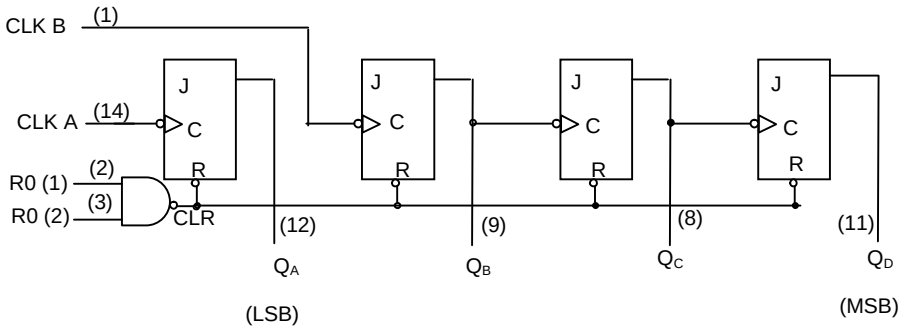
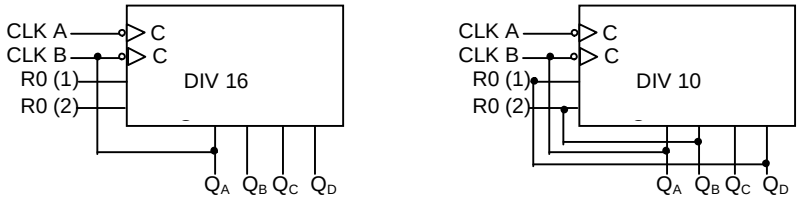


Fig. 13.4: The 7493A four-bit binary counter logic diagram. (Pin numbers are in parentheses, and all J-K inputs are internally connected HIGH.

Additionally, the 7493A can be used as a four-bit modules-16 counter (counts 0 through 151) by connecting Q_A output to the CLK B input as shown in fig. 13.5.



(a) 7493A connected as a modulus-16 counter

(b) 7493A connected as a decade counter

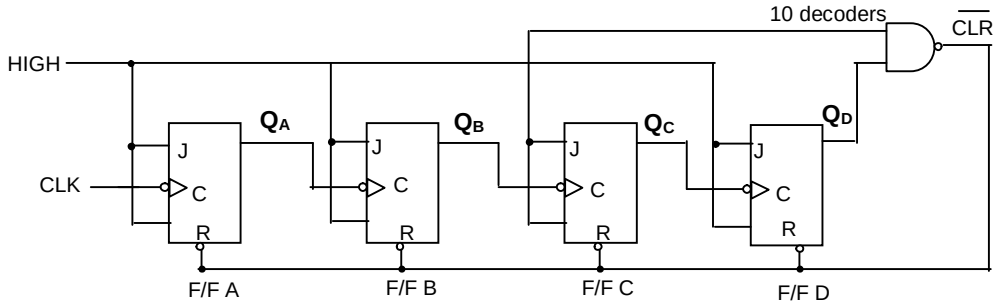
Fig. 13.5: Two Configurations of the 7493A Asynchronous Counter

It can also be configured as a decade counter with asynchronous recycling by using the gated reset inputs for partial decoding of count 10_{10}, as shown in fig. 13.5.

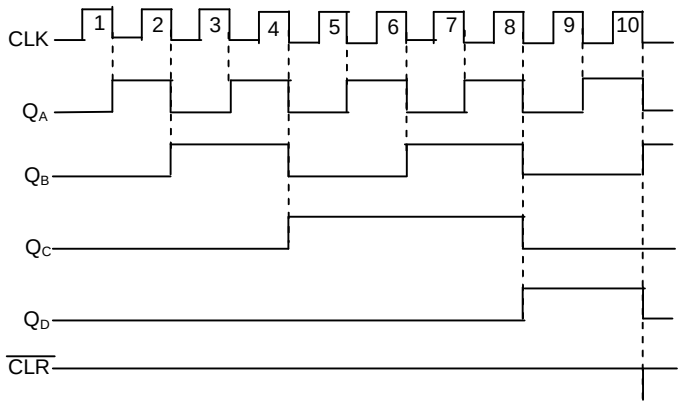
13.2.2.3. Asynchronous Decade Counters:

Counters with ten states in their sequence are called decade counter. A decade counter with a count sequence of 0(0000) through 9(1001) is a BCD decade counter because its ten-state sequence is the BCD code. This type of counter is very useful in display applications in which BCD is required for conversion to a decimal readout.

A decade counter requires four flip-flops. We will now take a four-bit asynchronous counter and modify its sequence in order to understand the principle of truncated counters. One method of achieving this recycling after the count of 9(1001) is to decode count 10_{10}(1010) with a NAND gate and connect the output of the NAND gate to the clear (CLR) inputs of the flip-flops, as shown in fig 13.6.



(a)



(b)

Fig. 13.6: An Asynchronously Clocked Decade Counter with Asynchronous Recycling

Notice that only Q_B and Q_D are connected to the NAND gate inputs. This is an example of partial decoding, in which the two unique states (Q_B = 1and Q_D = 0) are sufficient to decode the count of 10_{10} because none of the other states (0 through 9) have both Q_B and Q_D HIGH at

the same time. When the counter goes into count 10_{10}(1010), the decoding gate output goes LOW and asynchronously RESETS all of the flip-flops.

The resulting timing diagrams is shown in fig. Notice that there is a glitch on the Q_B wave form. The reason for this glitch is that Q_B must first go HIGH before the count of 10_{10} can be decoded. Not until several nano seconds after the counter goes to the count of 10_{10} does the output of the decoding gate go LOW. Thus, the counter is in the 1010 state for a short time before it is RESET back to 0000, thus producing the glitch on Q_B.

13.3 CASCADE COUNTERS:

- A counter will increment only when - The counter below it is at its terminal count, and it is being incremented.
  - That is the definition of the Rollover signal.
- Some people try to tie Rollover to the clk input of the next higher counter-– Bad idea… Very bad idea…
  - Violates our Globally Synchronous policy
  - Doesn't work as intended

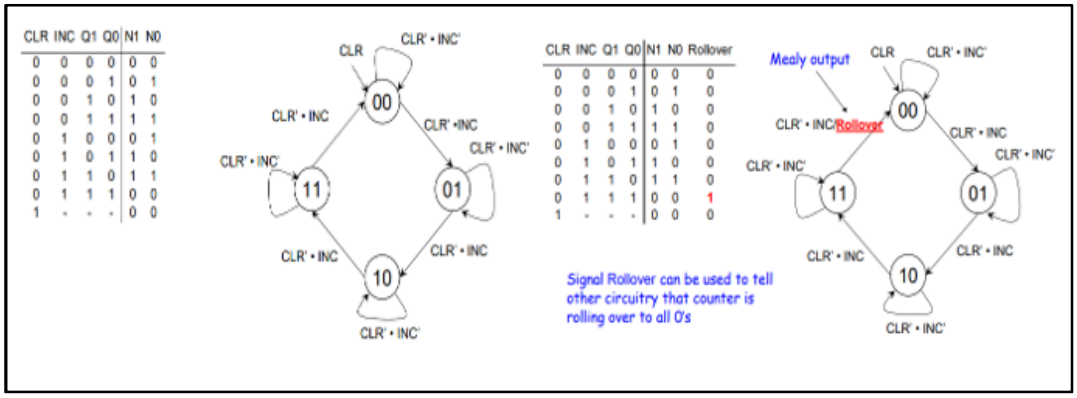


Fig. 13.7: Cascade Interface Signals

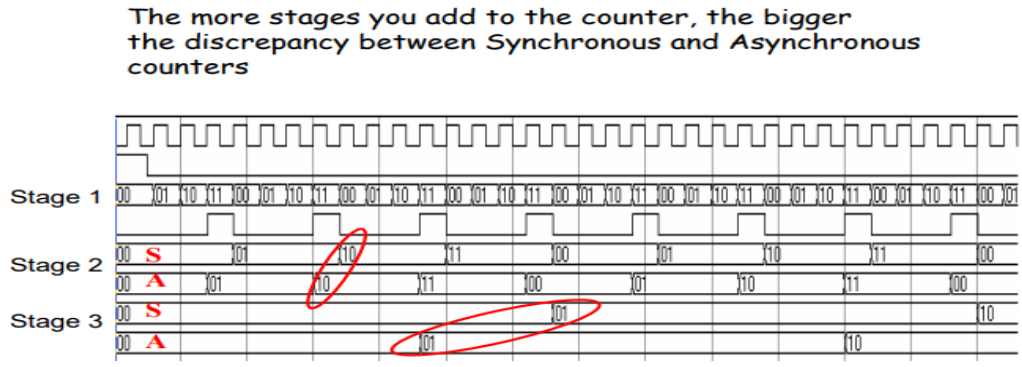


Fig. 13.8: Discrepancy between Synchronous and Asynchronous Counters

13.4 A/D TO D/A CONVERTERS:

13.4.1 Analog-to-Digital Converter:

Analog–to–digital conversion is slightly more complex than digital–to–analog conversion, and several different methods may be used. Four common conversion methods are described in this section. Of these, the successive approximation counter is the most widely used A/D converter, because it provides excellent performance for a wide range of applications at a reasonable cost.

The comparator circuit forms the basis of all A/D converters. This circuit compares an unknown voltage with a reference voltage and indicates which of the two voltages is larger. A comparator is essentially a multistage high gain differential amplifier, where the state of the output is determined by the relative polarity of the two input signals. If, for instance, input signal A is greater than input signal B, the output voltage is maximum, and the comparator is on. If input signal A is smaller than input signal B, the output voltage is minimum, and the comparator is off. Since the amplifier has a very high gain, it either saturates or cuts off at relatively low differential input levels, so that it acts as a binary device.

13.4.1.1 Simultaneous A/D Converter:

Simultaneous A/D converter is shown in Fig. 13.9, where three comparator circuits are used. Each of the three comparators has a reference input voltage, derived from a precision reference voltage source. A resistive divider consisting of four equal precision resistors is connected across the reference supply and provides output voltages of $\frac{3}{4}$V, $\frac{1}{2}$V, and $\frac{1}{4}$V, where V is the reference output voltage. The other input terminal of each comparator is driven by the unknown analog voltage.

In this example the comparator is on (providing an output) if the analog voltage is larger than the reference voltage. If none of the comparators is on, the analog input must be smaller than $\frac{1}{4}$V. If comparator C_1 is on and both C_2 and C_3 are off, the analog voltage must be between $\frac{1}{4}$V and $\frac{1}{2}$V. Similarly, if C_1 and C_2 are both on and C_3 is off, the analog voltage must be between $\frac{1}{2}$V and $\frac{3}{4}$V; if all comparators are on, the analog voltage must be greater than $\frac{3}{4}$V. In total, four different output conditions may exist from no comparators on to all comparators on. The analog input voltage can therefore be resolved in four equal steps. These four output conditions can be coded to give two binary bits of information. This is shown in

the table alongside the diagram of 12.6. Seven comparators would give three binary bits of information, fifteen comparators would give four bits, etc.

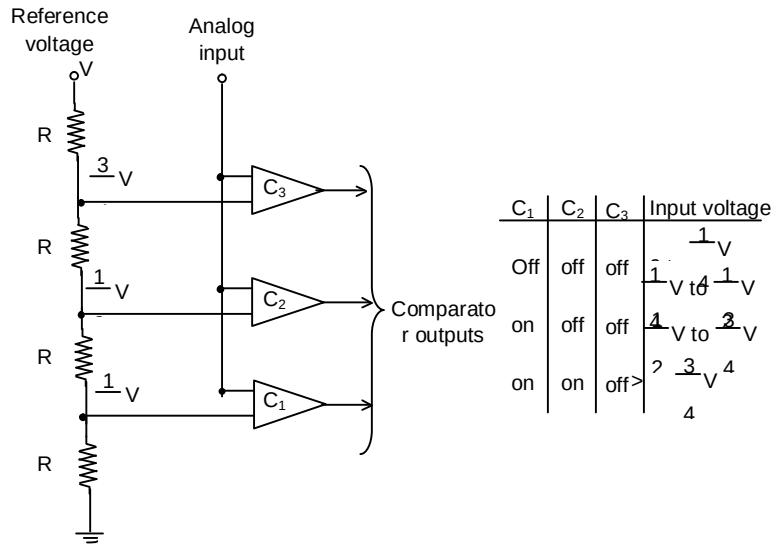


Fig. 13.9: Simultaneous Analog-to-Digital Converter

The advantage of the simultaneous system of A/D conversion is its simplicity and speed of operation, especially when low resolution is required. For a high – resolution system (many bits), this method requires so many comparators that the system becomes bulky and very costly.

13.4.1.2 Successive-Approximation A/D Converter:

This is one of the most popular A/D converters and has a resemblance with counter type A/D converters. Fig. 13.10 gives the outline of a successive approximation A/D converter.

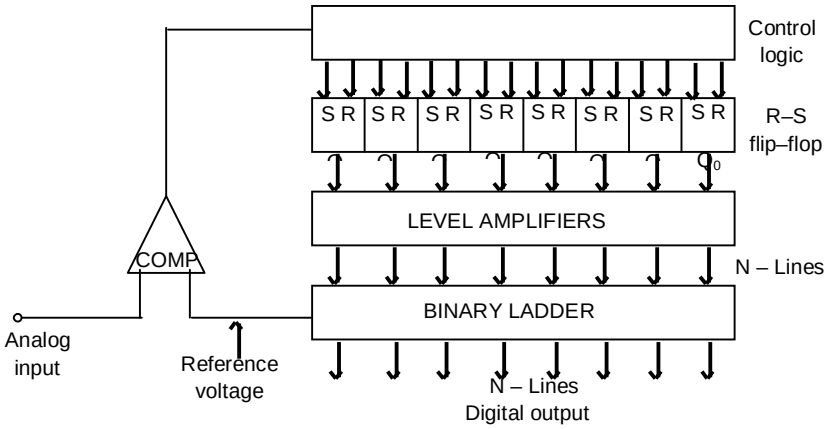


Fig. 13.11: Block Diagram of A Successive Approximation A/D

The converter circuit comprises of a control logic that is used to set or reset the flip-flops X_0, X_1….X_7. The output of these flip flops are given to the level amplifiers and then to the binary ladder which provides the digital output as well as the analog reference voltage.

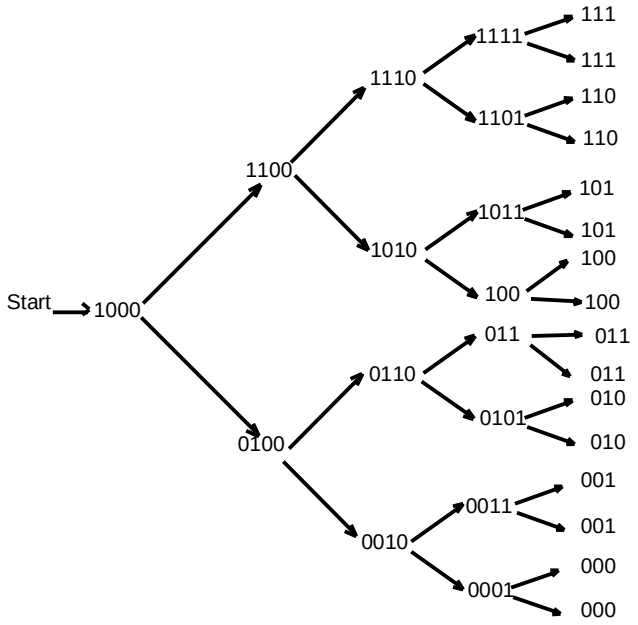


Fig. 13.12: Operation Diagram for a 4-bit Successive Approximation A/D Converter

Initially, the control logic resets all the flip-flops Q_0 to Q_7. Now its MSB is set to 1 so that digital signal reaching the binary ladder is 1000 0000. Analog voltage produced by binary ladder is now compared with the analog input signal. If the reference signal is smaller than the analog input, the comparator output gives a signal to the control logic which now switches the output of $Q_6$to 1 so that digital signal becomes 1100 0000 and its equivalent analog voltage is compared with the input signal. On the other hand, if in the first comparison, comparator finds that reference voltage is more than the analog input, then a signal is given to the control logic to reduce the reference voltage accordingly. $Q_{7\ is}$ reset to '0' and Q_6 is set to 1. The digital signal now becomes 0100 0000 which is again compared with the analog signal. This process is repeated until analog input and reference voltage produced become equal. At this stage, a stable digital output is produced. The complete approximation procedure is illustrated for a four-bit successive approximation A/D converter with the help of operation diagram fig 13.12.

An important feature of successive approximation A/D converter is that an 'n-bit'A/D converter requires only 'n' clock cycles for conversion. For a 8-bit A/D converter of this type, if a clock frequency of 1MHz is employed, the conversion time is only 8μs; which is quite low as compared to counter type or continuous type A/D converter.

13.4.1.2.1 Advantages:

1) It is more accurate than the staircase A/D converter.
2) It maintains a high resolution.

3) It is much faster.

4) There is much less.

13.4.1.2.2 Disadvantages: -

1) It requires a complex register called the successive-approximation register (SAR).

2) It is costly, as it contains more components.

13.4.1.3 Dual - Slope A/D Converter:

The Dual - Slope A/D converter is shown in fig. 13.13.

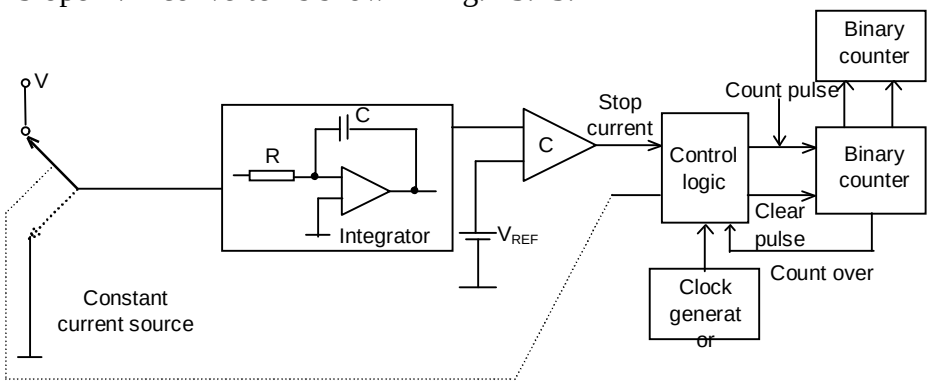


Fig. 13.13: Circuit of Dual–Slope A/D Converter

The basic principle of the operation of the circuit can be understood from fig. 13.13. In Fig. 13.13, we find that we charge a capacitor C for a fixed duration. This means that the integrator capacitor will charge to a voltage level depending on the value of the input analog voltage. Now, after the charging interval is over, the integrator capacitor is allowed to discharge at a constant (uniform). rate through a constant-current circuit. Depending on the level to which the capacitor had been initially charged to, the discharge time varies as shown, and hence the binary counter's counting time varies. This then reads the input voltage in terms of the count duration. The larger this duration, the more the count.

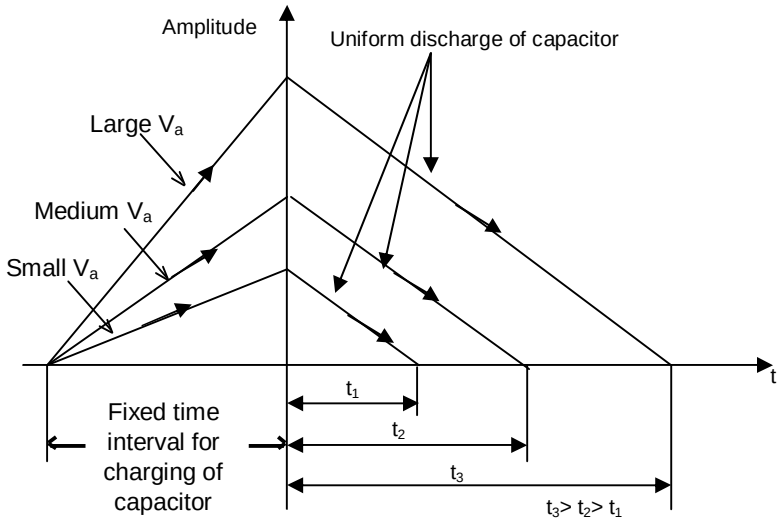


Fig. 13.14: Charge-Discharge Intervals of Integrator

In the circuit shown in fig 13.14, there is a toggle switch that connects the input data and the constant-current generator to the integrator, alternately. The toggle switch is controlled by a control-logic circuit. The output of the integrator, alternately. This toggle switch is controlled by a control-logic circuit. The output of the integrator is applied to a comparator, which gives a stop command once its output exceeds the reference voltage V_{ref}. The control logic controls the clock, start count and clear count pulses of the binary counter, as well as the signal-pole, double-throw (SPDT) toggle switch.

At the start command, the integrator starts charging to the level of the input voltage for a fixed interval. Digital counter is not energized in this interval. At the end of the fixed-time interval the control switch throws the SPDT switch back to the constant-current generator, and the binary counter starts counting. When the integrator voltage drops to a value such as $V_{ref} > V_{int}$, the comparator stops the count. The reading on the digital counter represents the value of the input V_a.

13.4.2 Digital - to - Analog Converter [D/A]:

The Digital to Analog converter accepts data in digital form and converts it to a voltage or current which is proportional to the digital value.

The black diagram of a n-bit D/A converter is shown in fig 13.16. It has no logic inputs and a single analog output, whose value varies discretely in response to the 2^n possible input bit patterns. The internal circuit can be all electronically switched ladder network.

Normally two types of networks are used:

i) Weighted resistor network ii) R – 2R ladder network.

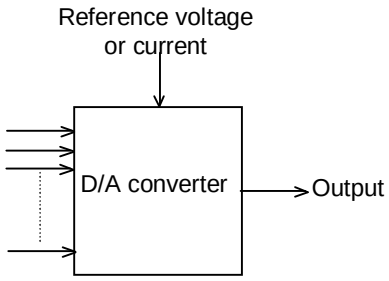


Fig. 13.16: D/A Converter–Block Diagram

13.4.2.1 The Weighted - Resistor Type D/A converter:

The diagram of the weighted – resistor D/A converter is shown in fig 13.17. The operational amplifier is used to produce a weighted sum of the digital inputs, where the weights are proportional to the weights of the bit positions of inputs. Since the op-amp is connected as an inverting amplifier, each input is amplified by a factor equal to the ratio of

the feedback resistance divided by the input resistance to which it is connected. The most significant bit (MSB) i.e. D_3 is amplified by R_f / R, D_2 is amplified by R_f / 2R, D_1 is amplified by R_f / 4R, and D_0, the LSB is amplified by R_f / 8R.

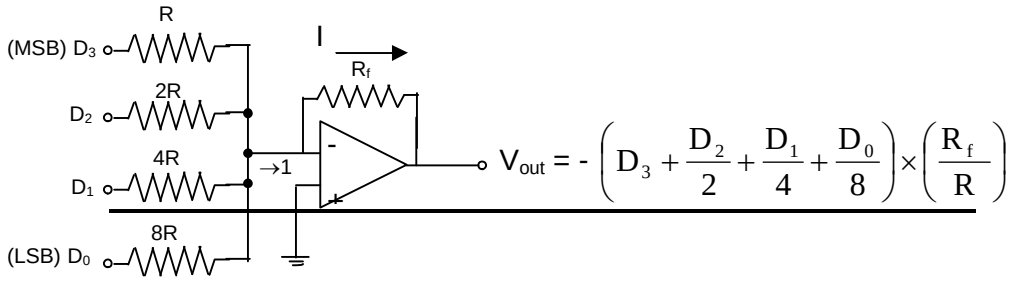


Fig 13.17 Weighted – Resistor Type D/A Converter

The inverting terminal of the op-amp in fig.13.17 acts a virtual ground. Since the op-amp adds and inverts.

$$V_{out} = - \left(D_3 + \frac{D_2}{2} + \frac{D_1}{4} + \frac{D_0}{8} \right) \times \left(\frac{R_f}{R} \right) \qquad \rightarrow (13.1)$$

The main disadvantage of this type of D/A Converter is, that a different-valued precision resistor must be used for each bit position of the digital input.

13.4.2.2 Binary R – 2R Ladder D/A Converter:

This D/A Converter circuit avoids the above problem and requires only two resistors R and 2R (hence the name R-2R), although each bit position requires two resistors instead of the single resistor per bit position of the binary weighted converter. The-n-bit R-2R ladder network with output operational amplifier is shown in fig. 13.18.

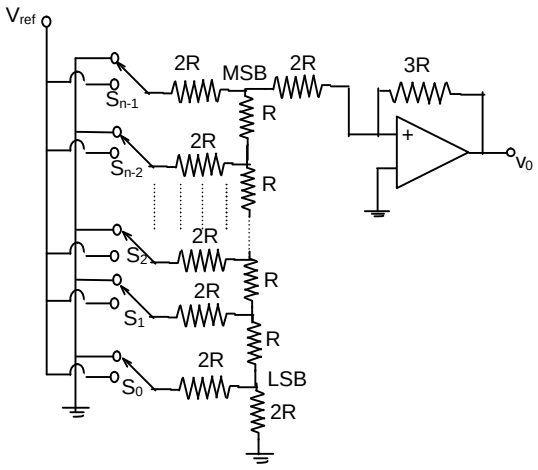


Fig. 13.18: Binary R–2R Ladder D/A Converter

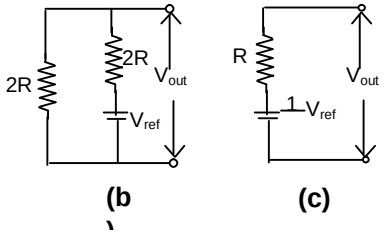


Fig. 13.19: Thevenin's Equivalent Circuits with only the LSB - Switch

The circuit may be analyzed using Thevenin's theorem.

Assume that $S_0 = 1$, $S_1 = S_2 =$ ---- $S_{n-1} = 0$. The R- 2R network above the LSB position is then effectively the load for the circuit which can be redrawn as in. The voltage across the open circuit terminals is thus V/2 and the value of series resistance is R (as 2R in parallel with 2R). The Thevenin's equivalent circuit is always the same as for any bit position (i.e., $S_1 = 1$, $S_0 = S_2 = S_3 =$……. $S_{n+1} = 0$). To find the contribution of each bit position to the output voltage, let us redraw the circuit for the 3-bit D/A Converter and use the Thevenin's theorem, Fig.13.2 shows the effective Thevenin's circuits for each bit position.

We see that the weighted value of voltage of LSB, 1st, 2nd …MSB are respectively $V_{ref}/2^n$, $V_{ref}/2^{n-1}$, $V_{ref}/2^{n-2}$ … $V_{ref}/2$ volts. Therefore, for n-bits D/A converter, the V_{out} at node X is given by

$$V_{out} = V_{ref}(S_0 2^{-n} + S_1 2^{-n-1} + S_2\, 2^{-(n-2)} + \ldots S_{n-2}\, 2^{-2} + S_{n-1}\, 2^{-1}) \text{ ---------- } (13.2)$$

In a common practice an additional resistor 2R is placed to the right of the MSB switch position so that the total load looking into the array is 3R. With a feedback resistor of value 3R, the gain of the operational amplifier is –1. Thus, the output of Op-amp or of the n-bit R-2R ladder with Op-amp is given by

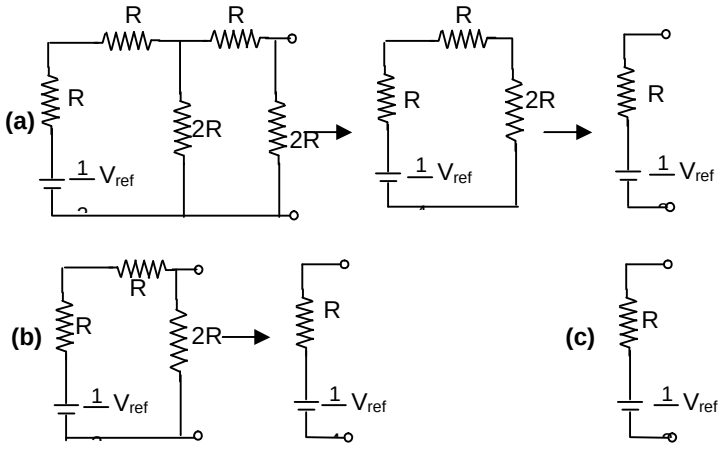


Fig. 13.20: Effective Thevenin's Circuits for (a) MSB, (b) Next MSB and (c) LSB of the 3-bit R–2R D/A Converter

$$V_0 = V_{ref}(S_0 2^{-n} + S_1 2^{-n-1} + S_2 2^{-(n-2)} + \ldots S_{n-2} 2^{-2} + S_{n-1} 2^{-1}) \quad \text{---------- (13.3)}$$

The negative sign corresponds to the inverting property of the Op-amp.

13.4.2.3 D/A converter Specifications:

1) **Resolution:** The resolution is specified as the value of 'n' or the number of bits which the D/A converter can accept. The resolution can also be expressed as % Resolution = $100/(2^n - 1)\%$. Thus a 3-bit D/A converter is of poor resolution. For a better resolution n must be large. The voltage conversion resolution is generally defined as $V_{ref}/2^n$.

2) **Linearity:** An ideal D/A converter would have equal increments of the analog output for equal increments in the digital input. The deviation from the linear behavior is called non-linearity. It is expressed as

% Non-linearity = 100D/M

Where D is the deviation of measured output from a straight line and M is the output range, usually the non-linearity is expressed as $\pm\frac{1}{2}$ LSB volts.

3) **Accuracy:** The accuracy of a D/A converter is a measure of the difference between the actual analog output voltage and the ideal output voltage. There are many factors contributing to accuracy including the linearity of the D/A converter, tolerances on reference voltages, amplifier gains, etc, and ability of the circuit to resist noise.

4) **Offset:** For zero input there may be a finite output, which is called zero error. This is quite small and may be offset by adjustment (trimming) at the Op-amp stage.

5) **Gain error:** The gain is the ratio of the D/A converter's full scale output value and the reference input value. The gain error is deviation of actual gain from the designed value. It can be minimized by using external preset resistors.

6) **Response or setting time:** The setting time is defined as the time required for the output to reach and remain within, a specified band of voltage (or current) after an input change has occurred. It is caused by transients set up by voltage switching and the effects on the stray capacitance in the circuit. It is of the order of nano to microseconds.

7) **Temperature Sensitivity:** For a given fixed input value, the output analog value varies with temperature, because of the sensitivity of the reference voltage, the resistors, Op-amp to temperature. It may be as high as ±50 ppm/^{0}C.

13.5 SUMMARY:

This lesson provides a comprehensive understanding of shift registers, counters, and A/D to D/A converters, which are fundamental components in digital electronics. Shift registers facilitate the storage and transfer of binary data, making them essential for data communication and processing. Counters, including synchronous and asynchronous types, are explored for their role in counting and timing applications, with cascade counters extending their range. The lesson also delves into the working principles of A/D and D/A converters, which enable the conversion between analog and digital signals. Various A/D conversion techniques, such as simultaneous, successive approximation, and dual-slope methods, are discussed alongside digital-to-analog conversion techniques like weighted-resistor and R-2R ladder networks. Understanding these concepts is crucial for designing digital circuits used in computing, communication, and control applications, ensuring efficient signal processing and system functionality.

13.6 TECHNICAL TERMS:

Shift Registers, Counters, Cascade Counters, A/D Converter (Analog-to-Digital Converter), D/A Converter (Digital-to-Analog Converter).

13.7 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are shift registers, and how are they used in digital circuits?
2) How do counters function, and what are their applications in electronic systems?
3) What is the purpose of cascade counters, and how do they extend counting capability?

Short Answer Questions:

1) How does an A/D converter work, and why is it important in signal processing?
2) What is the role of a D/A converter in digital systems?
3) How does the successive approximation method improve the accuracy of A/D conversion?

13.8 SUGGESTED READINGS:

1) **Morris Mano, M.** (2017). Digital Design: With an Introduction to the Verilog HDL, VHDL, and System Verilog (6th ed.). Pearson.
2) **Floyd, T. L.** (2018). Digital Fundamentals (11th ed.). Pearson.
3) **Sedra, A. S., & Smith, K. C.** (2020). Microelectronic Circuits (8th ed.). Oxford University Press.
4) **Malvino, A. P., & Brown, J. A.** (2018). Digital Computer Electronics (3rd ed.). McGraw-Hill.
5) **Razavi, B.** (2021). Design of Analog CMOS Integrated Circuits (2nd ed.). McGraw-Hill.

Prof. Ch. Linga Raju

LESSON-14

MICROCOMPUTERS AND 8085 MICROPROCESSORS

14.0 AIM AND OBJECTIVES:

.The aim of studying microcomputers and the 8085 microprocessors is to understand their architecture, functionality, and applications in computing and control systems. Microcomputers, being compact and application-specific, play a crucial role in various fields such as industrial automation, communication, and embedded systems. The 8085 microprocessors, a fundamental component of microcomputers, operate using instructions, addressing modes, and control signals to process data efficiently. By exploring its memory interfacing, I/O interfacing, and bus organization, one can comprehend how data is transmitted, stored, and manipulated. Additionally, understanding instruction sets and execution timing aids in programming and optimizing microprocessor-based systems. Learning about control and status signals, data transfer methods, and interfacing peripherals enables effective system design and troubleshooting. This knowledge is essential for developing microprocessor-based applications in robotics, instrumentation, and real-time embedded systems, making it a vital area of study in electronics and computer engineering.

STRUCTURE:

14.1 INTRODUCTION TO MICROCOMPUTERS:

Computer being a general-purpose device, can be used in different varieties of applications – word processing, computation or controlling industrial process are few of them. But microcomputers, as the name implies, are small computers that have limited capacity and are used for specific applications.

Figure (14.1) shows a block diagram of a simple microcomputer consists of a CPU, memory, input and output devices.

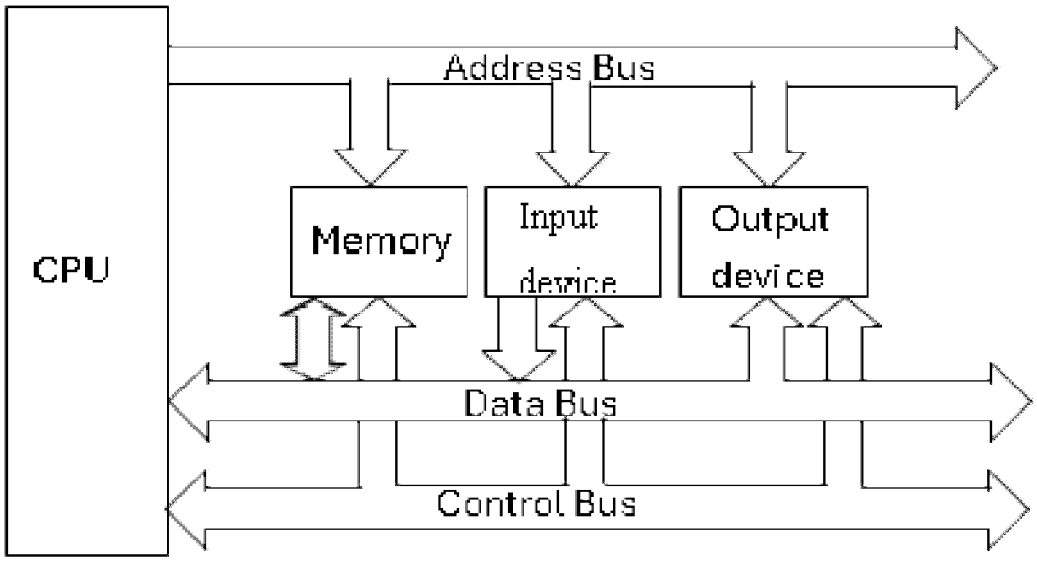


Fig 14.1 Block Diagram of A Simple Microcomputer Consists of a CPU

The function of a microcomputer is receiving data and information and processing it. It means performing arithmetic are logical computations on data, storing the results or data or information in memory and displaying the results of the computation.

The means used to receive data are known as input devices. Some of the input devices are keyboards, switches, mouse etc. A device which performs computations is known as arithmetic logic unit (ALU). In microcomputers, this task, together with control of all devices, is performed by a single chip called microprocessors. Some of the well-known microprocessors are: 8085, 8086, Z80, 650Z etc. Storage of data instructions is accomplished using memories. Cassettes, floppy disks, CD ROMs, semiconductor memories are some examples for memories. Presenting results is done by output devices. Commonly used output devices are monitor (VDU), printers, Seven segment LED displays, LCDs. Figure (13.1) gives a block diagram of a typical microcomputer.

Various peripherals-I/o devices, memories etc – are connected to the microprocessor by means of three types of buses: Address bus, data bus and control bus. A bus is a group of conducting wires.

Over the address bus, the microprocessor transmits the address of the device, with which it desires to communicate (access). The number of devices or memory locations that a CPU can address is determined by the number of address lines. For example, a CPU with 16 address lines can address 216 or 65,536 or 64K (210 = 1024 = 1K) locations. This bus is unidirectional. It means information can flow in one direction only. The CPU sends the address information on these lines.

Over the data bus, CPU can read in or write out the data to various memory or I/O devices. Therefore, this bus is bi-directional. Normally data bus widths of sizes 8, 16, 32 etc. are used.

Control bus may consist of some input lines, some output lines and some bi-directional lines depending upon the processor. CPU sends out some control signals like memory read, I/O write etc that are necessary for memory and I/O devices to work.

14.2 MEMORY:

Memory stores binary information such as instructions and data and provides that information to the microprocessor whenever necessary. To execute programs, the microprocessor reads instructions and data from memory and performs the computing operations in its ALU section. Results are either transferred to the output section for display or stored in memory for later use. The memory block has two sections: Read - Only Memory (ROM) and Read / Write Memory (R/WM), popularly known as Random Access Memory (RAM).

The ROM is used to store programs that d - not need alterations. The monitor program of a single - board microcomputer is generally stored in the ROM. Program stored in the ROM can only be read; they cannot be altered.

The Read / Write memory (R/WM) is also known as user memory. It is used to store user programs and data. The information stored in this memory can be read and altered easily.

14.3 INPUT/OUTPUT:

14.3.1 Input

The input section transfers data and instructions in binary from the outside world to the microprocessor. It includes devices such as keyboards, teletypes, and analog-to-digital converters. Typically, a microcomputer includes a keyboard as an input device. The keyboard has sixteen data keys (o to 9 and A to F) and some additional function keys to perform operations such as storing data and executing programs.

14.3.2 Output:

The output section transfers data from the microprocessor to output devices such as light emitting diodes (LEDs), cathode-ray-tubes (CRTs), printers, magnetic tape, or another computer. Typically, single-board computers include LEDs and seven-segment LEDs as output devices.

14.4 INTERFACING DEVICES 8085 CPU:

Interface is the path for communication between two components. Interfacing is of two types, memory interfacing and I/O interfacing.

14.4.1 Memory Interfacing:

When we are executing any instruction, we need the microprocessor to access the memory for reading instruction codes and the data stored in the memory. For this, both the memory and the microprocessor requires some signals to read from and write to registers. The interfacing process includes some key factors to match with the memory requirements and microprocessor signals. The interfacing circuit therefore should be designed in such a way that it matches the memory signal requirements with the signals of the microprocessor.

14.4.2 IO Interfacing:

There are various communication devices like the keyboard, mouse, printer, etc. So, we need to interface the keyboard and other devices with the microprocessor by using latches and buffers. This type of interfacing is known as I/O interfacing.

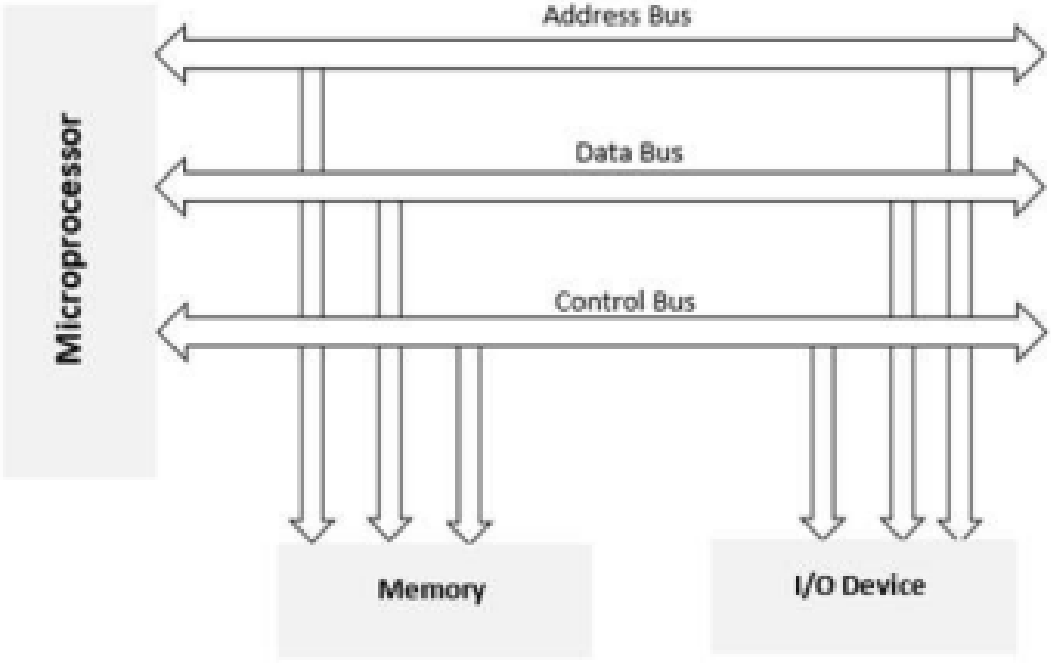


Fig. 14.2 Block Diagram of Memory and I/O Interfacing

8085 Interfacing Pins following is the list of 8085 pins used for interfacing with other devices:

- A15 - A8 (Higher Address Bus)
- AD7 - AD0(Lower Address/Data Bus)
- ALE
- RD
- WR
- READY

Ways of Communication:

Microprocessor with the Outside World? There are two ways of communication in which the microprocessor can connect with the outside world.

- Serial Communication Interface
- Parallel Communication interface

a) Serial Communication Interface:

In this type of communication, the interface gets a single byte of data from the microprocessor and sends it bit by bit to the other system serially and vice-a-versa.

b) Parallel Communication Interface:

In this type of communication, the interface gets a byte of data from the microprocessor and sends it bit by bit to the other systems in simultaneous (or) parallel fashion and vice-a-versa.

14.4.5 Architecture:

The internal logic design of the microprocessor is called its architecture, determines how and what various operations are performed by the microprocessor. As shown in the block diagram (figure 13.2), it consists of arithmetic and logic unit ALU, register section, a timing and control unit and other sections like address / date bus buffers, interrupt control and serial I / O control.

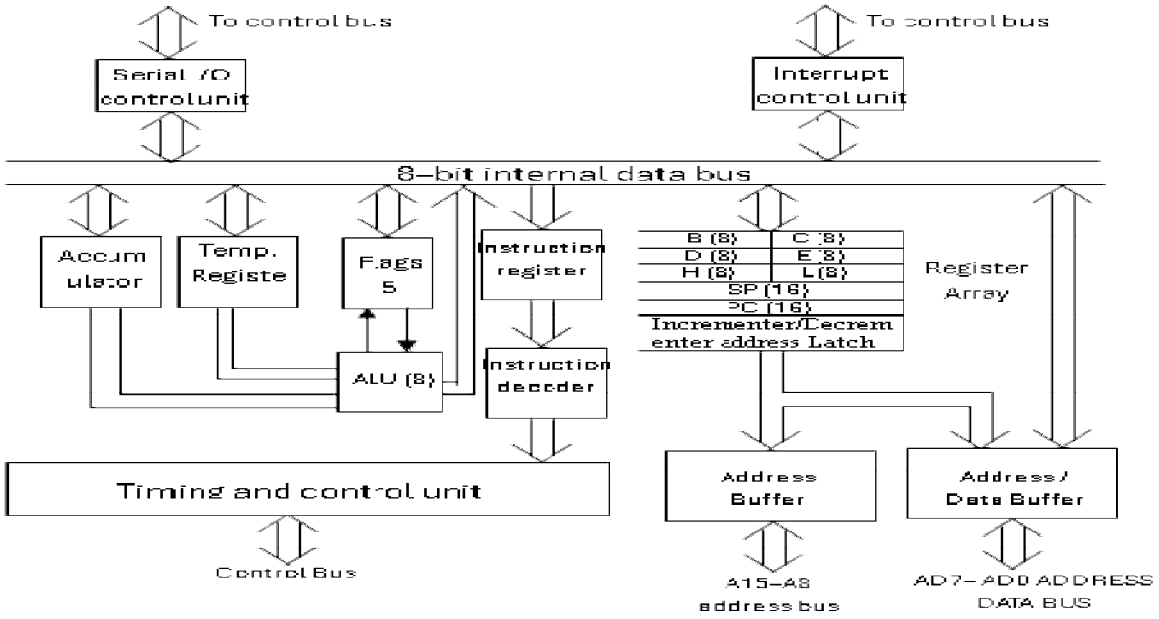


Fig. 14.3: Block Diagram of 8085 Microprocessor

14.5 BUS TIMINGS:

Timing diagram is the display of initiation of read/write and transfer of data operations under the control of 3-status signals IO/M', S1 and S0. Each machine cycle is composed of many clock cycles. Since, the data and instructions, both are stored in the memory, the µP performs fetch operation to read the instruction or data and then execute the instruction. The 3-status signals: IO / M', S1 and S0 are generated at the beginning of each machine cycle. The unique combination of these 3-status signals identifies read or write operation and remain valid for the duration of the cycle. Thus, time taken by any µP to execute one instruction is calculated in terms of the clock period. The execution of instruction always requires read and writes operations to transfer data to or from the µP and memory or I/O devices. Each read/ write operation constitutes one machine cycle. Each machine cycle consists of many clock periods/ cycles, called T-states.

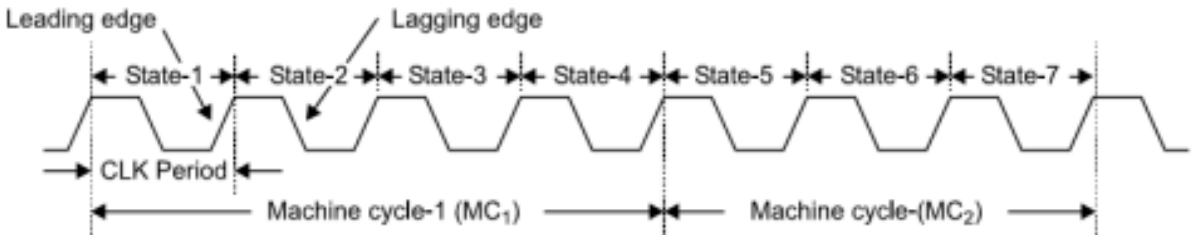


Fig. 14.4: Machine Cycle Showing Clock Periods

Each and every operation inside the microprocessor is under the control of the clock cycle. The clock signal determines the time taken by the microprocessor to execute any instruction. State is defined as the time interval between 2-trailing or leading edges of the clock. Machine cycle is the time required to transfer data to or from memory or I/O devices.

The 8085 microprocessor has 5 basic machine cycles. They are:

- Opcode fetch cycle (4T)
- Memory read cycle (3 T)
- Memory writes cycle (3 T)
- I/O read cycle (3 T)
- I/O write cycle (3 T)

14.6 DEMULTIPLEXING THE ADDRESS BUS:

- From the above description, it becomes obvious that the AD7-AD0 lines serve a dual purpose and that they need to be demultiplexed to get all the information.
- The high order bits of the address remain on the bus for three clock periods. However, the low order bits remain for only one clock period and they would be lost if they are not saved externally. Also, notice that the low order bits of the address disappear when they are needed most.
- To make sure we have the entire address for the full three clock cycles, we will use an external latch to save the value of AD7-AD0 when it is carrying the address bits. We use the ALE signal to enable this latch.
- Given that ALE operates as a pulse during T1, we will be able to latch the address. Then when ALE goes low, the address is saved and the AD7-AD0 lines can be used for their purpose as the bi-directional data lines.

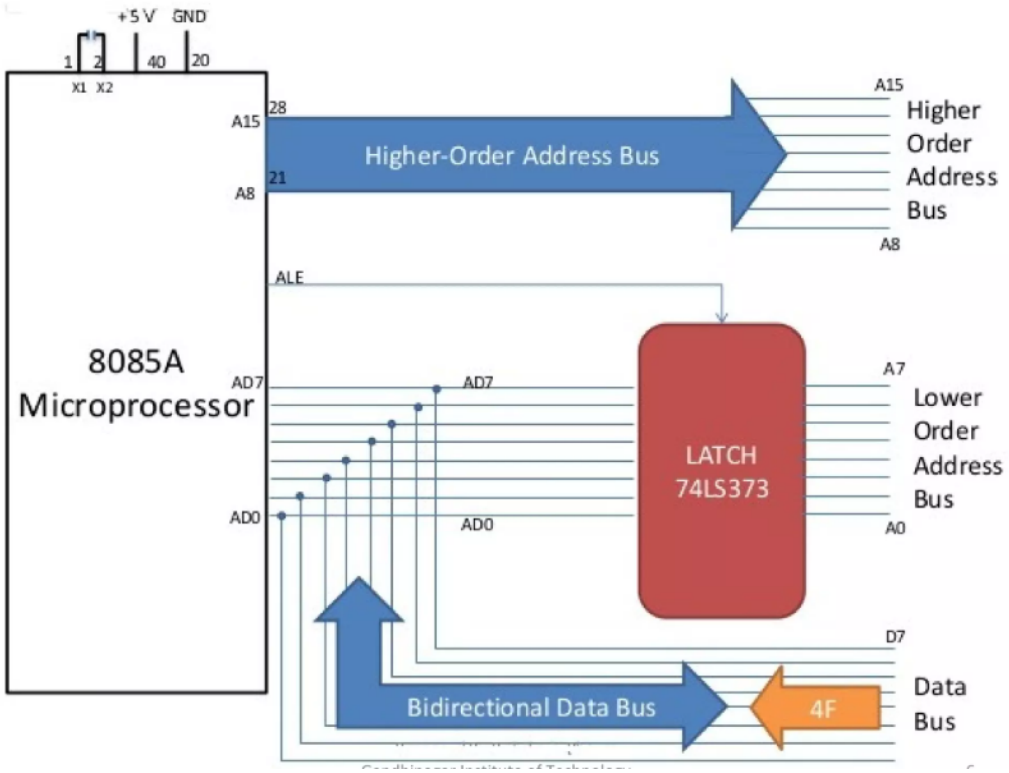


Fig. 14.5: 8085 A Microprocessor Address and Data Bus

14.7 GENERATING CONTROL SIGNALS:

14.7.1 Control and Status Signals:

This group of signals includes two control signals ($\overline{RD}$ and $\overline{WR}$), three status signals (IO/$\overline{M}$), S_1 and S0) to identify the nature of the operation and one special signal (ALE) to indicate the beginning of the operation. These signals are as follows:

- **ALE - Address Latch Enable:** This is a positive going pulse generated every time the 8085 begins an operation (machine cycle): it indicates that the bits on $AD_7 – AD_0$ are address bits. This signal is used primarily to latch the low- order address from the multiplexed bus and generate a separate set of eight address lines, A_7 -A_0.
- **$\overline{RD}$- Read:** This is a Read control signal (active low). This signal indicates that the selected I/O or memory device is to be read and data are available on the data bus.
- **$\overline{WR}$ – Write:** This is a write control signal (active low). This signal indicates that the data on the data bus are to be written into a selected memory or I/O location.
- **IO/$\overline{M}$:** This is a status signal used to differentiate between I/O and memory operations. When it is high, it indicates an I/O operation; when it is low, it indicated a memory operation. This signal is combined with $\overline{RD}$ (Read) and $\overline{WR}$ (write) to generate I/O and memory control signals.
- **S_1 and S_0:** These status signals, similar to IO/$\overline{M}$, can identify various operations, by they are rarely used in small systems.

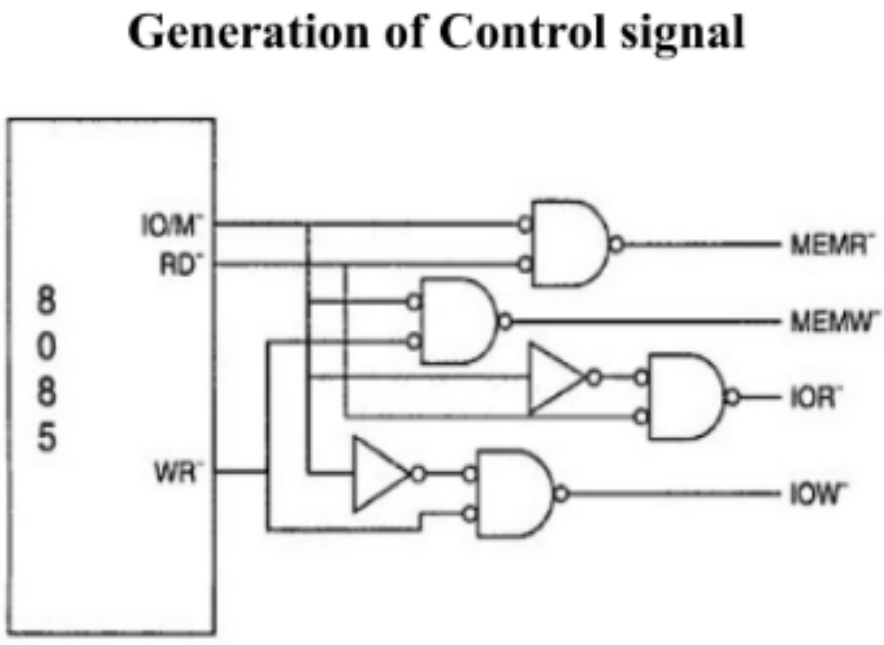


Fig. 14.6: Generation of Control Signal

Truth Table of Control Signal

Operation	Status		
	IO/$\overline{M}$	S1	S0
Opcode Fetch	0	1	1
Memory Read	0	1	0
Memory Write	0	0	1
I/O Read	1	1	0
I/O Write	1	0	1
INTR Acknowledge	1	1	1
Bus Idle	0	0	0

Fig. 14.7: Truth Table of Control Signal

14.8 INSTRUCTION SET:

It is clear from the previous studies that 8085 has a collection of 8-bit and 16-bit registers and 8-bit memory locations that can be used for programming purposes. The collection of these register and memory locations used by the programmer is called programmer's model. The programmer's model of 8085 consists of

i) 8-bit Accumulator

ii) Six 8-bit registers: B, C, D, E, H &L or three 16-bit registers: B-C, D-E & H-L

iii) 16-bit program PC

iv) 16-bit stack pointer SP

v) 8-bit flags register

vi) 64KBmemory locations divided into ROM, RAM and stack areas and

vii) 64 I/O ports.

An instruction is a command to the microprocessor to program a specific task. The entire group of instructions are called its instruction set. The 8085-instruction set has 74 opcodes that result in 246 instructions. Depending on their function, all these instructions can be classified into the following categories:

i) Data transfer group

ii) Arithmetic group

iii) Logical group

iv) Branch group

v) Stack, I/O and Machine-Control Group.

14.8.1 Data transfer group: These instructions copy data from a source into a destination without modifying the contents of the source and without affecting the flags.

8-bit data transfer as

1) MOV destination, source - Copies the contents of source into destination where the source may be

   - An immediate value (e.g.# 35H)
   - An 8-bit register (A, B, C, D, E, H & L)
   - The contents of a memory location whose address is specified in H-L reg. Pair

   The destination may be

   - An 8-bit register (reg.)
   - The contents of a memory location whose address is specified in H-L reg. pair. But both source and destination should not be memory contents, and the destination should not be an immediate value.

Examples:

Between registers:

MOV B, C - data of register C is copied into register B.

Specific data byte into a register or memory

MVI B, 2FH - data 2FH copied into register B

MVI M, 2FH - data 2FH into a memory location whose address is specified by H-L pair (Before this instruction H-L pair should be loaded with a memory address)

Data transfer between memory location and a register

MOV M, B - copy the contents of reg. B into a memory location address by H-L pair.

MOV B, M - copy the contents of memory location address by H-L pair into reg. B.

LDA 16-bit address: Load accumulator directly the contents of memory location whose address is specified in the second and third byte of instruction.

e.g.: LDA 2050 H - Load the contents of memory location 2050H into the acc.

STA 16-bit address: Store accumulator direct.

The accumulator is stored in a memory location whose address is specified with second and third byte of instruction.

STA 30A0H - Store accumulator content directly into memory location whose address is 30A0H.

LDAX rp: Load acc. indirect rp = B – C / D – C.

The contents of memory location whose address in the register pair rp is loaded into the acc.

STAX rp: Store acc. Indirect rp = B – C / D – E.

The contents of acc. is stored in memory, whose address is in the register pair rp.

e.g.: STAX B - Store the contents of acc. Into the memory location addressed by B – C.

14.8.1.1) 16-Bit Data Transfer:

1) **LXI rp, data 16: Load register pair immediately.**

   rp = B – C / D – C | H – C | sp.

   The 16-bit immediate data is loaded into the specified register pair.

   e.g. LXI H, 2450H – Loads 2450H into HL pair such that 50H is in L and 24H in H.

   e.g: LDAX B – This instruction will load the contents of the memory location, whose address is in B – C reg. Pair into acc.

2) **LHLD 16-bit addr.: Load H – L pair direct.**

   The contents of memory location, whose address is specified in second and third byte of instruction is loaded into register L. The contents of the next memory location is loaded into register H.

   e.g: LHLD 3050H – This instruction loads the contents of 3050H into register L and the contents of 3051H into register H.

3) **SHLD 16-bit addr: Store H – L pair direct.**

   The contents of register L is stored into the memory location whose address is specified in the instruction, and the contents of register H is stored in the next memory location.

   e.g: SHLD 3500H – This instruction will store the contents of register L in the memory location 3500H, and the contents of register H is the location 3501H.

4) XCHG: Exchange the contents of H – L with D – E.

The contents of register pair H – L are exchanged with that of the register pair D - E.

e.g. if DE=2050; HL=307F; after the execution of XCHG instruction DE=307F; HL=2050.

14.8.2 Arithmetic Group:

These instructions perform arithmetic operations such as addition, subtraction, increment and decrement.

14.8.2.1) 8-bit Addition and Subtraction: The 8085 can execute two types of addition / subtraction instructions - addition or subtraction without carry and addition or subtraction with carry. Both addition instructions can be used to add the contents of accumulator to the contents of one of the general-purpose registers or the contents of memory locations (address in H - L pair) or all immediate data byte.

a) These instructions assume that the acc. is one of the operands.

b) Place the result in the acc.

c) Do not affect the contents of another operand and

d) Modify all the flags according to the data condition in acc.

e.g.1: ADD B – Adds the contents of register B with the contents of the accumulator and the result is stored in the acc. If a=05H and B-0 7H, execution of ADD B modifies the content of A as 0CH. Reg B content is not altered. As 8085 handles data in binary in the display we see the hexadecimal equivalent of result OCH. Don't expect the decimal answer of 12.

e.g.2: ADD M – The contents of the memory location addressed by the pair is added to the contents of acc. one result is stored in the acc.

e.g.3: ADI 35H – The 8-bit immediate data 35H is added to the contents of the accumulator and the result is stored in acc.

e.g.4: ADC D – The contents of register D and carry flag are added to the contents of acc. and result is stored in the acc.

e.g.5: ADC M – The contents of memory location addressed by H – L and carry are added to the contents of the acc, and the result is stored in the acc.

e.g.6: ACI 35H – 35H with carry are added to the contents of acc.

e.g.7: SUB B – The contents of register B is subtracted from the contents of the acc. and the result is stored in the acc.

e.g.8: SBB M – The contents of the memory location addressed by H – L pair and carry are subtracted from the contents of acc. And the result is stored in the acc.

14.8.2.2) 16-bit Addition:

This instruction adds 16-bit data of H – L register to the 16-bit data of another 16-bit register contents. Result is stored in HL pair. Except carry no other flags are effected.

DAD rp: Add the content of register pair specified to the content of H-L pair and store the result in H-L pair.

rp = B-C / D-E / H-L / sp.

The contents of register pair r_P are added to the contents of H – L pair and the result is stored in H – L pair.

e.g.: DAD D – The contents of register pair D – E are added to the contents of HL pair and the result is stored in H – L pair. If the result is more than 16-bits carry flag is set. No other flags are affected.

14.8.2.3) 8-bit Increment/Decrement:

These instructions increment or decrement the contents of 8-bit registers or the contents of memory location whose address is specified in H – L register pair.

Carry flag is not affected.

e.g.: INR D – The contents of register D is incremented by one.

INR M – The contents of memory location addressed by H – L pair is incremented by one.

DCR C – The contents of register C is decremented by one.

DCR M – The contents of the memory location addressed by H – L pair is decremented by one.

14.8.2.4) 16-bit Increment/Decrement:

1N x rp: Increment register pair contents

rp= B – C / D – E / H – L / sp.

e.g.: 1NX H – The contents of the H – L pair is incremented by one. No flags are affected.

DCX rp: decrement register pair contents.

rp= B– C / D – E / H – L / sp.

e.g.: DCX B – The contents of the B – C pair is decremented by one. No flags are affected.

However, it becomes necessary in some programs to check whether the execution of DCX rp resulted in zero or not. After each decrement by checking the logical OR result of the registers involved one can ascertain it See how this is done in program no. .

14.8.2.5 Decimal Adjusting Data:

DAA: Decimal adjusting accumulator.

DAA instruction is used in the program after an additional instruction. The DAA instruction operates on the result in the accumulator and the result in decimal form. It uses curry and auxiliary curry for decimal adjustment. To use the DAA instructions the two numbers added should be in BCD form.

14.8.3 Logical Instruction:

These instructions perform logical operations such as AND, OR, EX-OR, compare, rotate and complement of data in register or memory.

14.8.3.1 Logical OR, AND, NOT and EX-OR Operations:

a) These instructions implicitly assume that the accumulator is one of the operands

b) Place the result in accumulator.

c) Do not affect the contents of the operand reg. (except NOT operation)

d) Some of the flags are affected according to the result.

e.g.1: ORA D – The contents of the register D is logically OR ed with the contents of the accumulator and the result is stored in the accumulator.

e.g.2: ANI 23H – The data 23H is logically AND ed with the contents of the accumulator and the result is stored in the accumulator.

e.g.3: XRA M – The contents of the memory location addressed by H – L pair is EX – OR ed with the contents of the accumulator and the result is stored in the acc.

e.g.4: CMA – The contents of the accumulator is complemented and the result stored in accumulator.

14.8.3.2 Comparison:

The contents of the accumulator is compared with the contents of the other register or the contents of the memory location addressed by H – L pair or an 8-bit data.

CMP R - Compare register with accumulator

CMP M - Compare memory contents with accumulator.

CPI 8-bit data - Compare immediate data with accumulator

Here the comparison is performed through subtraction. However, neither conten tis modified, but the result of the subtraction is only reflected in the form of conditional flags.

i) If the content of A is less than the data compared data, carry flag set 1.

ii) If the content of A is greater than compared data, carry flag is zero.

iii) If both are equal, zero flag is set to 1.

14.8.3.3 Rotate Instructions:

a) The rotate instructions are register specific, operates only on the accumulator

b) In these instructions, the bit that is shifted out is used as the new bit shifted in

c) There are four rotating instructions to rotate a bit pattern in the accumulator to the left or to the right with or without carry.

RLC - rotate accumulator left

RRC - rotate accumulator right

RAL - rotate all (including carry) left

RAR - rotate all (including carry) right.

14.8.4 Branching Instructions:

The flow of a program proceeds sequentially, from instruction to instruction, unless a control transfer command is executed. The branch instructions allow microprocessors to go to a different memory location, either unconditionally or conditionally (under certain test conditions) and μρ continues executing instructions from that new location.

The branch instructions are classified in three categories:

- Unconditional/conditional jump instructions.
- Unconditional / conditional call and return instructions
- Restart instructions.

14.8.4.1) Jump Instructions:

a) Unconditional Jump:

JMP addr – The program counter is loaded with 16-bit address and the program execution jumps to that address unconditionally.

b) Conditional Jump:

Conditional jump instructions first check for the condition and if the condition is satisfied then only execution jumps to new location. Flag(I), sign flag(S), and purity flag (P).

e.g.1: JC addr – Jump on Carry

JNZ addr – Jump on No zero

JPE addr – Jump on even parity.

14.8.4.2) Call and Return Instructions:

The call instruction is used in the main program to call a subroutine, and the return instruction is used at the end of the subroutine to return to the main program. When a subroutine is called, the contents of the program counter, which is the address of the instruction following the call instruction, is stored in the stack and the program execution is transferred to the subroutine address. When the return instruction is executed at the end of the subroutine, the memory address stored in the stack is retrieved, and the sequence of execution is resumed in the main program.

a) Unconditional Call:

CALL addr – unconditional call

b) Conditional Call:

CC addr – call if the carry flag set

CNC addr – call of the carry flag not set

CNF addr – call on No zero

a) Unconditional return

RET – unconditional return

b) Conditional return

RC – Return on carry

RNC – Return on No carry

RP – Return on positive

14.8.4.3) Restart Instructions:

Restart is a one-byte CALL instruction. Because these instructions have specific call addresses for specific restart instructions as specified bellow. The program jumps to the instruction starting at restart location.

Restart Instruction Location

RST 0 0000H

RST 1 0008H

RST 2 0010H

RST 3 0018H

RST 4 0020H

RST 5 0028H

RST 6 0030H

RST 7 0038H

One interesting use of RST instruction is to create break points. Finding mistakes (bugs) in a big program can be made simple by inserting break points at desired places. The program terminates when RST instruction is executed. If there are no bugs up to that point, the break point can be shifted to some other place. The process of removing bugs is called debugging. Instead of HALT instruction one may use Break point for the safe transfer of control to monitor program.

14.8.5 Stack, I/O and Machine-Control Group:

a) Stack Instructions:

PUSH rp – PUSH register pair onto stack

rp = B – C / D – C / H – C / psw.

The content of the specified register pair is pushed onto stack after the stack pointer is decremented by two.

e.g.: PUSH B – The contents of register pair B – C is pushed on to stack.

POP rp – POP off stack to register pair.

rp = B – C / D – C | H – C | psw.

The top two elements of the stack are POPed into the specified register pair. The stack pointer is incremented by two.

e.g.: POP D – The contents of stack top POPed into register pair D – E.

In some programs the available internal registers may not be sufficient to write the program. However, some of the registers may not be needed immediately. In such situations the contents of immediately not used register contents are pushed on to the stack. These registers are used for the present need and once the purpose is served, they are reloaded with their earlier values with POP instruction. If a series of push instructions like PUSH B, PUSH D, PUSH H are used, to restore their original contents the POP instructions must be in the order POP H, POP D, POP B as stack operations follow first in last out policy.

b) I/O Instructions:

IN port-addr – Input to accumulator from input port

The data available on the port is moved into the accumulator. The port address is 8-bit.OUT port-addr – Output from accumulator to output port. The contents of the accumulator is moved to the port whose 8-bit address is specified in the instruction

External devices with suitable interfacing can be connected to input and output ports. The power of microprocessors lies in its ability communicating with external devices. Industrial digital control systems, Robots and other smart devices work under microprocessor control. All together

256 I/O devices can be connected. The devices which use IN and OUT instruction for data transfer are said to be operating in I/o mapped I/o mode. If more than 256 devices are to be connected, Memory mapped I/O mode can be used. In this mode each device is assigned with a 16-bit address.

c) Machine Control Instructions:

These instructions are used to control microprocessor operation.

e.g.: HLT – Halt

The execution of the HLT instruction stops the microprocessor.

8085 Instruction set

Data transfer group

MOV r1,r2 Cycles:1 States:4 Flags: none

MOV r, M Cycles:2 States:7 Flags: none

MOV	B,B 40	MOV	C,B 48	MOV	D,B 50	MOV	E,B 58
	B,C 41		C,C 49		D,C 51		E,C 59
	B,D 42		C,D 4A		D,D 52		E,D 5A
	B,E 43		C,E 4B		D,E 53		E,E 5B
	B,H 44		C,H 4C		D,H 54		E,H 5C
	B,L 45		C,L 4D		D,L 55		E,L 5D
MOV	H,B 60	MOV	L,B 68	MOV	M,B 70	MOV	A,B 78
	H,C 61		L,C 69		M,C 71		A,C 79
	H,D 62		L,D 6A		M,D 72		A,D 7A
	H,E 63		L,E 6B		M,E 73		A,E 7B
	H,H 64		L,H 6C		M,H 74		A,H 7C
	H,L 65		L,L 6D		M,L 75		A,L 7D

MVI r, data Cycles:2 States:7 Flags: none

MVI M, data Cycles:3 States:10 Flags: none

Move immediate	Load immediate
MVI A,data 3E	LXI B,data 01
B,data 06	D,data 11
C,data 0E	H,data 21
D,data 16	SP,data 31
E,data 1E	

Load / Store Opcode Cycles States Flags

LDAX B 0A 2 7 none

LDAX D 1A 2 7 none

LHLD addr 2A 5 16 none

LDA addr 3A 4 13 none

STAX B 02 2 7 none

STAX D 12 2 7 none

Branch CONTROL Instructions

	Cycles	States	Flags
JMP addr	C3	3	10 none
JNZ addr	C2	2/3	7/10 none
JZ addr	CA	2/3	7/10 none
JNC addr	D2	2/3	7/10 none
JC addr	DA	2/3	7/10 none
JPO addr	E2	2/3	7/10 none
JPE addr	EA	2/3	7/10 none
JP addr	F2	2/3	7/10 none
JM addr	FA	2/3	7/10 none
PCHL	E9	1	6 none

	Cycles	States	Flags
CALL addr	CD	5	18 none
CNZ addr	C4	2/5	9/18 none
CZ addr	CC	2/5	9/18 none
CNC addr	D4	2/5	9/18 none
CC addr	DC	2/5	9/18 none
CPO addr	E4	2/5	9/18 none
CPE addr	EC	2/5	9/18 none
CP addr	F4	2/5	9/18 none
CM addr	FC	2/5	9/18 none

RETURN Cycles States Flags

RET	C9	3	10 none
RNZ	C0	1/3	6/12 none
RZ	C8	1/3	6/12 none
RNC	D0	1/3	6/12 none
RC	D8	1/3	6/12 none
RPO	E0	1/3	6/12 none
RPE	E8	1/3	6/12 none
RP	F0	1/3	6/12 none
RM	F8	1/3	6/12 none

RESTART Cycles States Flags

RST 0	C7	3	12 none
RST 1	CF	3	12 none
RST 2	D7	3	12 none
RST 3	DF	3	12 none
RST 4	E7	3	12 none
RST 5	EF	3	12 none
RST 6	F7	3	12 none
RST 7	F8	3	12 none

STACK operation Cycles States Flags

PUSH B	C5	3	12 none
PUSH D	D5	3	12 none
PUSH H	E5	3	12 none
PUSH PSW	F5	3	12 none
POP B	C1	3	12 none
POP D	D1	3	12 none
POP H	E1	3	12 none
POP PSW	F1	3	12 none
XTRL	E3	5	16 none
SPHL	F9	1	6 none

I / O and Machine

INPUT / OUTPUT Cycles States Flags

OUT data	D3	3	10 none
IN data	DB	3	10 none

CONTROL

DI	F3	1	4 none
EI	FB	1	4 none
NOP	00	1	4 none
HLT	76	1	5 none
SIM	30	1	4 none
RIM	20	1	4 none

Arithmetic group

	Cycles	States	Flags
ADD r, ADC r, SUB r, SBB r	1	4	All
ADD M, ADC M, SUB M, SBB M	2	7	All
ADI data, ACI data, SUI data, SBI data	2	7	All
INR r, DCR r,	1	4	Z, S, P, AC
INR M, DCR M	3	10	Z, S, P, AC
INX rp, DCX rp	1	6	none
DAD rp,	3	10	CY
DAA	1	4	All

ADD	ADC	SUB	SBB
B 80	B 88	B 90	B 98
C 81	C 89	C 91	C 99
D 82	D 8A	D 92	D 9A
E 83	E 8B	E 93	E 9B
H 84	H 8C	H 94	H 9C
L 85	L 8D	L 95	L 9D
M 86	M 8E	M 96	M 9E
A 87	A 8F	A 97	A 9F

INR	DCR	INX	DCX		
A 3C	A 3D	B 03	B 0B	ADI data C6	DAA 27
B 04	B 05	D 13	D 1B	ACI data CE	DAD B 09
C 0C	C 0D	H 23	H 2B	SUI data D6	DAD D 19
D 14	D 15	SP 33	SP 3B	SBI data DE	DAD H 29
E 1C	E 1D				
H 24	H 25				
L 2C	L 2D				
M 34	M 35				

Logical group

	Cycles	States	Flags
ANA r, ORA r, XRA r, CMP r	1	4	All
ANA M, ORA M,XRA M, CMP M	2	7	All
ANI data, ORI data, XRI data, CPI data	2	7	All
RLC, RRC, RAL, RAR	1	4	CY
CMA	1	4	none
CMC, STC	1	4	CY

ANA	XRA	ORA	CMP
B A0	B A8	B B0	B B8
C A1	C A9	C B1	C B9
D A2	D AA	D B2	D BA
E A3	E AB	E B3	E BB
H A4	H AC	H B4	H BC
L A5	L AD	L B5	L BD
M A6	M AE	M B6	MBE
A A7	A AF	A B7	A BF

ANI data E6	RLC 07	DAA 27
XRI data EE	RRC 0F	CAN 2F
ORI data F6	RAL 17	STC 37
CPI data FE	RAR 1F	CMC 3F

14.9 ADDRESSING MODES:

A microprocessor system operates under the control by a series of instructions that make up a program. An instruction is a command to the microprocessor to perform a specific operation on certain data.

Let us consider the following instructions:

a) MOV C, B – This instruction moves the contents of register B into register C.

b) LDA 2050H – Loads the contents of the memory location 2050H into accumulator.

Each instruction has two parts: first part indicates the task to be performed and it is called Mneumonic (code in English characters that stands for binary op code) The second part is calledoperand.

In the first example MOV specifies the tasks of instruction (to move the content of one register to another)and is therefore called op code. In the remaining part of the instruction two data registers (source , destination) called operands are specified

In the second instruction, LDA is the op code(in this instruction Acc is one of the operands). A 16 bit number is specified for operand . This number stands for a memory location and represents the second operand. According to INTEL notation if 2050 location is intended the instruction has to be entered in the order as op code, lower byte of address, upper byte of address i.e 3A 50 20 . Here 3A is the hexadecimal equivalent for the mneumonic of LDA op code.

Above examples indicates that each instruction specifies an operation to be performed on certain data. There are certain techniques by which the address of the data to be operated upon may be specified. The various ways of specifying the operand are called the addressing modes.

There are five addressing modes

i) Direct addressing mode ii) Register addressing mode iii) Register indirect addressing mode

iv) Immediate addressing mode v) Implicit addressing mode

14.9.1 Direct addressing modes: In this mode, the address of the operand is specified directly within the second and third byte of the instruction itself.

Examples:

a) STA 1850H: Store the contents of accumulator into memory location 1850H.

b) LHLD 2015H: Load the contents of memory location 2015H into register L and the contents 2016H into register H.

14.9.2 Register addressing mode: In this addressing mode, the source or destination or both operands are located in the microprocessor registers

Examples:

a) MOV A, B – The contents of register B is copied into register A

b) ADD B – The contents of register B is added to the contents of accumulation and the result stored in accumulator.

14.9.3 Register indirect addressing mode: In this addressing mode, the address of the operand is specified by a register pair.

Examples:

a) MOV A, M – The contents of the memory location whose address is specified in H-L register pair is copied into accumulator.

b) LDAX B – The contents of the memory location whose address is specified by B-C register pair is copied into register A.

14.9.4 Immediate addressing mode: In this addressing mode, the operand is specified in the instruction itself.

Examples:

1) MVI B, 25H – The 8-bit data 25H is moved in to register B

2) LXI H, OF5AH – The 16-bit data OF5AH is moved into H-L register pair such that 5AH is moved into register L and OFH is moved into register H.

14.9.5 Implicit addressing mode: Some instructions do not require the address of the operand. The operands are implicit in the instruction itself. Most of these instructions operate on the contents of the accumulator.

Examples:

a) CMA – Compliment the contents of accumulator.

b) STC – Set carry flag.

14.10 SUMMARY:

Microcomputers, small yet powerful computing devices, use microprocessors like 8085 to perform various tasks, including data processing, control operations, and communication. The 8085 microprocessor consists of an ALU, registers, control unit, and buses for efficient data transfer. It operates using different addressing modes, such as direct, register, and immediate addressing, to execute instructions. Memory is classified into ROM for permanent storage and RAM for temporary data handling. Input and output devices facilitate interaction with

external systems, while interfacing techniques enable communication with peripherals. Control and status signals, along with bus timings, govern data flow and execution cycles. Understanding these concepts helps in designing and optimizing microprocessor-based applications in automation, robotics, and embedded systems, making microcomputers essential in modern technology.

14.11 TECHNICAL TERMS:

Microprocessor, Addressing Modes, Bus System, Instruction Set, Memory Interfacing

14.12 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) 1.What are the main components of a microcomputer, and how do they interact?
2) 2.Explain the different types of addressing modes used in a microprocessor with examples.
3) 3.How does the 8085 microprocessor communicate with memory and I/O devices?

Short Answer Questions:

1) What is the purpose of the address, data, and control buses in a microcomputer system?
2) Differentiate between ROM and RAM in terms of functionality and usage in a microcomputer.
3) What are control signals in the 8085 microprocessor, and how do they regulate data flow?

14.13 SUGGESTED READINGS:

1) Ramesh S. Gaonkar, Microprocessor Architecture, Programming, and Applications with the 8085, Prentice Hall.
2) B. Ram, Fundamentals of Microprocessors and Microcontrollers, Dhanpat Rai Publications.
3) Douglas V. Hall, Microprocessors and Interfacing: Programming and Hardware, McGraw-Hill.
4) P. Malvino, Digital Computer Electronics, McGraw-Hill.
5) William Stallings, Computer Organization and Architecture, Pearson Education.

Prof. Ch. Linga Raju

LESSON-15

MICROCONTROLLERS

15.0 AIM AND OBJECTIVES:

The aim of this study is to understand the architecture, functionality, and applications of microcontrollers, with a focus on the 8051 microcontrollers. The objective is to explore the types of memory used in microcontrollers, including RAM, EEPROM, and Flash Memory, and analyze their role in data storage and processing. Additionally, the study aims to examine the control unit, CPU operations, oscillator functions, buses, and interrupt handling mechanisms in the 8051 microcontrollers. A detailed understanding of the pin configuration and register functionalities will also be covered to highlight their significance in embedded systems. By achieving these objectives, this study will provide a comprehensive understanding of microcontrollers and their applications in various electronic and automation systems, contributing to efficient hardware design and development in embedded technology.

STRUCTURE:

15.1 ILLUSTRATIVE PROGRAMS:

Example 1: To give a better idea of how the data transfer group of instructions are used in programs, we will now write a simple program to input a value from the keyboard, store this value in memory location and display it. Here we assume that the keyboard is connected through input port 1 whose address is 05 H, display is connected through output port of address 02H and memory location address is 1050H.

Solution:

Algorithm for the above program:

1) Input a value from key board connected to the port at address 05H.
2) Store the value in memory location 1050H.
3) Output the value to a display connected to the port at address 02H.

Assembly Language Program:

TABLE 15.1

Label	Mnemonics		Comments
	Op Code	Operand	
	IN	05H	; Input from port 05H
	STA	1050H	; Store in memory
	OUT	02H	; Send it to display
	HLT		; End of the program

Example 2:

To use the arithmetic group and branching group of instructions and to have better concept about counters and loops, we will write a program to find the sum of a series of numbers stored in the memory locations, starting from 0850H to 0856H. Store the 16-bit sum in 0860H and 0861H locations.

In this example, we use the concept of looping. Looping is a technique used to instruct the processor to execute certain instructions repeatedly. The number of repetitions are specified in a counter.

Flow Chart:

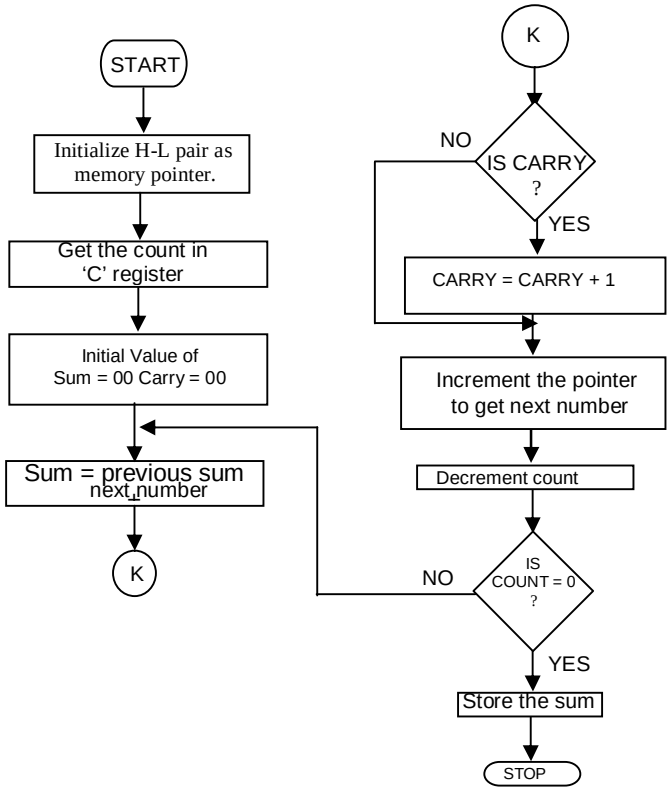


Fig. 15.1: Flow Chart

Algorithm:

1) Take initial sum as zero
2) Add the first number to the previous sum (initial sum)
3) Reduce the number of items by one.
4) Add the next number of the given series to the previous sum
5) Reduce the number of items by one
6) Repeat steps (4) and (5) till all the numbers are added
7) Final sum is the result
8) Stop the process.

Program:

TABLE 15.2

Label	Mnemonics		Comments
	Op code	Operand	
Repeat:	LXI	H, 0850	; Initialize the memory
	MVI	C, 07	; Get the count
	MVI	A, 00	; Clear A register
Next:	MVI	B, 00	; Clear B register
	ADD	M	; Add the number to previous sum
	JNC	Next	; Jump to 'next' if there is no carry
	INR	B	; Register carry
	INX	H	; Increment the memory pointer
	DCR	C	; Decrements the counter
	JNZ	Repeat	; Jump to 'repeat' if counter is not zero.
	STA	0861	; Store the sum
	MOV	A, B	; Save the carry
	STA	0860	; Store the carry
	HLT		; End of the program

The above program can be executed on a microprocessor kit, after converting the program into hex code and assigning one address to each byte of the program as in the previous example.

Example 3: 8085 microprocessor does not have a multiply instruction. So, when we want to multiply two numbers, we have to write a program for this. There are various algorithms for the multiplication of two numbers. Here we adopt a simple repeated addition algorithm .The method is to initialize a register with zero value and call it RESULT and go on adding one number to the RESULT second number times. For example if 5 is to be multiplied by 6 it can be done adding 6 to itself 5 times or adding 6 to itself 5 of times.

Algorithm:

1) Take the multiplicand (I number)
2) Take the initial sum as zero
3) Add the multiplicand to the previous sum
4) Decrement the multiplier
5) Repeat the steps (3) and (4) till the multiplier becomes zero
6) The final sum gives the products of two numbers
7) Stop the process

Flow Chart:

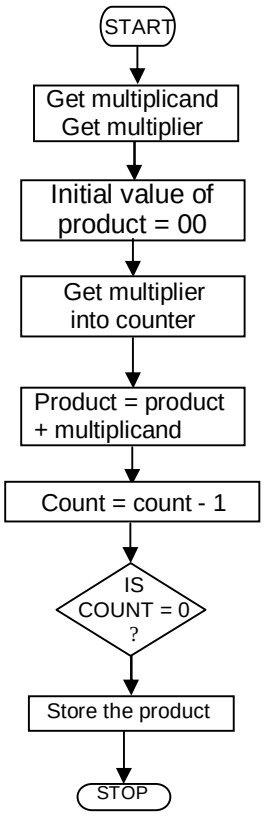


Fig. 15.2

Let the numbers (multiplicand, multiplier) be stored at memory locations with the address 0850, 0851; let the result (product) be of 16-bit be stored at locations with the address 0852, 0853 (for higher and lower order byte of the product). Product is found by using repetitive addition.

Program:

TABLE 15.3

Label	Mnemonics		Comments
	Op code	**Operand**	
Next:	LXI	H, 0850	; Initialize the memory pointer
	MOV	B, M	; Get the multiplicand
	INX	H	; Increment the pointer
Loop:	MOV	C, M	; Get the multiplier
	SUB	A	; Make the initial product as 00
	MOV	D, A	; Clear reg 'D' for carry
	ADD	B	; Add the multiplicand
	JNC	Loop	; Check for carry
	INR	D	; Register the carry
	DCR	C	; Decrement the multiplier
	JNZ	Next	; If (C) $\neq$ 0, go to 'next'
	INX	H	; Increment the pointer
	MOV	M, D	; Save the carry
	INX	H	; Increment the pointer
	MOV	M, A	; Save the sum
	HLT		; End of the program

EXAMPLE 4: To have the better concept of using compare instruction in finding the smaller or bigger or equality of two given numbers we write a program to find the largest number among the given set of n- numbers.

Let the list starts from a memory location with the address 0850H and let the largest number be stored at a memory location 0870H

Algorithm:

1) Take the first number from the given series
2) Compare this number, with the next number of given series
3) If the first number is larger than the next number then retain the first number
4) If the first number is smaller than the next number then retain the next number
5) Decrement the number of comparisons
6) Repeat the process of comparison till the largest number in the given list is found.
7) Stop the process.

Flow Chart:

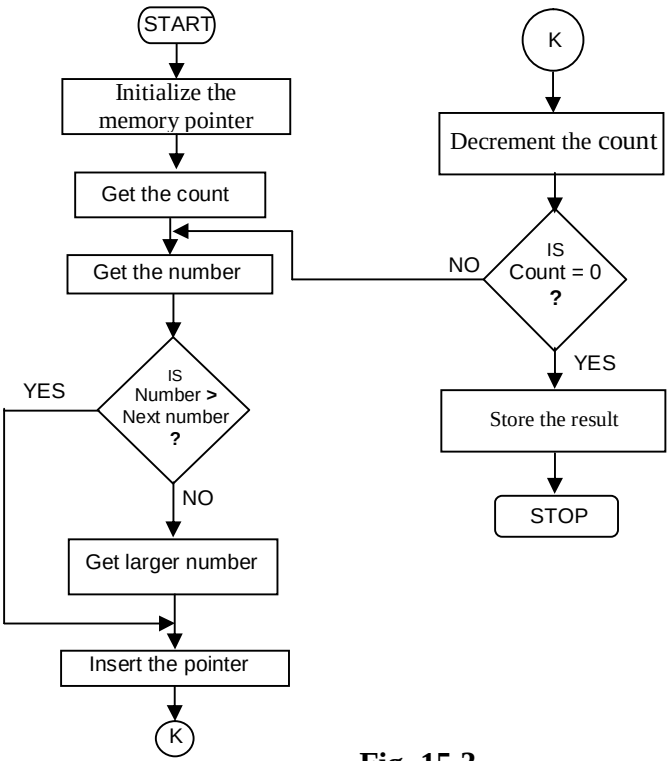


Fig. 15.3

PROGRAM:

TABLE 15.4

Label	Mnemonics		Comments
	Op Code	**Operand**	
Next:	MVI	C, n – 1	; Get the count
	LXI	H, 0850	; Initialize the memory pointer
	MOV	A, M	; Get the number
	INX	H	; Increment the pointer
Loop:	CMP	M	; Compare with memory for largeness
	JNC	Loop	; If larger, jump to 'loop'
	MOV	A, M	; If smaller change with memory
	DCR	C	; Reduce the counter
	JNZ	Next	; If comparisons are not over, go to 'next'
	STA	0870	; Store the largest
	HLT		; End of the program

Here, after converting the AL program into its hex codes and assigning with suitable addresses, the actual series of data items is to be to loaded into the memory, whose starting address is specified in the second instruction of above program. The method of loading the data items into the memory is same as that of loading the program. As an example, suppose the number of items, n is 8, and let the data items are:

0A, 29, 02, 0B, 0E, 22, 39, 49

After executing the program using the above data, the largest number i.e. OE can be viewed at the specified memory location 0870.

TABLE 15.5

Address	Data
0850	0A
0851	29
0852	02
0853	0B
0854	0E
0855	22
0856	39
0857	49

Result	
Address	**Data**
0870	0E

Note that the same program can be used to find the smallest number simply using JC instruction instead of JNC in the above program.

Example 5: In most of the microprocessor systems, the key board input is in BCD format (in BCD format a decimal digit, 0 to 9, is represented by its 4-bit binary equivalent). But the processing of data inside of the microprocessor is performed in binary format. Hence there is a need to convert BCD number into its equivalent binary value. In doing so we appreciate the usage of logical AND ROTATE instructions.

Algorithm:

1) Take the given 8-bit packed BCD number,
2) Separate the number into two 4-bit unpacked BCD digits BCD, and BCD_2
3) Convert each digit into its binary value according to its position.
4) Add both binary numbers to obtain the binary equivalent of the given BCD number
5) Stop the process.

Flow Chart:

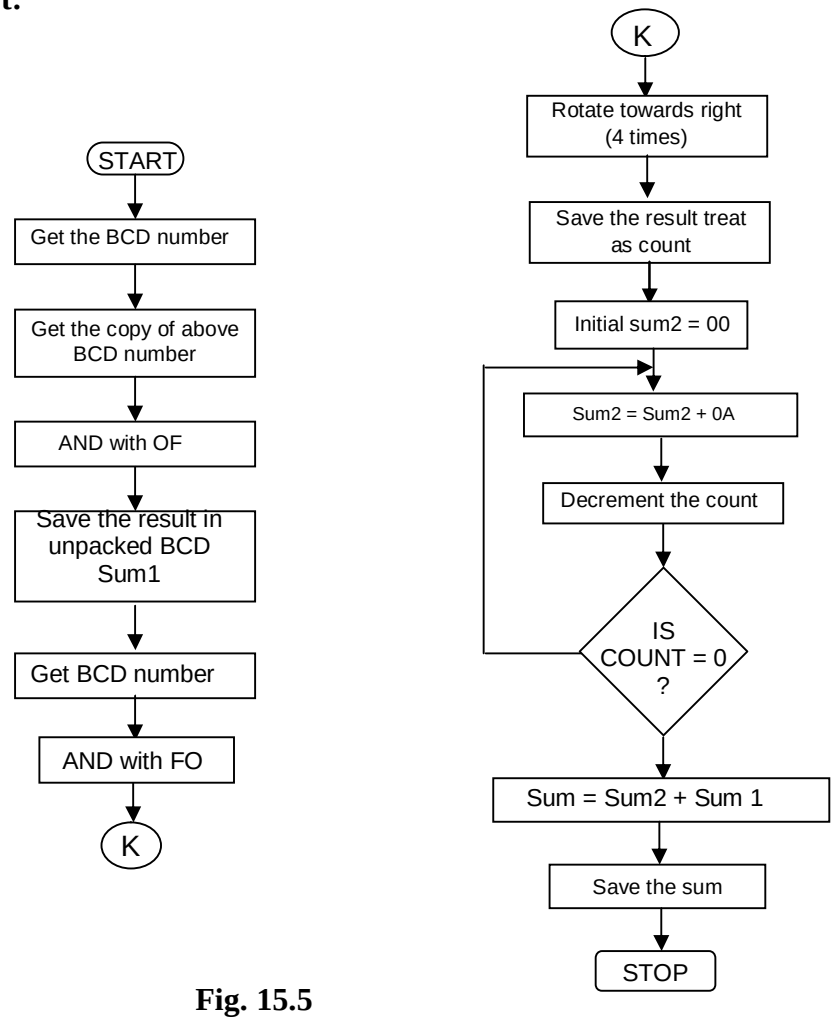


Fig. 15.5

In this example we assume that the packed BCD number stored in memory location 1050H and we store the result in 1060H.

The conversion of a BCD number into its equivalent binary number employs the principle of positional weighting in a given number. For example in a number 96_{10}, the digit 9 represents 90 based on its second position from the right. So while converting 96_{BCD} into its binary equivalent requires the multiplication of 9 by 10 and adding the first digit of 6.

TABLE 15.6

Label	Mnemonics		Comments
	Op code	**Operand**	
	LDA	1005H	; Get packed BCD number
	ANI	0FH	; Mask most significant four bits
	MOV	C, A	; Save unpacked BCD, in regC
	LDA	1050H	; Get BCD again
AGAIN:	ANI	FOH	; Mask least significant four bits
	RRC		
	RRC		
	RRC		
	RRC		
	MOV	D, A	; Save BCD_2 in D
	XRA	A	
	MVI	E, 0AH	
	ADE	E	
	DCR	D	
	JNZ		
	ADD	C	
	STA	1060H	
	HLT		

Example 6: Accurate time delays between two events can be generated using assembly language programming. For example, clock generators, digital clocks, flashing of lights etc. require these time delays. The process of generating time delays using software instructions is more flexible than the generation of time delays using hardware. To illustrate this, we writea program to generate time delay by using counters.

ALGORITHM:

1) Load a register with required delay.
2) Decrement the register.
3) Repeat it till register becomes zero.
4) Get out of the loop.

Flow Chart:

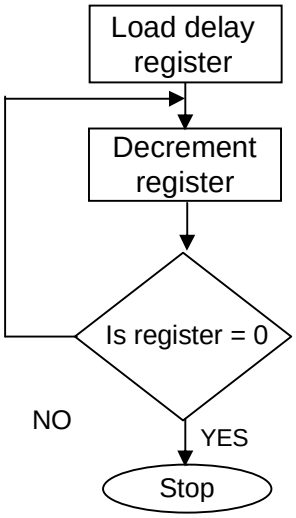


Fig. 15.6

TABLE 15.7

Label	Mnemonics		Comments
	Op code	**Operand**	
BACK:	MVI	C, N	; Load RegC with number N
	DCR	C	; Decrement C
	JNZ	BACK	;If C ≠ 0 Jump to BACK to decrement
	HLT		

In the above program, the number N in register C determines the duration of the time delay. For the calculation of N, we should know the number of states required for the execution of each instruction and the number of times the instruction is executed. The number of states required can be found from the data supplied by the microprocessor manufacturer (see the instruction set given) and the number of times the instruction is executed is determined by the program as shown below.

Instruction States Total number of states

MVI C, N 7 7 x 1

DCRC 4 4 x N

JNZ 10/7 10(N-1)+7 x 1

HLT 5 5 x 1

$\therefore$ Total states = 7x1 + 4N + 10(N-1) + 7x1 + 5x1

= 7+4N+10N-10+7+5

= 14N + 4

Time for one state if the processor is operative with 2MHz clock = .5μs

Therefore Time delay = $.5*10^{-6}$ (N-1)

Using the above relation, the N value can be calculated for the required time delay.

Example 7:

To have the concept of subroutine, we write an assembly language program to switch ON and OFF an LED connected to an out port of the address 05H.

To visualize the ON and OFF states of LED we include a delay program in between these ON and OFF states.

Algorithm:

1) Switch ON the LED
2) Insert delay
3) Switch OFF the LED
4) Insert delay
5) Loop continuously

Flow Chart:

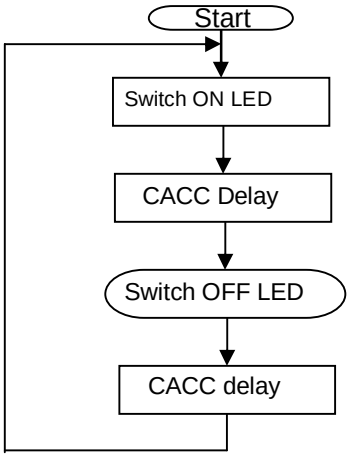


Fig. 15.7

Program:

START: MVI A, FFH	Delay
OUT O5H	MVI B, FFH
CALL DELAY	Back: DCR B
MVI A, 00H	JNZ Back
OUT 05H	RET
CALL DELAY	
JMP START	

Table 1.8a **Table 1.8b**

15.2 WRITING ASSEMBLY LANGUAGE PROGRAMS:

A Simple Program:

ABSTRACT: This Program adds two 8-bit words in the memory.

: Locations called NUM1 and NUM2.

: The result is stored in the memory location called RESULT.

: If there was a carry from the addition, it will be stored as 0000 0001 in the location CARRY.

ALGORITHM:

get NUM 1

add NUM 2

Put sum into memory at SUM

Position carry in LSB of byte registers

Mask off upper seven bits

Store the result in the carry location

REGISTERS Uses CS, DS, AX

```
DATA        SEGMENT
            NUM DB      16h ;   first number stored here
            NUM DB      19h ;   second number stored here
            RESULT      DB  ?   ; put sum here
            CARRY       DB  ?   ; put any carry here
DATA ENDS
CODE SEGMENT
        ASSUME      CS: CODE,   DS :    DATA
START       :MOV    AX,     DATA        ;       Initialize data segment
             MOV    DS,      AX         ;       register
             MOV    AL,     NUM1        ;       Get the first number
            ADD     AL,     NUM2        ;       add it to 2nd number
            MOV     RESULT, AL          ;       Store the Result
            RCL     AL, 01              ;       Rotate carry into LSB
            AND     AL, 00000001B       ;       mask out all but LSB
            MOV     CARRY, AL           ;       store the carry result
            INT 3h
CODE ENDS
        END START
```

This program contains, certain additional mnemonics, in addition to the instructions you have studied so far. These are called as **assembler directives or pseudo operations.** These are the directions for the assembler. Their meaning is valid only till the assembly time. There is no code generated for them. We shall now study the complete program step by step.

15.3 LOOPING:

The programming technique used to instruct the microprocessor to repeat tasks is called looping. This task is accomplished by using jump instructions.

Classification of Loops:

1) Continuous Loop
2) Unconditional Loop

Continuous Loop:

Repeats a task continuously. A continuous loop is set up by using the unconditional jump instruction. A program with a continuous loop does not stop repeating the tasks until the system is reset.

Conditional Loop:

A conditional loop is set up by a conditional jump instructions. These instructions check flags (Z, CY, P, S) and repeat the tasks if the conditions are satisfied. These loops include counting and indexing.

Conditional Loop and Counter:

- A counter is a typical application of the conditional loop.
- A microprocessor needs a counter, flag to accomplish the looping task.
- Counter is set up by loading an appropriate count in a register.
- Counting is performed by either increment or decrement the counter.
- Loop is set up by a conditional jump instruction.
- End of counting is indicated by a flag.

Conditional Loop, Counter and Indexing:

Another type of loop which includes counter and indexing.

15.4 COUNTING AND INDEXING:

15.4.1 Counting

This technique allows programmer to count how many times the instruction/set of instructions are executed.

15.4.2 Indexing:

Pointing of referencing objects with sequential numbers. Data bytes are stored in memory locations and those data bytes are referred to by their memory locations.

Example:

Steps to add ten bytes of data stored in memory locations starting ata given location and display the sum.

The microprocessor needs

- A counter to count 10 data bytes.
- An index or a memory pointer to locate where data bytes are stored.
- To transfer data from a memory location to the microprocessor (ALU)
- To perform addition
- Registers for temporary storage of partial answers
- A flag to indicate the completion of the stack
- To store or output the result.

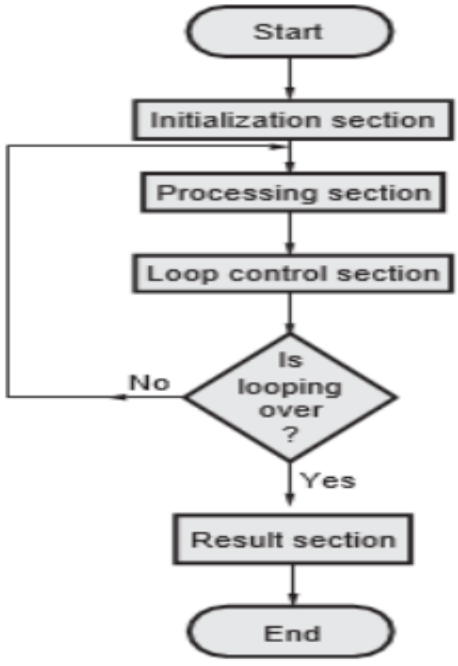


Fig. 15.8: Looping Flow Chart

1) The initialization section establishes the starting values of loop counters for counting how many times loop is executed, Address registers for indexing which give pointers to memory locations and other variables.

2) The actual data manipulation occurs in the processing section. This is the section which does the work.
3) The loop control section updates counters, indices (pointers) for the next iteration.
4) The result section analyzes and stores the results.

The processor executes initialization section and result section only once, while it may execute processing section and loop control section many times. Thus, the execution time of the loop will be mainly dependent on the execution time of the processing section and loop control section. The flowchart 1 shows typical program loop. The processing section in this flowchart is always executed at least once. If you position of the processing and loop control section then it is possible that the processing section may not be executed at all, if necessary

15.5 COUNTERS AND TIMING DELAYS:

In real time applications, such as traffic light control, digital clock, process control, serial communication, it is important to keep a track with time. For example, in traffic light control applications, it is necessary to give time delays between two transitions. These delays are in few seconds and can be generated with the help of executing group of instructions number of times. This software timers are also called time delays or software delays. Let us see how to implement these time delays or software delays.

As you know microprocessor system consists of two basic components, Hardware and software. The software component controls and operates the hardware to get the desired output with the help of instructions. To execute these instructions, the microprocessor takes fix time as per the instruction, since it is driven by constant frequency clock. This makes it possible to introduce delay for specific time between two events. In the following section we will see different delay implementation techniques.

15.5.1 Timer Delay Using NOP Instruction:

NOP instruction does nothing but takes 4T states of processor time to execute. So by executing NOP instruction in between two instructions we can get delay of 4 T-state

$$\textbf{1T state} = \frac{\textbf{1}}{\textbf{operating frequency of 8085}}$$

15.5.2 Timer Delay Using Counters:

Counting can create time delays. Since the execution times of the instructions used in a counting routine are known, the initial value of the counter, required to get specific time delay can be determined.

Using 8 bit counter :

```
                                                    Number of T-states
        MVI C, count    ; Load count                7 T-states
BACK :  DCR C           ; Decrement count           4 T-states
        JNZ BACK        ; If count ≠ 0, repeat      10/7 T-states
```

Using 16 Bit Counter :

```
                                                    Number of T-states
        LXI B, count    ; load 16 bit count         10 T-states
BACK :  DCX B           ; Decrement count           6 T-states
        MOV A, C        ;                           4 T-states
        ORA B           ; logically OR B and C      4 T-states
        JNZ BACK        ; If result is not 0, repeat 10 T-states
```

15.5.3 Timer Delay Using Nested Loops:

In this, there are more than one loops. The innermost loop is same as explained above. The outer loop sets the multiplying count to the delays provided by the innermost loop.

Table 15.8: Timer Delay Using Nested Loops

START:	MVI B, Multiplier count	Initialize multiplier	Number of T states
BACK:	MVI C, Delay count	Initialize delay count	7 T-states
	DCR C	Decrement delay count	7 T-states
	JNZ BACK	If not 0, repeat	4 T-states
	DCR B	Decrement multiplier count	10/7 T-states
	JNZ START	If not 0, repeat	4 T-states
			10/7 T-states

15.6 STACK AND SUBROUTINE:

The stack is a reserved area of memory in RAM where we can store temporary information. Interestingly, the stack is a shared resource as it can be shared by the microprocessor and the programmer. The programmer can use the stack to store data. And the microprocessor uses the stack to execute subroutines. The 8085 has a 16-bit register known as the 'Stack Pointer.'

This register's function is to hold the memory address of the stack. This control is given to the programmer. The programmer can decide the starting address of the stack by

loading the address into the stack pointer register at the beginning of a program. The stack works on the principle of First In Last Out. The memory location of the most recent data entry on the stack is known as the Stack Top.

How does a stack work in assembly language?

We use two main instructions to control the movement of data into a stack and from a stack. These two instructions are PUSH and POP.

PUSH – This is the instruction we use to write information on the stack.

POP – This is the instruction we use to read information from the stack.

There are two methods to add data to the stack: Direct method and Indirect method.

15.6.1 Direct Method:

In the direct method, the stack pointers address is loaded into the stack pointer register directly**.**

Explanation of the code:

LXI SP, 8000H – The address of the stack pointer is set to 8000H by loading the number into the stack pointer register.

LXI H, 1234H – Next, we add a number to the HL pair. The most significant two bits will enter the H register. The least significant two bits will enter the L register.

PUSH H – The PUSH command will push the contents of the H register first to the stack. Then the contents of the L register will be sent to the stack. So the new stack top will hold 34H.

POP D – The POP command will remove the contents of the stack and store them to the DE register pair. The top of the stack clears first and enters the E register. The new top of the stack is 12H now. This one clears last and enters the D register. The contents of the DE register pair is now 1234H.

HLT – HLT indicates that the program execution needs to stop.

15.6.2 Indirect Method:

In the indirect method, the stack pointers address is loaded into the stack pointer register via another register pair.

LXI H, 8000H

SPHL

LXI H, 1234H

PUSH H

POP D

HLT

Explanation of the Code:

LXI H, 8000H – The number that we wish to enter into the stack pointer, 8000H, is loaded into the HL pair register.

SPHL – This is a special command that we can use to transfer data from HL pair to Stack pointer (SP). Now, the contents of the HL pair are in the SP.

LXI H, 1234H – Next, we add a number to the HL pair. The most significant two bits will enter the H register. The least significant two bits will enter the L register.

PUSH H – The PUSH command will push the contents of the H register first to the stack. Then the contents of the L register will be sent to the stack. So the new stack top will hold 34H.

POP D – The POP command will remove the contents of the stack and store them to the DE register pair. The top of the stack clears first and enters the E register. The new top of the stack is 12H now. This one clears last and enters the D register. The contents of the DE register pair is now 1234H.

HLT – HLT indicates that the program execution needs to stop.

Both the methods can be shown diagrammatically with the following diagram.

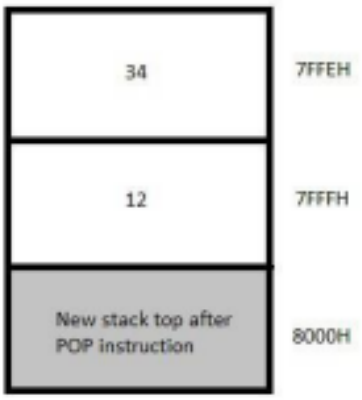


Fig. 15.9

15.6.3 Subroutine:

A subroutine is a small program written separately from the main program to perform a particular task that you may repeatedly require in the main program. Essentially, the concept of a subroutine is that it is used to avoid the repetition of smaller programs. Subroutines are

written separately and are stored in a memory location that is different from the main program. You can call a subroutine multiple times from the main program using a simple CALL instruction.

What are the conditional CALL statements in assembly language?

You can use conditional CALL statements, too, according to your needs. These statements enter a subroutine only when a certain condition is met.

CC Call at address if cy (carry flag) = 1

CNC Call at address if cy (carry flag) = 0

CZ Call at address if ZF (zero flag) = 1

CNZ Call at address if ZF (zero flag) = 0

CPE Call at address if PF (parity flag) = 1

CPO Call at address if PF (parity flag) = 0

CN Call at address if SF (signed flag) = 1

CP Call at address if SF (signed flag) = 0

The subroutine can be exited from using a return (RET) instruction.

What are the conditional return (RET) statements in assembly language?

Akin to the CALL instruction, we have conditional RET statements too.

These statements ensure that a subroutine is exited only when a certain condition is met.

RC Return from subroutine if cy (carry flag) = 1

RNCReturn from subroutine if cy (carry flag) = 0

RZ Return from subroutine if ZF (zero flag) = 1

RNZReturn from subroutine if ZF (zero flag) = 0

RPE Return from subroutine if PF (parity flag) = 1

RPOReturn from subroutine if PF (parity flag) = 0

RN Return from subroutine if SF (signed flag) = 1

RP Return from subroutine if SF (signed flag) = 0

After the completion of the subroutine, the main program begins from the instruction immediately following the CALL instruction.

What are the advantages of using subroutines?

- Subroutines avoid the repetition of instructions.
- They give an aspect of modular programming to the entire program.
- Improves efficiency by reducing the size of the memory needed to store the program.

15.7 INTRODUCTION TO MICRO CONTROLLERS:

A microcontroller is a small and low-cost microcomputer, which is designed to perform the specific tasks of embedded systems. Microprocessors are different than microcontrollers in their design. Microprocessors have the only CPU inside them and no in-memory support while Microcontrollers, on the other side, has CPU, RAM, ROM and other peripherals which are all embedded on the chip.

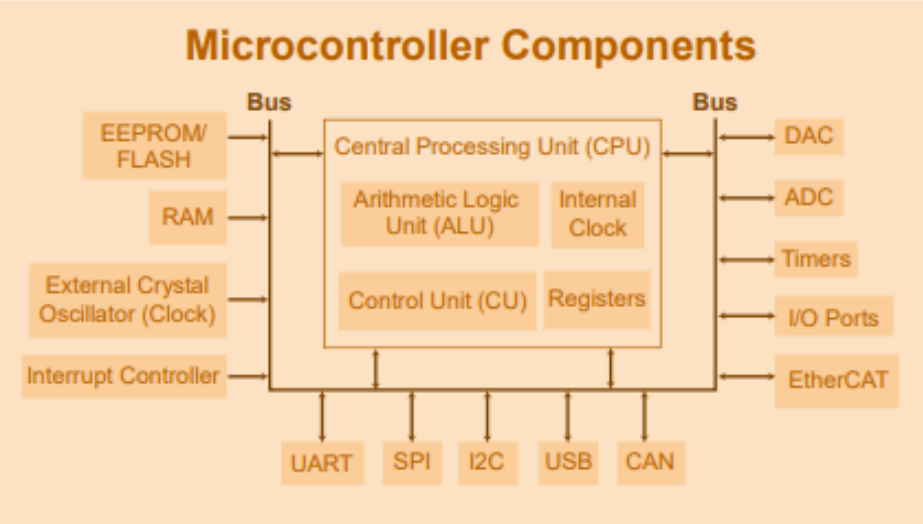


Fig. 15.10 Microcontroller Components

15.7.1 Types of Memory:

Three types of memories are commonly used in modern microcontrollers.

1) Random Access Memory (RAM)
2) Electrically Erasable Programmable Read-Only Memory (EEPROM)
3) Flash Memory Random Access Memory (RAM)

- It is a volatile memory, which means it can only be accessed when the system is powered-up. As the system power shuts down, so do the contents inside the RAM.
- It has a faster access time, but this is also expensive when compared with the other ones. It is mainly used to store temporary data, variables, and constants during the execution of user instructions.

Electrically Erasable Programmable Read-Only Memory (EEPROM)

- It is a non-volatile memory, which means that contents in such a memory remain intact even when the power is removed.
- The main use of EEPROM in the microcontroller is to store user instructions and data. The microcontroller then accesses the instructions one-by-one and sequentially executes them.

Flash Memory:

- Flash Memory is also non-volatile memory.
- Unlike EEPROM, it can only program a whole sector or block of bytes. Some microcontroller manufacturers use either Flash or EEPROM, while others use both to provide more flexibility to the end-user.

15.7.2 Control Unit:

A control unit controls the operations of all parts of the computer but it does not carry out any data processing operations. Some main functions of the control unit are listed below:

- Controlling of data and transfer of data and instructions is done by the control unit among other parts of the computer.
- The control unit is responsible for managing all the units of the computer.
- The main task of the control unit is to obtain the instructions or data which is input from the memory unit, interprets them, and then directs the operation of the computer according to that.
- The control unit is responsible for communication with input and output devices for the transfer of data or results from memory.
- The control unit is not responsible for the processing of data or storing data.

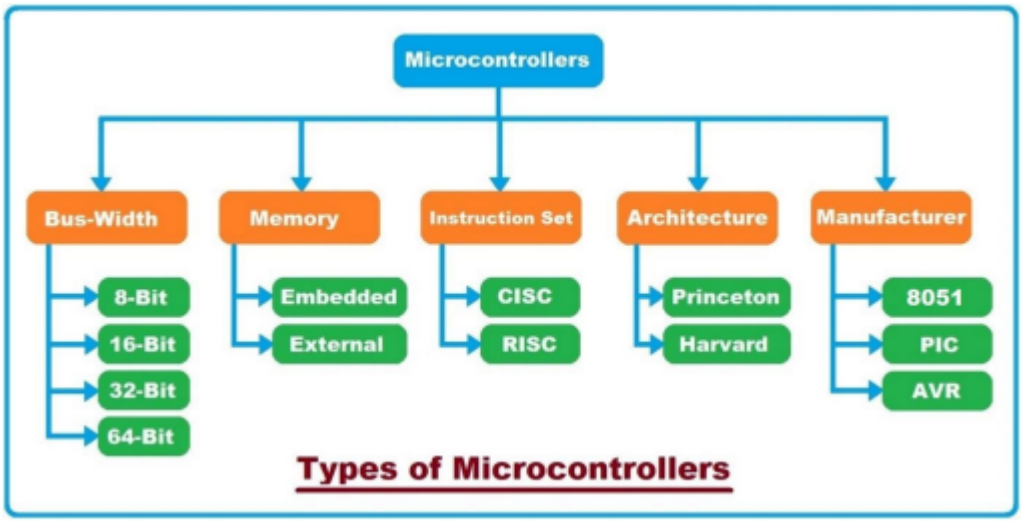


Fig. 15.11: Types of Microcontrollers

15.8 8051 MICROCONTROLLERS:

Introduction of 8051:

Salient features of 8051 microcontroller are given below:

- Eight bit CPU
- On chip clock oscillator
- 4Kbytes of internal program memory (code memory) [ROM]
- 128 bytes of internal data memory [RAM]
- 64 Kbytes of external program memory address space.
- 64 Kbytes of external data memory address space.
- 32 bi directional I/O lines
- Two 16 Bit Timer/Counter :T0, T1
- Full Duplex serial data receiver/transmitter
- Four Register banks with 8 registers in each bank.
- Sixteen bit Program counter (PC) and a data pointer (DPTR)
- 8 Bit Program Status Word (PSW)
- 8 Bit Stack Pointer
- Five vector interrupt structure (RESET not considered as an interrupt.)
- 8051 CPU consists of 8 bit ALU with associated registers like accumulator 'A', B register, PSW, SP, 16 bit program counter, stack pointer.
- ALU can perform arithmetic and logic functions on 8 bit variables.

In the following diagram, the system bus connects all the support devices to the CPU. The system bus consists of an 8-bit data bus, a 16-bit address bus and bus control signals. All other devices like program memory, ports, data memory, serial interface, interrupt control, timers, and the CPU are all interfaced together through the system bus.

15.8.1 CPU (Central Processor Unit):

As you may be familiar that Central Processor Unit or CPU is the mind of any processing machine. It scrutinizes and manages all processes that are carried out in the Microcontroller. User has no power over the functioning of CPU. It interprets program printed in storage space (ROM) and carries out all of them and do the projected duty. CPU manages different types of registers in 8051 microcontroller.

15.8.2 Oscillator:

As we all make out that Microcontroller is a digital circuit piece of equipment, thus it needs timer for its function. For this function, Microcontroller 8051 consists of an on-chip oscillator which toils as a time source for CPU (Central Processing Unit).

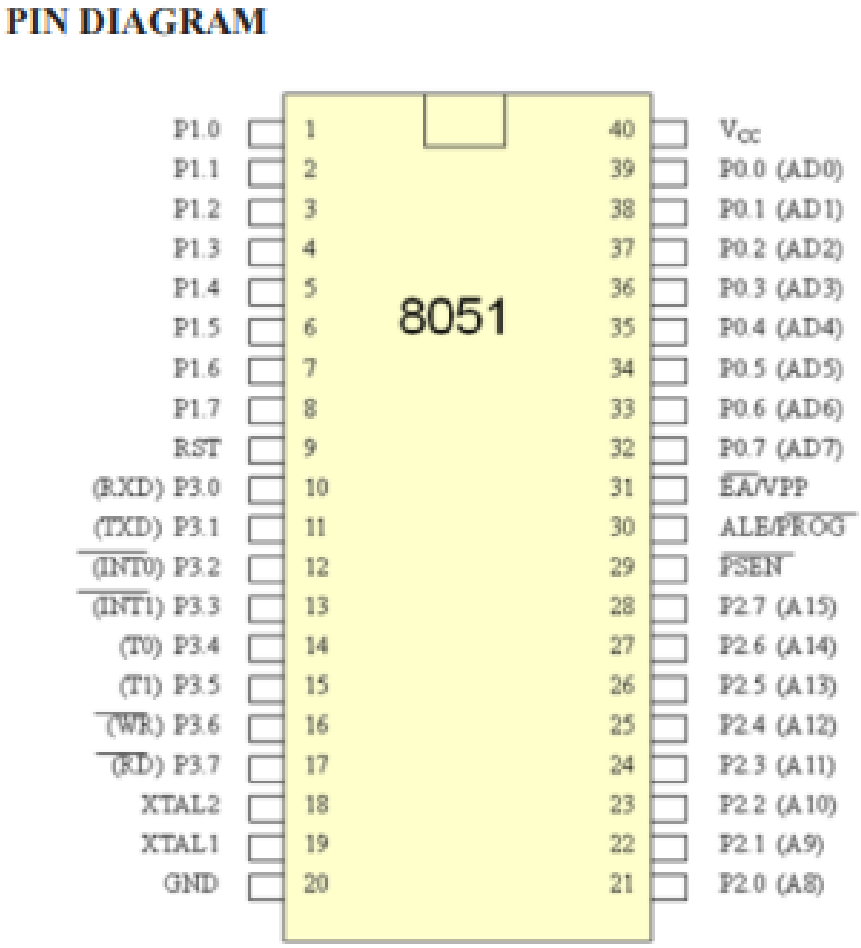


Fig. 15.12: PIN Diagram

Pins 1-8: Recognized as Port 1. Unlike other ports, this port doesn't provide any other purpose. Port 1 is a domestically pulled up, quasi bidirectional Input/output port.

Pin 9: As made clear previously RESET pin is utilized to set the micro-controller 8051 to its primary values, whereas the micro-controller is functioning or at the early beginning of application. The RESET pin has to be set elevated for two machine rotations.

Pins 10-17: Recognized as Port 3. This port also supplies a number of other functions such as timer input, interrupts, serial communication indicators TxD & RxD, control indicators for outside memory interfacing WR & RD, etc. This is a domestic pull up port with quasibi directional port within.

Pins 18 and 19: These are employed for interfacing an outer crystal to give system clock.

Pin 20: Titled as Vss - it symbolizes ground (0 V) association.

Pins 21-28: recognized as Port 2 (P 2.0 - P 2.7) Other than serving as Input/output port, senior order address bus indicators are multiplexed with this quasi bi directional port.

Pin 29: Program Store Enable or PSEN is employed to interpret sign from outer program memory.

Pin 30: External Access or EA input is employed to permit or prohibit outer memory interfacing. If there is no outer memory need, this pin is dragged high by linking it to Vcc.

Pin 31: Aka Address Latch Enable or ALE is brought into play to de-multiplex the address data indication of port 0 (for outer memory interfacing). Two ALE throbs are obtainable for every machine rotation.

Pins 32-39: Recognized as Port 0 (P0.0 to P0.7) - other than serving as Input/output port, low order data & address bus signals are multiplexed with this port (to provide the use of outer memory interfacing). This pin is a bidirectional Input/output port (the single one in microcontroller 8051) and outer pull up resistors are necessary to utilize this port as Input/output.

Pin 40: Termed as Vcc is the chief power supply. By and large it is +5V DC.

15.8.3 Registers in 8051:

- The 8051 contains 34 general purpose or working registers.
- Two of these Register A and B.
- Register A (Accumulator) is a 8 bit register used by all arithmetic and logical operation. It can store 8 bit operand before execution of an instruction. The immediate result is stored in the accumulator register (Acc) after execution of instruction.
- The B register is a register just for multiplication and division operation which requires more register spaces for the product of multiplication and the quotient and the remainder for the division.

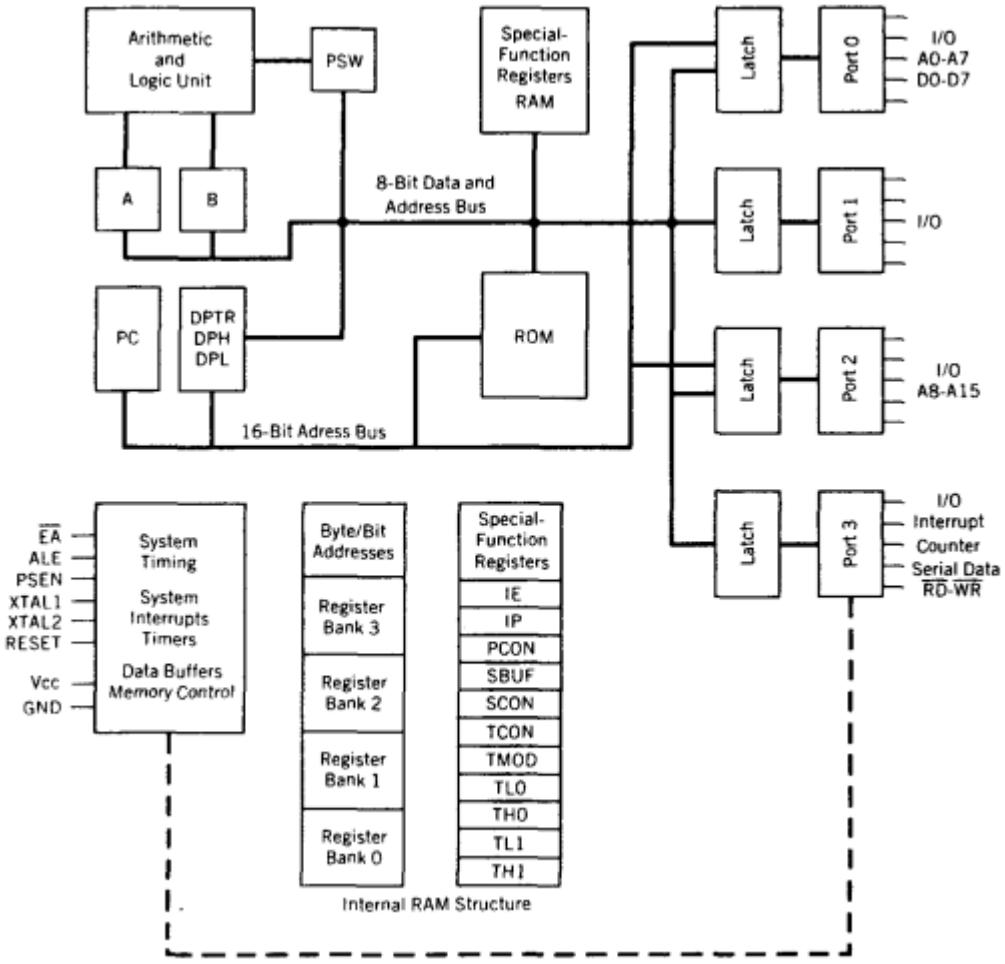


Fig. 15.13: 8051 Block Diagram

15.9 ARCHITECTURE & PIN DESCRIPTION:

15.9.1 The 8051 Architecture:

The salient features of 8051 microcontroller are given below.

- Eight bit CPU
- On chip clock oscillator
- 4k bytes of internal program memory (code memory) [ROM]
- 128 bytes of internal data memory [RAM]
- 64 bytes of external program memory address space
- 64 bytes of external data memory address space
- 32 bi directional I/O lines
- Two 16 bit Timer/Counter: T0,T1

- Full duplex serial data receiver/ transmitter
- Four register banks with 8 registers in each bank
- Sixteen-bit program counter (PC) and a data pointer (DPTR)
- 8 Bit program status word (PSW)
- 8 Bit stack Pointer
- Five vector interrupt structure (RESET not considered as an interrupt)
- 8051 CPU consists of 8 bit ALU with associated registers like accumulator 'A', B register, PSW,SP, 16 bit program counter, stack pointer.
- ALU can perform arithmetic and ligic functions on 8 bit variables.

  In the following diagram, the system bus connects all the support devices to the CPU. The system bus consists of an 8 bit data bus, a 16 bit address bus and bus control signals. All other devices like program memory, ports, data memory, serial interface, interrupt control and the CPU are all interfaced together trough the system bus

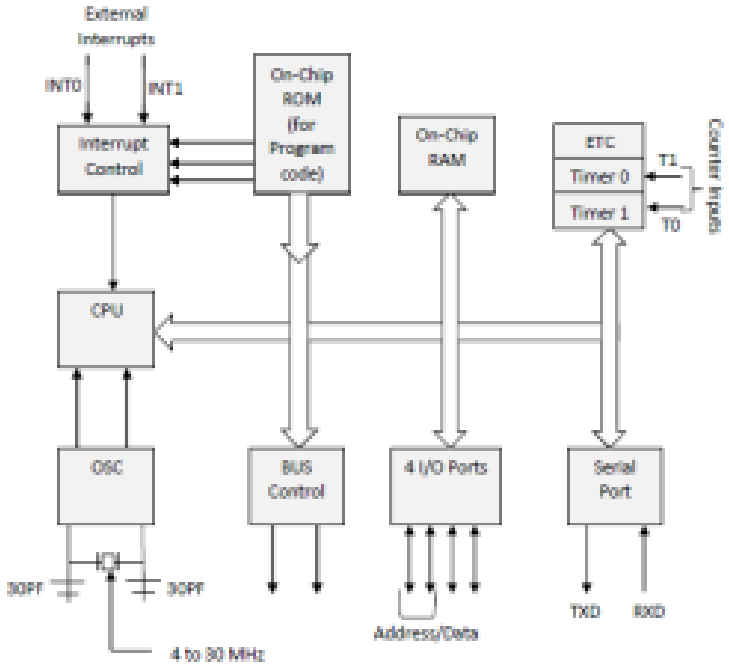


Fig. 15.14: 8051 Architecture Diagram

CPU (Central Processor Unit): As you may be familiar that Central Processor Unit or CPU is the mind of any processing machine. It scrutinizes and manages all processes that are carried out in the Microcontroller. User has no power over the functioning of CPU. It interprets program printed in storage space (ROM) and carries out all of them and do the projected duty. CPU manages different types of registers in 8051microcontrollerOscillator:As we all make out that Microcontroller is a digital circuit piece of equipment, thus it needs

timer for its function. For this function, Microcontroller 8051 consists of an on-chip oscillator which toils as a time source for CPU (Central Processing Unit). Bus: Fundamentally Bus is a group of wires which functions as a communication canal or mean for the transfer Data. These buses comprise of 8, 16 or more cables. As a result, a bus can bear 8 bits, 16 bits all together. There are two types of buses:

1) **Address Bus:** Microcontroller 8051 consists of 16 bit address bus. It is brought into play to address memory positions. It is also utilized to transmit the address from Central Processing Unit to Memory.
2) **Data Bus:** Microcontroller 8051 comprise of 8 bits data bus. It is employed to cart data.

Interrupts:

As its name suggests, Interrupt is a subroutine call that interrupts of the microcontrollers main operations or work and causes it to execute any other program, which is more important at the time of operation.

The Microcontroller 8051 can be configured in such a way that it temporarily terminates or pause the main program at the occurrence of interrupts. When a subroutine is completed, then the execution of main program starts. Generally five interrupt sources are there in 8051 Microcontroller. There are 5 vectored interrupts are shown in below

- INTO
- TFO
- INT1
- TF1
- R1/T1

Memory:

The memory which is used to store the program of the microcontroller is known as code memory or Program memory of applications. It is known as ROM memory of microcontroller also requires a memory to store data or operands temporarily of the micro controller. The data memory of the 8051 is used to store data temporarily for operation is known RAM memory. 8051 microcontroller has 4K of code memory or program memory,that has 4KB ROM and also 128 bytes of data memory of RAM.

Input/Output Port:

Normally microcontroller is used in embedded systems to control the operation of machines in the microcontroller. Therefore, to connect it to other machines, devices or peripherals we require I/O interfacing ports in the microcontroller interface. For this purpose microcontroller

8051 has 4 input, output ports to connect it to the other peripherals Timers/Counters 8051 microcontroller has two 16 bit timers and counters. These counters are again divided into a 8 bit register.

The timers are used for measurement of intervals to determine the pulse width of pulses.

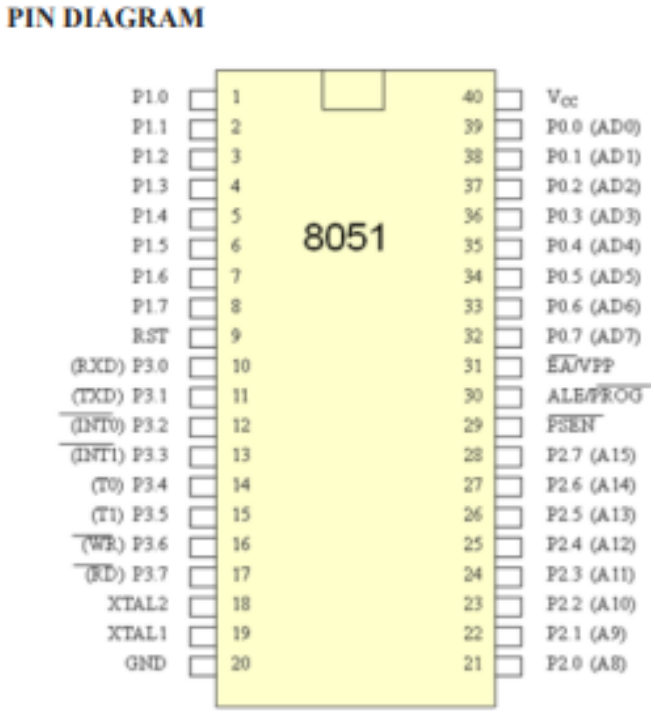


Fig. 15.15: PIN Diagram of 851 Microcontroller

Pin 1-8: Recognized as port 1. Unlike other ports, this port does not provide any other purpose. Port 1 is a domestically pulled up, quasi bidirectional Input/output port.

Pin 9: As made clear previously RESET pin is utilized to set the micro controller 8051 to its primary values, whereas the micro controller is functioning or at the early beginnings of application. The RESET pin has to be set elevated for two machine rotations.

Pin 10-17: Recognized as port 3. This port also supplies a number of other functions such as timer input, interrupts, serial communication indicators TxD & RxD, control indicators for outside memory.

Pin 18 and 19: These are employed for interfacing an outer crystal to give system clock.

Pin 20: Titled as Vss- it symbolizes ground (0 V) association.

Pins 21-28: Recognized as port 2 (P 2.0 - P 2.7) other than serving as input/output port, senior order address bus indicators are multiplexed with this quasi-bidirectional port.

Pin 29: Program store enables PSEN is employed to interpret signs from outer program memory.

Pin 30: External access or EA input is employed to permit or prohibit outer memory interfacing. If there is no outer memory need, this pin is dragged high by linking it to Vcc.

Pin 31: Aka address latch enables or ALE is brought into play to multiplex the address data indication of port 0(for outer memory interfacing). Two ALE throbs are obtainable for every machine rotation.

Pins 32-39: Recognized as port 0 (P0.0 to P0.7) – other than serving as input/output port, low order data & address bus signals are multiplexed with this port (to provide the use of outer memory interfacing). This pin is a bidirectional input/output port (the single one in microcontroller 8051) and outer pull up resistors are necessary to utilize this port as inpt/output.

Pin 40: Termed as Vcc is the chief power supply. By and large it is +5V DC.

15.10 SUMMARY:

This study explores microcontrollers, focusing on their architecture, memory types, and operational components, with a particular emphasis on the 8051 microcontrollers. It covers the distinctions between microcontrollers and microprocessors, highlighting the integration of CPU, RAM, ROM, and peripherals in microcontrollers. The study details various memory types such as RAM, EEPROM, and Flash Memory, explaining their roles in data storage and execution. Key components like the control unit, CPU, oscillator, buses, interrupts, and registers are examined to understand their functions in processing and execution. Additionally, the study provides insights into the pin configuration and interfacing capabilities of the 8051 microcontrollers. Through this, it establishes a strong foundation for understanding microcontrollers and their applications in embedded systems, automation, and digital electronics, making them essential in modern technological advancements.

15.11 TECHNICAL TERMS:

Microcontroller, EEPROM (Electrically Erasable Programmable Read-Only Memory), ALU (Arithmetic Logic Unit), Oscillator.

15.12 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are the key differences between a microcontroller and a microprocessor?
2) Explain the three main types of memory used in microcontrollers and their characteristics.
3) What is the function of the control unit in a microcontroller, and how does it manage data flow?

Short Answer Questions:

1) Describe the role of the oscillator in the 8051 microcontroller and why it is essential for operation.
2) How do interrupts work in a microcontroller, and why are they important for real-time processing?
3) Identify and explain the functions of at least four important pins in the 8051 microcontrollers.

15.13 SUGGESTED READINGS:

1) **Mazidi, M.A., Mazidi, J.G., & McKinlay, R.D.** (2006). The 8051 Microcontroller and Embedded Systems: Using Assembly and C. Pearson Education.
2) **Kenneth J. Ayala** (2004). The 8051 Microcontroller. Cengage Learning.
3) **Raj Kamal** (2011). Microcontrollers: Architecture, Programming, Interfacing and System Design. Pearson.
4) **Krishna Kant** (2014). Microprocessors and Microcontrollers: Architecture, Programming and System Design 8085, 8086, 8051, 8096. PHI Learning.

Prof. Ch. Linga Raju

LESSON-16

DATA INTERPRETATION AND ANALYSIS

16.0 AIM AND OBJECTIVES:

The aim of this lesson is to develop a comprehensive understanding of data interpretation and analysis, enabling effective decision-making and problem-solving. The objectives include understanding data sources, structures, and types, ensuring data accuracy through cleaning and preprocessing, and applying exploratory data analysis (EDA) techniques to identify patterns, trends, and correlations. Additionally, learners will explore statistical methods, predictive modeling, and hypothesis testing to extract meaningful insights. The lesson also emphasizes precision and accuracy in measurements, addressing systematic and random errors, and understanding error propagation. By mastering least-squares fitting and data visualization, learners will improve their ability to communicate results effectively. Ultimately, the goal is to enhance analytical skills for diverse applications, from scientific research to business intelligence, ensuring data-driven strategies and informed decision-making.

STRUCTURE:

16.1 DATA INTERPRETATION AND ANALYSIS:

Data interpretation and analysis involve extracting meaningful insights from raw data to support decision-making and problem-solving. This process begins with understanding the data - identifying its source, structure, and type (qualitative or quantitative). Proper context is critical to ensure the analysis aligns with the goals. For example, a business may analyze

customer feedback to improve products, while a researcher might study experimental data to confirm a hypothesis. The preparation stage includes cleaning the data by addressing missing values, outliers, and inconsistencies to ensure its reliability and accuracy.

Once the data is prepared, exploratory data analysis (EDA) is conducted to uncover patterns, relationships, and trends. This step uses descriptive statistics (like mean and standard deviation) and visualizations (such as histograms, scatter plots, and boxplots) to summarize and understand the data. Analysts may identify correlations, anomalies, or groupings that suggest further investigation. For instance, a retailer could spot seasonal purchasing trends or a scientist might notice unexpected variables influencing an outcome.

Advanced techniques like statistical analysis or predictive modeling follow EDA to deepen insights. Statistical methods, such as regression analysis or hypothesis testing, help quantify relationships and test assumptions. Predictive analytics, often powered by machine learning models, forecasts future trends based on historical data. These approaches enable informed decision-making, such as optimizing supply chains, targeting marketing campaigns, or projecting financial performance.

The final step is interpretation and reporting, where raw findings are transformed into actionable insights. This requires clear communication, often supported by visual tools like dashboards or charts, to ensure stakeholders understand the results. For example, a report might illustrate how a new strategy increased sales or highlighted factors contributing to employee satisfaction. Effective interpretation bridges the gap between complex data and practical applications, enabling informed strategies and driving success across industries.

16.2 PRECISION AND ACCURACY:

Having determined the best values of the parameters, your problems are not yet solved! You also wish to assess the validity of the fit and you wish to have an estimate of the inaccuracies in the parameters. The key to the answers to these problems lies in the value of $\chi 2$ as a function of the parameters. The next two sections explain all this.

Having determined the best values of the parameters, your problems are not yet solved! You also wish to assess the validity of the fit and you wish to have an estimate of the inaccuracies in the parameters. The key to the answers to these problems lies in the value of $\chi 2$ as a function of the parameters. The next two sections explain all this.

Consider another example in which a researcher is collecting gravitational acceleration data. If five measurements are collected then the following scenarios are possible.

(1) High accuracy and high precision: $9.83 m/s^2 \pm 0.01 m/s^2$.

The measured value is close to known value of 9.8 m/s^2 and has a low uncertainty of $0.01m/s^2$.

(2) High accuracy and low precision: 9.7 $m/s^2 \pm 0.01m/s^2$.

The measured value is close to known value of 9.8 m/s^2 and has a not-so-low uncertainty of $0.1m/s^2$, compared to $0.01m/s^2$.

(3) Low accuracy and high precision: 12.64 $m/s^2 \pm 0.01m/s^2$

The measured value is not close to known value of 9.8 m/s^2 and has a low uncertainty of 0.01 m/s^2.

(4) Low accuracy and low precision: 12.64 $m/s^2 \pm 0.01m/s^2$

The measured value is not close to known value of 9.8 m/s^2 and has a not-so-low uncertinity of 0.1 m/s^2, compared to 0.01 m/s^2.

Both accuracy and prcision will play a role in explaining the discrepancy between a final measured value and a known or accepted value. Both must always be considered in an error analysis discussion.

If a final measured value is found using combinations of several individual measured values, the uncertainty or precision of the final value is dependent on the uncertainty or precision of the individual measured values. The final uncertainty tells us to what degree we known our measurement is correct.

If a final measured value is found using combinations of several individual measured values, the accuracy of the final measured values. The final accuracy tells us how correct the final value is.

16.3 ERROR ANALYSIS:

There are errors and uncertainties. The latter are unavoidable; eventually it is the omnipresent thermal noise that causes the results of measurements to be imprecise. After trying to identify and correct avoidable errors, this chapter will concentrate on the propagation and combination of uncertainties in composite functional relations.

Classification of Errors:

There are several types of error in experimental outcomes:

(i) Accidental, stupid or intended) mistakes

(ii) Systematic deviations

(iii) Random errors or uncertainties

The first type we shall ignore. Accidental mistakes can be avoided by careful checking and double checking. Stupid mistakes are accidental errors that have been overlooked. Intended mistakes (e.g. selecting data that suit your purpose) purposely mislead the reader and belong to the category of scientific crimes.

16.3.1 Systematic Errors:

Systematic errors have a non-random character and distort the result of a measurement. They result from erroneous calibration or just from a lack of proper calibration of a measuring instrument, from careless measurements (uncorrected parallax, uncorrected zero-point deviations, time measurements uncorrected for reaction time, etc.), from impurities in materials, or from causes the experimenter is not aware of. The latter are certainly the most dangerous type of error; such errors are likely to show up when results are compared to those of other experimentalists at other laboratories. Therefore, independent corroboration of experimental results is required before a critical experiment (e.g. one that overthrows an accepted theory) can be trusted.

16.3.2 Random Errors or Uncertainties:

Random errors are unpredictable by their very nature. They can be caused by the limited precision of instrumental readings, but are ultimately due to physical noise, i.e. by natural fluctuations due to thermal motions or to the random timing of single events. Since such errors are unavoidable and unpredictable, the word “error” does not convey the proper meaning, and we prefer to use the term uncertainty for the possible random deviation of a measured result from its true value. If a measurement is repeated many times, the results will show a certain spread around an average value, from which the estimated inaccuracy in the average can be determined. The probability distribution, from which the measured values are random samples, is supposed to obey certain statistical relations, from which rules to process the uncertainties can be derived. In the case of a single measurement one should estimate the uncertainty, based on knowledge of the measuring instrument. For example, a length read on a ruler will be accurate to ±0.2 mm; a length read on a vernier caliper will be accurate to ±0.05 mm. Chemists reading a liquid level on a buret or graduated cylinder can estimate volumes with a precision of ±0.3 scale divisions. Be aware of the precision of digital instruments: they usually display more digits than warranted by their precision. The precision of reliable commercial instruments is generally indicated by the manufacturer, sometimes as an individual calibration report. Often the maximum error is given, which can have a (partly) systematic character and which exceeds the standard deviation.

$$\sigma_f = \left|\frac{df}{dx}\right|\sigma_x \quad \textbf{.(16.1)}$$

16.4 PROPAGATION OF ERRORS:

16.4.1 Propagation through Functions:

In general the required end result of a series of measurements is a function of one or more measured quantities. For example, if you measure the length l and width w of a rectangular plane object, both the circumference C = 2(l + w) and the area A = lw are (simple) functions of l and w. Assume the deviations in l and w are independent of each other with standard uncertainties "l and "w, respectively, what then is the standard uncertainty in C or in A? A somewhat more complicated relation is the determination of the change in standard Gibbs function "G0 for an equilibrium reaction with measured equilibrium constant K:

$$\Delta G^0 = -RT \ln K, \qquad (16.2)$$

where R is the gas constant and T the absolute temperature. What is the standard uncertainty in "G0 given the standard uncertainty in KAnd if the equilibrium constant K of a dimerization reaction 2A ! A2 is determined by measuring concentrations [A] and [A2]:

$$K = \frac{[A_2]}{[A]^2} \qquad (16.3)$$

How can we determine the standard uncertainty in K given those in [A] and [A2], assuming the deviations of [A] and [A2] to be independent? How will the latter be modified if the deviations are not independent, e.g. if we measure both the total concentration [A] + 2[A2] and [A2] independently? What we need to establish is the propagation of uncertainties. The clue is differentiation: If the standard uncertainty in x equals σx, then the standard uncertainty σf in f(x) equals

16.5 LEAST SQUARES FITTING:

When the function to be fitted f(x, θ) with parameters θ = θ1, ... , θn is more general than ax + b, the following cases should be distinguished:

i) f linear in θ. For functions linear in all parameters, such as an analytical (weighted) least-squares solution minimizing S is still possible but requires some matrix algebra. Appendix A9 gives details.

$$f(x) = ax_2 + bx + c \qquad (16.4)$$

$$f(x) = a + b \exp(-k_1 x) + c \exp(-k_2 x); \quad (k_1, k_2 \text{ known constants})$$

$$f(x) = ax + b/x + c, \qquad (16.5)$$

ii) f linear in several variables. Functions linear in more than one independent ("explanatory") variable, such as

$$\boldsymbol{f(\xi,\eta,\zeta) = a\xi + b\,\eta + c\,\zeta + d} \qquad \textbf{(16.6)}$$

where ξ , η, ζ are the independent variables and a, b, c, d are the parameters, similarly yield analytical least-squares solutions using some matrix algebra.

iii) f nonlinear but linearizable. Functions that are nonlinear in the parameters can often be transformed to linear functions, as was done in order to obtain linear graphs. For example, the function f(t) = a exp(−kt) is not linear in the parameter k. But if you take the logarithm:

$$\boldsymbol{\ln f(t) = -kt + \ln a,} \qquad \textbf{(16.7)}$$

you obtain a function that is linear in k and has the form ax + b. So you can apply linear regression to the points (ti, ln yi) and determine k and ln a. But take proper care of the weights: if all values of y have equal standard deviations σ, the values ln yi have different weights:

$$\boldsymbol{\sigma_{\ln y} = \left|\frac{d \ln y}{dy}\right| \sigma_x = \sigma_x / y_i} \qquad \textbf{(16.8)}$$

In this case you should take the weights wi equal to (or proportional to) y2 i. Negative values of y, which may occur by random deviations for large values of t, cannot be handled. It is not allowed to selectively omit negative values, as that will bias the result. The best way to proceed is to omit all points with larger values for t than the value for which a negative y first occurred. Even better is the use of a general nonlinear fitting procedure.

iv) f nonlinear: general case. Functions that are nonlinear in the parameters cannot always be linearized. For example, the function

$$\boldsymbol{f(t) = a \exp(-kt) + b} \qquad \textbf{(16.9)}$$

cannot be transformed to a linear function of the parameters a, b, and k. A nonlinear least-squares fitting procedure must then be used. In this case there are no analytical solutions and solutions are obtained by iterative function minimizers. There are several minimizers available, some that require analytical derivatives and some that do not. The latter are easier to use. In all cases an initial guess of the parameters is required; for some functions a bad guess may lead to failure of the minimization

procedure. A graphical analysis is a good source for a reasonable initial guess. Below an example is given of a nonlinear least square minimization using Python.

Example: Nonlinear Fit

Consider the data on enzyme kinetics given. Given are six data points Si, vi with equal weights and our task is to fit two parameters p0 = vmax and p1 = Km in the function

$$f(S,p) = \frac{p_0 S}{p_1 + S} \qquad (16.10)$$

$$p = [V_{max}, K_m] \qquad (16.11)$$

such that the sum of squares

$$SSQ = \sum_i [v_i - f(S_i, p)]^2 \qquad (16.12)$$

is minimal. One possibility is to use the Python least-squares minimization procedure leastsq that comes with the module optimize of SciPy. This function requires a specification of the residues yi − fi (or (yi − fi)/σi if s.d.'s are known), which are a function of the parameters p, but does not require any derivatives. It must be called with an initial guess for p,

$$p_{init} = [15, 105] \qquad (16.13)$$

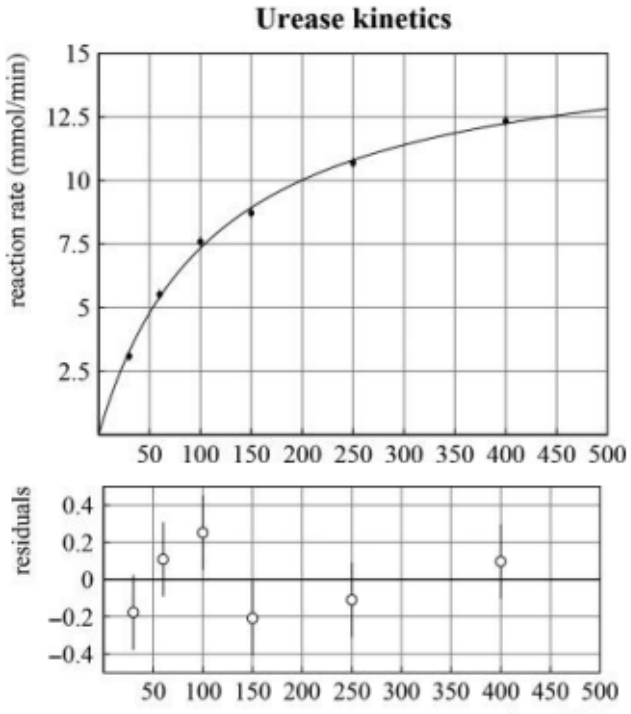


Figure 16.1 Upper panel: The urease reaction rate data plotted together with the least sqare fitted function. Lower panel: Plotting the residuals yi-fi with error bars more clearly show whether the deviations have a random character or not.

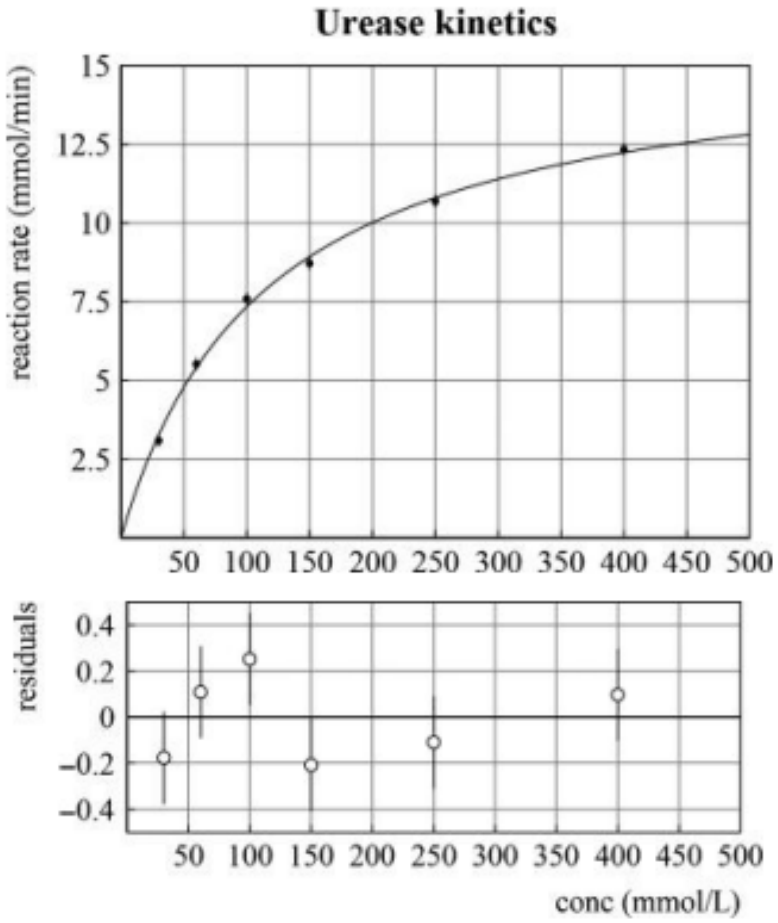


Figure 16.2 Upper panel: The urease reaction rate data plotted together with the least square fitted function. Lower panel: Plotting the residuals yi-fi with error bars more clearly show whether the deviations have a random character or not

The SSQ using this initial guess equals 0.375. After applying the minimization procedure, the parameters appear to be

$$\mathbf{p_{init}} = [\mathbf{15.75}, \mathbf{114.65}] \qquad \textbf{(16.14)}$$

and the minimal SSQ is 0.171. Figure 16.2 shows the fit, together with a plot of the residues. The latter plot is able to show the size of the error bars and gives a visual impression of any systematic deviations. Later we shall see how large the uncertainties in the parameters appear to be. Another possibility is to use the Python procedure fmin_powell, also in the module optimize of SciPy. With this minimizer the function to be minimized must be specified. This routine is less accurate than leastsq and should preferably be applied more than once.

Having determined the best values of the parameters, your problems are not yet solved! You also wish to assess the validity of the fit and you wish to have an estimate of the inaccuracies in the parameters. The key to the answers to these problems lies in the value of $\chi 2$ as a function of the parameters. The next two sections explain all this.

16.6 SUMMARY:

Data interpretation and analysis involve extracting meaningful insights from raw data to support decision-making. The process starts with understanding data sources, structures, and types, followed by cleaning and preprocessing to ensure accuracy. Exploratory Data Analysis (EDA) helps identify patterns, trends, and correlations using statistical tools and visualizations. Advanced techniques like regression analysis, hypothesis testing, and predictive modeling enhance insights. Precision and accuracy play a crucial role in measuring validity, while error analysis addresses systematic and random errors. Understanding error propagation helps estimate uncertainties in computed values. Least-squares fitting is essential for modeling relationships in data. Effective interpretation and communication of findings, supported by visual tools, ensure informed strategies in fields like research, business, and engineering.

16.7 TECHNICAL TERMS:

Exploratory Data Analysis (EDA), Hypothesis Testing, Error Propagation, Least-Squares Fitting, Systematic Error

16.8 SELF-ASSESSMENT QUESTIONS:

Essay Questions:

1) What are the key steps involved in data interpretation and analysis?
2) How does Exploratory Data Analysis (EDA) help in identifying patterns and trends in a dataset?
3) What is the difference between precision and accuracy in measurements, and why are both important in data analysis?

Short Answer Questions:

1) Explain the types of errors in experimental data and how they impact the reliability of results.
2) How does error propagation affect the final computed values in an experiment or analysis?
3) What is least-squares fitting, and how is it used in data modeling to minimize errors?

16.9 SUGGESTED READINGS:

1) Montgomery, D.C., & Runger, G.C. (2018). Applied Statistics and Probability for Engineers. Wiley.

2) Jolliffe, I.T., & Cadima, J. (2016). "Principal component analysis: a review and recent developments." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.

3) Box, G. E.P., Hunter, J.S., & Hunter, W.G. (2005). Statistics for Experimenters: Design, Innovation, and Discovery. Wiley-Interscience.

4) Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

5) Devore, J.L. (2015). Probability and Statistics for Engineering and the Sciences. Cengage Learning.

6) Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories. Cambridge University Press.

Prof. Ch. Linga Raju