BASIC RADIO AND TELEVISION

SECOND EDITION

About the Author

S P Sharma completed diploma course in radio engineering at Marconi College, Chelmsford (England) after obtaining his M.Sc. (Hons.) degree in Physics. He completed a special course in television practice at the South East Technical Institute, London. He received practical training in the technical departments of Marconi Works, British Broadcasting Corporation and British Post Office Research Station, Great Britain. He has more than 25 years of experience in the All India Radio, from where he retired as Engineer-In-Charge of a group of High Power Transmitters at Delhi. He then taught diploma and certificate courses at Datamatics Corporation and Disco Institute of Television Technology, New Delhi. For a number of years he worked as a visiting Professor in electronics at Delhi College of Engineering and Netaji Subhash Institute of Technology, New Delhi.

Mr Sharma's special blend of technical knowledge and practical experience are reflected in this valuable book. He is also the author of the book *VCR—Principles, Maintenance and Repair* published by Tata McGraw-Hill.

BASIC RADIO AND TELEVISION

SECOND EDITION

S P Sharma
Formely Senior Engineer
All India Radio

Tata McGraw-Hill Publishing Company Limited
NEW DELHI

McGraw-Hill Offices
New Delhi New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

Information contained in this work has been obtained by Tata McGraw-Hill, from sources believed to be reliable. However, neither Tata McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither Tata McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that Tata McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

Tata McGraw-Hill

© 2003, 1987, Tata McGraw-Hill Publishing Company Limited

No part of this publication can be reproduced in any form or by any means without the prior written permission of the publishers

This edition can be exported from India only by the publishers,
Tata McGraw-Hill Publishing Company Limited

ISBN 0-07-047335-8

Published by Tata McGraw-Hill Publishing Company Limited
7 West Patel Nagar, New Delhi 110 008, typeset in Times at The Composers,
20/5 Old Market, West Patel Nagar, New Delhi 110 008 and printed at
S P Printers, E-120, Sector-7 NOIDA

Cover: Meenakshi

RADYCRCDRQXXQ

*The **McGraw-Hill** Companies*

To
my wife ***Vimla***
and
daughter ***Rekha***

Preface to the Second Edition

The first edition of the book has been completely revised and brought up to date to present the latest 'state of the art technology' in modern electronics and communication system including Radio, TV and Video Technology.

In order to make the book completely solid state oriented, the chapter on vacuum tubes has been eliminated and all circuits using vacuum tubes have been replaced by equivalent circuits using solid state devices. However, in view of the fundamental importance of vacuum tubes and their continued use in some special types of electronic equipment, a brief description of vacuum tubes and their characteristics has been provided in the form of an appendix at the end of the book.

A new format has been adopted by dividing the revised text into six parts each containing chapters relevant to a particular topic, namely Introduction, Semiconductor Devices and Circuits, Radio Transmission and Reception, Monochrome and Colour Television, Audio and Video Systems, and Electronic Measuring Equipments, Maintenance and Troubleshooting. Different parts of the book and the order of chapters in each part has been so arranged as to facilitate a progressive and systematic study of the various topics dealt with in the book.

For better comprehension and proper understanding of the subject by the students, chapter summary and review questions are provided at the end of each chapter as in the first edition. The scope and number of Review Questions has been enlarged by including more numerical problems (with answers) and adding objective type of questions like 'True or False' questions. The number of solved examples in each chapter has also been increased.

A large number of illustrations, properly labelled sketches, and block diagrams have been used to aid the users of the book in understanding the text fully.

The subject of Television has been very comprehensibly dealt with by expanding the portion on Colour Television to include relevant topics such as Satellite communication, Cable TV, Teletext, Video Recording, (VCR) and TV games, closed circuit TV, etc.

A new chapter on Digital Electronics has been added keeping in view the importance of digital electronics in modern electronic technology.

A separate chapter has been provided on troubleshooting, maintenance and repair of electronic equipment for the benefit of those engaged in the maintenance and repair of electronic equipment.

Topics like Safety Precautions and Internet have been included in the form of appendices at the end of the book.

The reader-friendly character of the book has been maintained by the use of simple language and avoiding higher mathematics. The book is particularly helpful for those who have no previous knowledge of electricity and electronics.

The revised edition of the book is unique in providing under one cover all that is latest on radio, audio and video technology.

In revising the book, suggestions and comments received from a large section of teachers, engineers and technicians engaged in the professions have been implemented to the maximum extent possible.

Further suggestions of improvement are most welcome.

S P SHARMA

Preface to the First Edition

This book is intended to provide an elementary course in radio and television technology for students who have no previous knowledge of electricity and electronics. It has been specially designed to take the student step-by-step from the principles of electricity and electronics to the application of these principles to modern radio and television techniques. The approach is both logical and progressive, so that one topic leads to another in a smooth and systematic manner. Written in very simple and easy-to-understand language, the text can be easily grasped by a reader of average ability. Mathematics has been held to a minimum to make the book useful to as wide a range of readers as possible. A large number of carefully devised illustrations and circuit diagrams inserted at suitable places lend greater clarity to the text.

The book is aimed at providing a complete and comprehensive coverage to the syllabus approved by the Central Government for trainees in the trade of Mechanic (Radio and Television) being trained at the Industrial Training Institutes and other recognized centres under the Craftsmen Training Scheme of the Government. A thoroughly practical approach, based on the use of indigenous components and circuits, should make the book not only suitable as a textbook for ITI students but also provide an ideal training manual for all other institutions engaged in training raw hands in the art of ratio, TV and audio servicing. Students of communication engineering at Diploma level will also find the book interesting and informative.

The book can be broadly divided into three main parts.

The first part, comprising Chapters 1 to 11, explains the basic principles of electricity and magnetism together with the properties of important circuit components such as resistors, coils, capacitors, vacuum tubes and transistors. The use of simple meters, explained in Chapter 7, has been introduced at a fairly early stage to help in practical work.

The second part, consisting of Chapters 12, 13 and 14 describes the characteristics of amplifiers, oscillators and power supply systems which form the building blocks of all electronic circuits.

The third part includes Chapters 15 to 21 which are devoted to radio transmission and reception and the application of electronic circuits to radio and TV technology. This part provides useful information on the maintenance, servicing and fault-finding procedures for radio, TV and other common electronic equipment.

Colour television has been given a special coverage in the book because of its growing popularity in India. Chapter 21 has been entirely devoted to the principles and practice of colour television with special reference to the PAL system adopted in India.

Chapter organization in the book is such that the book can prove equally useful for self-study and for classroom teaching. Each chapter is divided into a number of short paragraphs carrying bold and prominent headings indicating the topic under discussion. This will help the student select at a glance any desired topic for study. A short summary at the end of each chapter provides a review of the key concepts developed in the chapter. Each chapter ends with a set of test questions including numerical problems with answers, meant for self-examination by those preparing for various tests and examinations.

The book is supplemented by a set of five appendices containing additional information and technical data. Appendix D on *soldering* and Appendix E on *printed circuit boards (PCB)* should prove particularly useful for service mechanics and technicians.

With so much information and technical material packed in one volume, the book will adequately meet the requirements of students, instructors and technicians alike.

Suggestions for improvement will be gratefully accepted.

S P SHARMA

Acknowledgements

It gives me great pleasure to express my grateful thanks to the following firms and organisations for their kind permission to publish circuit diagrams, photographs and other data so generously supplied by them.

Weston Electronics, New Delhi
Philips India Ltd., New Delhi
Jetking Kits India, Bombay
Moraj Electronic Industries, New Delhi
Bharat Electronics Ltd., Bangalore
Director General, Doordarshan, New Delhi

I am also thankful to the authors and publishers of the following works which were consulted during the preparation of the book.

Dhake, Arvind M, *Television Engineering,* Tata McGraw-Hill Publishing Co. Ltd., New Delhi, 1979.

Grob, Bernard, *Basic Electronics,* Fourth Edition, International Student Edition, McGraw-Hill Kogakusha Ltd., Tokyo, 1977

Grob, Bernard, *Basic Television: Principles and Servicing,* Fourth Edition, International Student Edition, McGraw-Hill Kogakusha Ltd., Tokyo, 1975

Huston, Geoffrey H., *Colour Television Theory,* McGraw-Hill Book Co. (UK) Ltd., 1971

Kevir, Milton S. and Milton Kaufman, *Television Simplified,* Seventh edition, Van Nostrand Reinhold Company, 1973

Marcus, Williams and Alex Levy, *Elements of Radio Servicing*, Tata McGraw-Hill Publishing Co. Ltd., New Delhi, 1974

Mileaf, Harry, *Electricity,* 1-7, Heydon Book Co. Inc., New Jersey, 1966

Terman, F.E., *Electronic and Radio Engineering,* Fourth Edition, Asian Students Edition, McGraw-Hill Kogakusha Co. Ltd., Tokyo, 1955

Zbar, Paul B., *Basic Electronics*, A Text-Lab Manual, Fourth Edition, Tata McGraw-Hill Publishing Co. Ltd., New Delhi, 1977

Zbar, Paul B., *Basic Television—Theory and Servicing,* A Text-Lab Manual, Fourth Edition, Tata Mc-Graw-Hill Publishing Co. Ltd., New Delhi, 1973

Kennedy George, *Electronic Communication Systems*, Third Edition, McGraw-Hill International Edition.

Mal-Vino, Albert Paul, *Digital Computer Electronics,* Second Edition, Tata McGraw-Hill Publishing Co. Ltd., New Delhi 1983.

William H-Gothmann, *Digital Electronics*, Prentice Hall Inc, 1977.

Clyde N. Henderric, *Electronic Troubleshooting*, Second Edition, Reston Publishing Company Inc—1977.

Gulati R.R., *Monochrome and Colour Television*, Wiley Eastern.

Khandpur R.S., *Modern Electronic Equipment—Troubleshooting, Repair and Maintenance,* TMR, New Delhi.

Gupta R.G., *Audio and Video Systems,* Tata McGraw-Hill Publishing Company Ltd, New Delhi.

Dhir S.M., *Electronic Components and Materials*, Tata McGraw-Hill Publishing Co. Ltd., New Delhi.

Sharma S.P., *VCR—Principles, Maintenance and Repair*, TMH, New Delhi.

Finally I express my sense of gratitude and deep appreciation for the excellent work done by my daugther Rekha Sharma, Principal, Bal Bharati Public School, Rohini in checking the manuscript and illustrations.

S P Sharma

Contents

PART I

INTRODUCTION

chapter 1

Introduction to Radio Communication Systems

Radio communication has made wonderful progress over the past 100 years. Starting with the early experiments on radio waves performed by Hertz and Marconi, the present state of the art in Radio Communication technology is a big leap forward in the Radio Communication field.

This chapter contains a brief review of the various topics which will be discussed in detail in other chapters of this book.

1.1 RADIO BROADCASTING

Sound broadcasting for entertainment and educational purposes constitutes the biggest network of radio communication for circuits all over the world including India. Amplitude Modulation (AM) Broadcast transmitters will continue to provide the basic service for national as well as international large area coverage of broadcasting by the use of high power Long wave, Medium wave and Short wave transmitters. The new generation of AM transmitters, either fully transistorised for Medium Power Low Frequency (LF) and Mid Frequency (MF) service; or single output tube type for short wave and high power MF and LF broadcasting, based on low level digital processing and control, will prove much reliable and economical service both from the consumption and operational points of view.

Use of Frequency Modulation (FM) services for internal broadcasting has added another dimension to the information and broadcasting service.

1.2 TELEVISION

Rapid development of monochrome and colour television has completely revolutionized the entire field of audio visual technology and has made television one of the most effective means of entertainment, education and research. Helped by such devices as Video Casette Recorder (VCR) (or video tape

recorder (VTR)) and Closed Circuit Television (CCTV), TV finds wide application in medical, scientific and industrial fields. Cable TV combined with satellite communication system has made TV programmes available to wide variety of subscribers.

1.3 SATELLITE COMMUNICATION

External coverage of TV programmes by satellites to replace the use of microwave links and coaxial cables has made the relay of international sports events and distribution of national programmes over large areas more economical and technically viable. Satellite communication is useful for many other applications such as telephone circuits, weather forecasts and remote sensing, etc.

1.4 DIGITAL ELECTRONICS

Progressive replacement of analog technology by digital electronics and the rapid growth of computer aided technology through digital electronics has resulted in considerable improvement in the quality and control of Radio, Video and Audio programmes.

1.5 SOLID STATE ELECTRONIC DEVICES

Solid state devices like transmitters and Integrated Circuits (ICs) have almost completely replaced vacuum tubes in all electronic equipment including the equipment used for Radio and TV Broadcasting and reception. This has greatly improved the quality and performance of these circuits. Although the entire electronic industry has become solid-state oriented, but the continued use of vacuum tubes in certain specialized equipments like High Power Broadcast transmitters, makes some basic knowledge of vacuum tubes and their circuits useful, though not essential, for a modern electronic engineer.

1.6 AUDIO AND VIDEO RECORDING SYSTEMS

Both audio and video recording and playback techniques have considerably improved particularly with the advent of tape recording. Extensive use of audio and video recordings in Radio and TV broadcasting has not only resulted in improved quality of these broadcasts but has also brought about better control and care of operations in these broadcasts. Video and audio tape recorders have become one of the most popular consumer items in electronic industry. Improved designs in microphones and loudspeakers has not only improved the quality of PA systems but has helped in the acoustic design of theaters, auditoria and cinema halls. Use of digital recording techniques has popularised the use of CD (Compact Discs) for long play quality recordings.

1.7 FUTURE TRENDS

Use of solid state scanning devices to replace the vacuum type picture tube and the introduction of HDTV to improve the quality of pictures are some of the

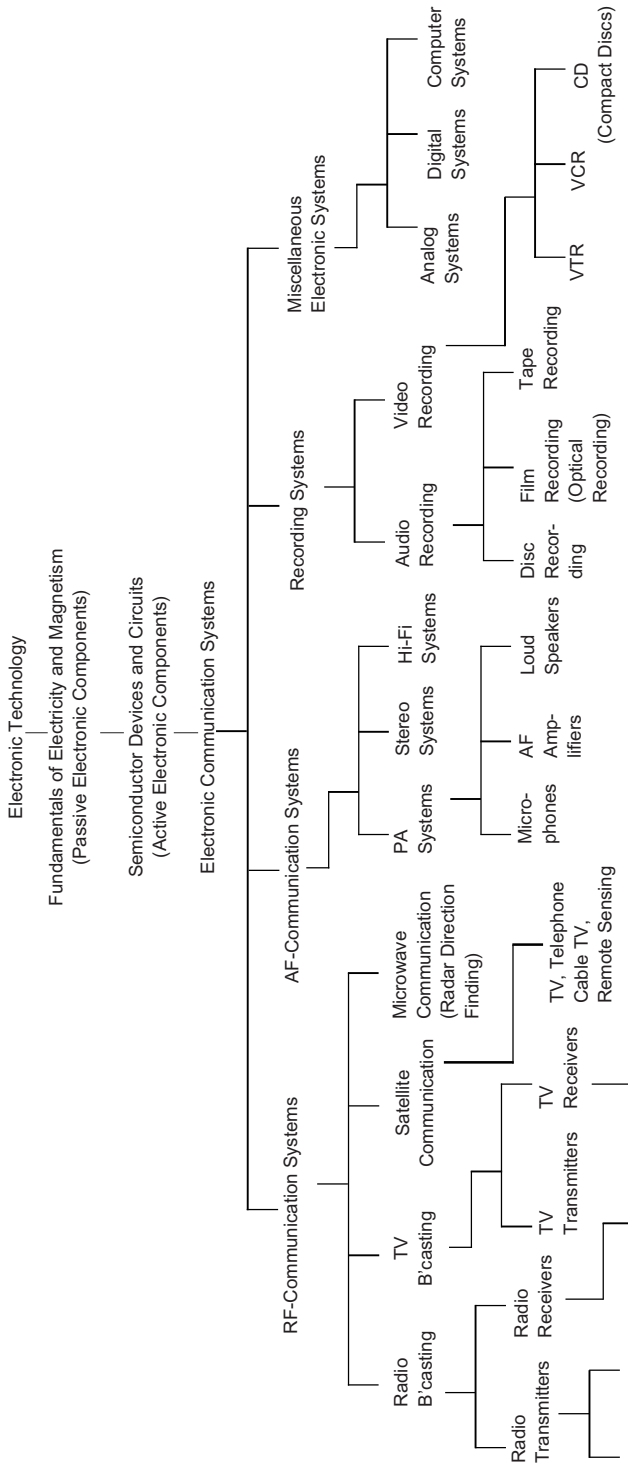


Fig.1.1 Radio Communication System

new trends in the design of TV sets. Use of stereophonic sound and 3-D pictures in the telecasts are almost round the corner. Teletext, Video text and FM Radio paging are some of the latest trends in future Radio and TV broadcasting.

Figure 1.1 provides a view of the Radio Communication System and its offshoots which will be discussed in this book.

Summary

There has been wonderful progress in Radio Communication over the past 100 years. This chapter contains a brief review of the various topics which will be discussed in this book. These topics include Radio broadcasting, Television, Satellite Communication, Digital Electronics, Solid state electronic devices, Audio and Video Recording Systems. Use of solid state scanning devices, HDTV, Teletext, Video text and FM Radio paging are some of the latest trends in future Radio and TV Broadcasting.

Review Questions

1. Give a brief review of the various topics under which progress has been made in Radio Communication.
2. What are the future trends in Radio and TV broadcasting?
3. Give in a tabular form the Radio communication and allied subjects discussed in this book.

chapter 2

Fundamentals of Electricity and Magnetism

2.1 ELECTRICITY—STATIC ELECTRICITY

The word electric is derived from the Greek term *electron* which means amber. An amber rod when rubbed with a piece of fur or cat skin acquires the property of attracting small pieces of paper or straw and is said to be charged with electricity. Similarly, a glass rod rubbed with a piece of silk cloth also gets charged with electricity. However, the polarity of the charge of electricity acquired by a glass rod when rubbed with silk is opposite to the polarity of the charge acquired by an amber or ebonite rod when rubbed with fur or cat skin. The charged glass rod has positive charge and the charged ebonite rod has negative charge. It can further be shown that a body charged positively attracts another body charged negatively and repels a body charged positively. This is illustrated in Fig. 2.1 where a positively charged glass rod attracts a negatively charged ebonite rod suspended freely with a silk thread but repels a positively charged glass rod. This can be explained by saying that opposite charges attract and like charges repel each other.

Electricity produced by friction or by rubbing two bodies with each other is called static electricity. It is a form of energy like heat, light, sound and mechanical energy. Any of these forms of energy can be changed into electricity or vice versa. In fact, static electricity is produced as a result of the conversion of mechanical energy due to friction into electrical energy.

2.2 ELECTRIC CHARGE AND ELECTRON

According to the modern concept of electricity, the production of electric charge can be explained in terms of the transfer of a tiny and invisible particle from one body to another called an *electron* which carries one unit of negative charge. Another basic particle which carries one unit of positive charge is

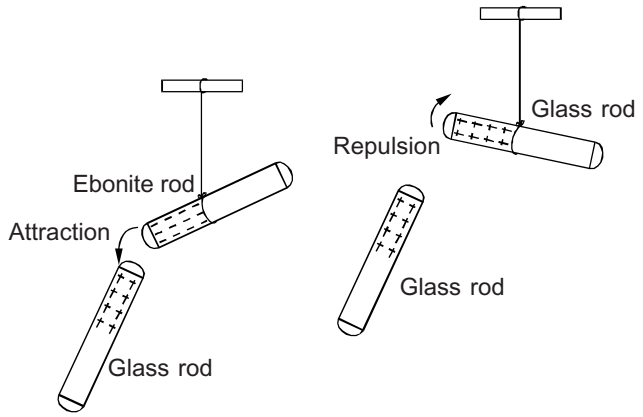


Fig. 2.1 Attraction and Repulsion between Charged Bodies

called a *proton*. All matter in whatever form, solid, liquid or gas, basically consists of electrons and protons. When electrons and protons exist in equal numbers in a body, it exhibits neither positive nor negative polarity and the body is said to be neutral. When electrons are removed from one body and transferred to another body the two bodies get opposite charges. The body from which the electrons are removed is left with an excess of protons and is said to be positively charged. The body to which the electrons have been transferred gets an excess of electrons and is said to be negatively charged. In rubbing a glass rod with a silk cloth, electrons are transferred from the glass rod to the silk cloth thereby producing an excess of electrons in the silk cloth which gets a negative charge. The glass rod which loses electrons gets an excess of protons or a deficit of electrons and thus has a positive charge. If the glass rod and the silk cloth are again put in contact with each other, as in Fig. 2.2, the electrons from the silk cloth return to the glass rod and both become neutral again.

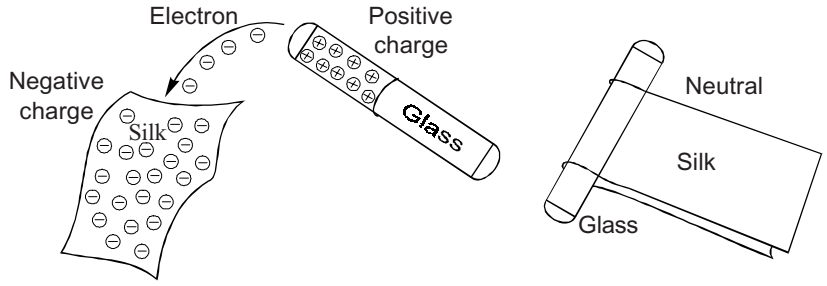


Fig. 2.2 How Bodies get Charged and Neutralised

For the transfer of electrons from one body to another, some force or energy in the form of heat, light, chemical or mechanical energy is necessary. Thus, a source of electricity has to be devised for converting any type of energy into electric energy or electricity.

Electrons and protons being the basic constituents of all types of matter, it is the arrangement of these charged particles in a substance that determines the electrical characteristics of that substance. In order to understand the electrical properties of different types of matter, it is necessary to have some idea of the structure of matter in terms of its basic constituents.

2.3 STRUCTURE OF MATTER

Matter can exist in the form of solid, liquid or gas as in the case of stone, water and air. The smallest particle of matter that can retain the properties of the original matter is called a *molecule*. A molecule can be split up into still smaller particles called *atoms* which are the basic building blocks for all matter. Matter exists in the form of elements and compounds.

Elements: A molecule may consist of a group of two or more atoms. If all atoms in the molecule of a substance are identical, the substance is called an element. Carbon, silver, hydrogen, oxygen and copper are examples of elements. There are about 106 known elements of different types. An atom is the smallest part of an element which will retain the properties of the element.

Compounds: If the molecule of a substance consists of two or more atoms which are not identical, the substance is called a compound. Water (H_2O) is a compound because it consists of atoms of hydrogen and oxygen which are not similar. In the same way common salt or sodium chloride (NaCl) is a compound which can be broken into dissimilar atoms of sodium and chlorine. Compounds are formed by a chemical combination of elements. The large number of elements that exist in nature can combine with each other and form any number of compounds.

An atom is the simplest particle of matter, but it contains within itself three fundamental particles known as *electrons, protons* and *neutrons*. As already stated, electrons are negatively charged particles with a unit negative charge and protons are positively charged particles with a unit positive charge but protons are much heavier than electrons. A proton is nearly 1840 times heavier than an electron. A neutron is an electrically neutral particle which is almost as heavy as a proton. It is the number and the arrangement of these particles in an atom that determine its physical and chemical properties.

2.4 ATOMIC STRUCTURE

The three important particles constituting an atom are electrons, protons and neutrons. These particles can combine in any number of ways to form an atom but there is a specific arrangement that gives rise to a stable structure called the atomic structure. Each stable arrangement of these fundamental particles makes a particular type of atom. A different stable arrangement of electrons, protons and neutrons will make a different type of atom with different physical and chemical characteristics.

The best picture of an atom that explains satisfactorily most of the physical and chemical properties of known elements was provided by a Danish scientist,

Neils Bohr. According to Bohr, the structure of an atom is similar to our planetary system in which planets like the Earth, Venus and Mars revolve round the sun in different orbits. In the case of an atom, there is a central part called the nucleus which contains all the protons with their positive charge. The electrons with their negative charge revolve round the nucleus in rings or orbits at different distances from the nucleus. In the neutral state of an atom the number of electrons in orbit is exactly equal to the number of protons in the nucleus so that the negative and the positive charges cancel each other. The force of attraction between the positive charge at the nucleus and the negative charge in the orbit is balanced by the mechanical force acting outward on the rotating electron. This makes the electron stable while rotating in its orbit. Neutrons which are neutral particles also form a part of the nucleus and though they have no electrical properties they add to the weight of the atom.

The simplest case of atomic structure is provided by the hydrogen atom. It consists of one proton in the nucleus and one electron revolving around it as in Fig. 2.3(a). The next possible arrangement is with two protons at the nucleus and two electrons orbiting round it as shown in Fig. 2.3(b). This is the atom of the element helium.

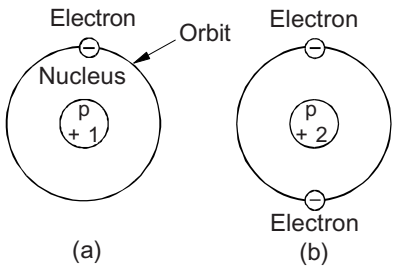


Fig 2.3 Atomic Structure (a) Hydrogen (b) Helium

When there are more than one electron and one proton in an atom, all the protons cling together in the nucleus like a bunch of grapes but the electrons revolve round the nucleus in one or more orbits at varying distances from the nucleus. The number of electrons in the orbit, however, is always equal to the number of protons in the nucleus for a neutral atom. The orbits of the revolving electrons are also called *shells* or energy levels. These successive shells from the nucleus outwards are known as K, L, M, N, O and P shells. Each of these shells can accommodate a maximum number of electrons for stability which is given by the formula $2 \times n^2$ where n = 1, 2, 3, 4, etc. respectively for shells, K, L, M, N and so on. For a K shell, n = 1 and the maximum number of electron for K shell is 2. This is the case for helium as shown in Fig. 2.3(b). The atomic structure of an element like helium which has all its orbits full upto the maximum number of electrons they can contain is that of an inert gas. Neon is another example of an inert gas having a maximum number of 8 electrons in the L shell (n = 2) as shown in Fig. 2.4

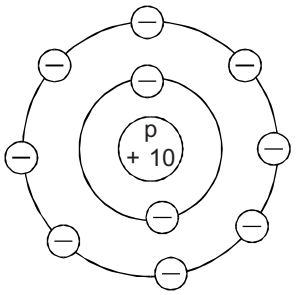
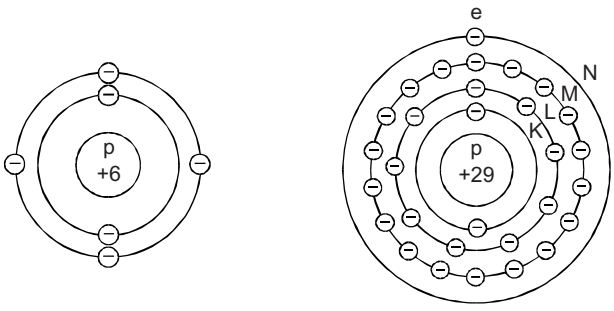


Fig. 2.4 Neon Atom

After any shell is complete with the maximum number of electrons it can contain, the remaining electrons arrange themselves in higher orbits—the maximum number of electrons in the K, L, M, N,



(a)
Carbon atom—6 electrons
K = 2 electrons (complete)
L = 4 electrons (incomplete)

(b)
Copper atom—29 electrons
K = 2 electrons (complete)
L = 8 electrons (complete)
M = 18 electrons (complete)
N = 1 electron (incomplete)

Fig. 2.5 Carbon and Copper Atoms

orbits being 2, 8, 18, 32, etc. However, the maximum number of electrons in the outermost shell does not exceed 8. Figures 2.5(a) and (b) show the structure of carbon and copper atoms which have 6 and 29 electrons, respectively arranged in various orbits. It will be seen that the carbon atom has four electrons in the outermost ring against the maximum number of 8 and shows a stable structure. In the case of copper, however, there is only one electron in the outermost ring which is not very stable. This electron is free to jump from atom to atom and hence called a free electron. This forms a carrier of charge or electricity.

2.5 ATOMIC NUMBER AND ATOMIC WEIGHT

The atomic number of an element is given by the total number of electrons revolving round the nucleus or by the total number of protons in the nucleus. The atomic number of carbon is 6 and that of copper is 29. The atomic weight of an atom is the total weight of the atom compared to the weight of the hydrogen atom. The atomic weight of an atom is indicated by the weight of protons plus the weight of neutrons in the nucleus. Thus a carbon atom with an atomic number of 6 has an atomic weight equal to 12 (6 protons and 6 neutrons). This simply means that a carbon atom is 12 times as heavy as an atom of hydrogen which has an atomic weight of 1.

2.6 CONDUCTORS, INSULATORS AND SEMICONDUCTORS

A study of the atomic structure of various elements shows that certain metals like copper have a single free electron in their outermost shell which can easily move from atom to atom and thus allow an easy flow of electricity through them. Such materials are called *conductors*. In general, all metals are good

conductors but certain metals are better conductors than others. Silver, copper and aluminium are examples of good conductors. Silver is a better conductor than copper but copper is mostly used as a conductor for wires because it is less expensive than silver. Similarly, copper is a better conductor than aluminium but the latter is being freely used as a conductor for electrical wiring because the cost of copper also has gone up considerably compared to the cost of aluminium. Because of their property to allow current to flow through them with a minimum of opposition, the conductors are used for making various types of cables and parts of electrical equipments and accessories like plugs, sockets and switches.

Substances like glass, mica and rubber do not have a large number of free electrons in their atoms and so do not allow current to flow through them easily. Such substances are called *insulators*. As in the case of conductors certain substances are better insulators than others. Mica is a better insulator than rubber because it has fewer free electrons than rubber. Because they do not allow electricity to flow through them easily, insulators are used to separate conductors in electrical appliances so that electricity does not jump from one conductor to another. In fact, insulators form as important a part of electrical equipment and appliances as conductors themselves.

Conductors and insulators are relative terms. There is no sharp dividing line between conductors and insulators. It is only a question of how freely the electrons can move about in a material. No material is a perfect conductor and no material is an absolute insulator. A good insulator is a bad conductor and a bad insulator is a good conductor.

While conductors are useful for allowing the flow of electricity from one point to another, insulators are able to store electricity in them and are employed as dielectrics for capacitors as will be discussed later.

Given below is a list of common conductors and insulators in the order in which they conduct or oppose the flow of electricity through them:

Conductors	*Insulators*
Silver	Dry air
Copper	Glass
Aluminium	Mica
Zinc	Rubber
Brass	Asbestos
Iron	Bakelite

There is also a class of materials like silicon, germanium and carbon in which the electrons are neither very free to move about nor completely tied down to the atom. These materials are neither conductors nor insulators and are known as *semiconductors*. They partially allow the flow of electricity through them. These elements have four electrons in their outermost orbit against the maximum number of 8 which makes these orbits neither completely stable nor very unstable. Semiconductors form the basic material for the construction of transistors and other semiconductor components which have revolutionised the

world of electronics because of the many advantages transistors possess over their rival electronic components—the vacuum tubes.

2.7 CURRENT, VOLTAGE, RESISTANCE AND POWER

2.7.1 Current

A neutral body can be charged with electricity by either removal of electrons or addition of electrons to the atoms of the neutral body. The amount of charge, positive or negative, depends on the number of electrons that are removed or added to the body. The unit for the measurement of charge is called *Coulomb*. It is equal to the charge on 6.28×10^{18} electrons (or protons). A coulomb is a practical unit of charge. When the free electrons in a conductor like copper are made to move in a particular direction in the conductor, this flow of electrons constitutes the electric current which is defined as the rate of flow of charge. The practical unit of electric current is ampere (A) which is equal to one coulomb per second (C/s). In terms of the movement of electrons, a current of one ampere flowing through a conductor will mean the movement of 6.28×10^{18} electrons in one second across any section of the conductor.

Smaller units of current in practical use are milliamperes (mA) and micro amperes (μA):

$$1\text{A} = 1000\ \text{mA} = 10^3\ \text{mA}$$
$$1\text{A} = 1{,}000{,}000\ \mu\text{A}$$
$$= 10^6\ \mu\text{A}$$

The alphabet I is generally used to represent current. The direction of current is shown by arrows.

EXAMPLE: How many electrons will flow through a wire in 5 s if the current flow is 10 mA?

$$1\text{A} = 1000\ \text{mA} = 10^3\ \text{mA}$$
$$1\text{A} = 6.28 \times 10^{18}\ \text{electrons per second}$$
$$1\text{m A} = \frac{6.28 \times 10^{18}}{10^3} = 6.28 \times 10^{15}\ \text{electrons per second}$$
$$10\text{mA} = 6.28 \times 10^{15} \times 10 = 6.28 \times 10^{16}\ \text{electrons per second}$$

Number of electrons flowing in 5 s

$$= 5 \times 6.28 \times 10^{16} = 31.40 \times 10^{16}$$

2.7.2 Voltage

Electrical current flows through conductors or wires in the same way as water flows through pipes. In the water circuit shown in Fig. 2.6(a) water flows through the pipes due to a difference of pressure between the input side and the discharge side of the water pump while the amount of water flowing through the pipes depends on the difference of pressure at the two ends of the water pump. In the electrical circuit shown in Fig. 2.6 (b) the electrons are made to

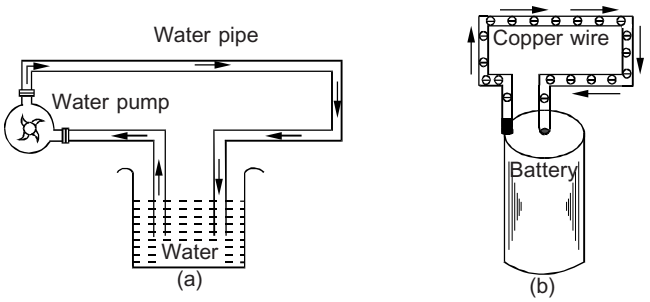


Fig. 2.6 Analogy between Water Flow and Electric Current (a) Water Flow and (b) Current

move in the copper wire by the electrical pressure provided by the battery. This electrical pressure or electromotive force is provided between the negative and positive poles of the battery due to the chemical action inside the battery cell. The negative pole which develops an excess of electrons repels the free electrons in the copper wire and the positive pole which develops deficiency of electrons attracts the free electrons to its side and a current starts flowing through the wire in the same way as water flows through the pipes. The greater the difference of electrical pressure between the negative and positive poles of the battery the greater will be the current flow. This difference of electrical pressure which has been referred to as electromotive force (emf) is also known as the *voltage* or *potential difference* (pd), and the unit for its measurement is termed a volt (V). This is a practical unit for the measurement of voltage, pd or emf. Units larger than a volt are called kilovolt (kV) and megavolt (MV).

$$1 \text{ kilovolt} = 1000 \text{ V or } 10^3 \text{ V}$$

$$1\text{MV} = 1000{,}000 \text{ V or } 10^6 \text{ V}$$

We also come across units of voltage which are smaller than a volt. These are millivolts (mV) and microvolts (μV).

$$\text{IV} = 1000 \text{ mV} \quad \text{or} \quad 10^3 \text{ mV}$$

$$1\text{V} = 1000{,}000\ \mu\text{V or } 10^6\ \mu\text{V}$$

The meter used for the measurement of voltage, pd or emf is called a Voltmeter. The letter E is used to denote voltage, pd or emf. The source of voltage is represented by a cell or a number of cells (battery) as shown in Fig. 2.7.

EXAMPLE: A current of 10 mA flows through a wire when a voltage of 100 mV is applied across it. Find the current that will flow through the wire when a voltage of 1.5 V is applied to it.

$$1.5 \text{ V} = 1.5 \times 10^3 = 1500 \text{ mV}$$

Current through the wire with 1500 mV

$$= 10 \times \frac{1500}{100} = 150 \text{ mA}$$

$$= \frac{150}{1000} = 0.15\text{A}$$

2.7.3 Resistance

The free electrons in a copper wire start moving under the influence of the potential difference applied, resulting in the flow of current. The free electrons do not, however, start moving in a particular direction without offering any opposition. The opposition to the flow of current is called *resistance*. In a material like copper, the atoms have a large number of free electrons and the resistance is less. Compared to this, carbon has fewer free electrons and the resistance will be greater than in the case of copper. Conductors have less resistance than insulators. The unit of resistance is ohm (Ω). A wire has a resistance of 1 Ω when a potential difference of 1 V applied across it makes a current of 1 A flow through it. The bigger practical units of resistance are kilohm (kΩ) and (MΩ).

$$1k\Omega = 1000\ \Omega \quad \text{or} \quad 10^3\Omega$$

$$1M\Omega = 1000{,}000\ \Omega \quad \text{or} \quad 10^6\ \Omega$$

Resistance is denoted by the letter *R*. When a resistance *R* is connected to a source of voltage *E* by copper or any conductor, a current *I* flows through this resistance as shown in Fig. 2.7. Such an arrangement is called an electrical circuit.

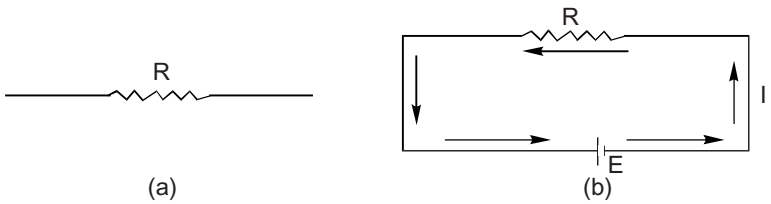


Fig. 2.7 Electrical Circuit

2.7.4 Power

When a voltage is applied to a conductor, the electrons start moving in a particular direction and a current flows through the conductor. The electrons collide with other atoms of the conductor and produce heat. The amount of heat developed depends on the applied voltage and the current. The production of heat indicates that energy is consumed in the process of the flow of current and work is done to overcome the friction due to the movement of the electrons. The amount of work done in one second is called *power*. It is equal to the product of current *I* and voltage *E* in a particular circuit. Thus

$$\text{Power } (P) = E \times I \tag{2.1}$$

The unit of power is watt (W). In the above formula *P* is equal to 1 W when $E = 1$ V and $I = 1$ A. Thus the power consumed in a circuit is 1 W when the voltage applied is 1V and the current flowing through the circuit is 1 A.

2.8 OHM'S LAW

The current flowing through a conductor increases when the electrical pressure

or voltage applied across the conductor increases and the current decreases if the voltage decreases. In other words, the current flowing through a conductor is directly proportional to the voltage, provided the resistance of the conductor is constant. Similarly, for the same voltage, the current decreases when the resistance of the conductor increases and the current increase when the resistance of the conductor decreases. In other words, the current is inversely proportional to the resistance when the voltage is constant. Thus a definite relationship exists between the applied voltage E and the current I that flows through a conductor with a fixed resistance R. This relationship was first established by George Simon Ohm and is known as *Ohm's* law after its discoverer. Ohm's law is expressed by the mathematical formula:

$$I = \frac{E}{R} \tag{2.2}$$

where I is in amperes, E is in volts and R is in ohms. By a slight mathematical manipulation, this formula can be expressed in two other forms:

$$E = I \times R \tag{2.3}$$

$$R = \frac{E}{I} \tag{2.4}$$

Given any two of these quantities (E, I, R), the third one can be found out by any of the three relations given above.

To determine current (I) *when voltage* (E) *and resistance* (R) *are known*: The formula to be used in this case is Eq. (2.2).

Thus in the circuit in Fig. 2.8 the applied voltage E = 12 V and resistance R = 6 Ω; so the current I is given by:

$$I = \frac{12V}{6\Omega} = 2\text{A}$$

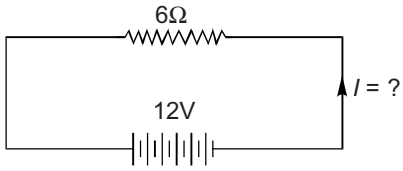


Fig. 2.8 Application of Ohm's law to (find I when E and R are known)

Current can be expressed in milliamperes or microamperes when the resistance is in kilohms and megaohms.

EXAMPLES: Find the current flowing through a circuit in which the applied voltage is 12 V and the resistance of the circuit is

(i) 6 k Ω; (ii) 6M Ω

$$I = \frac{E}{R}, \quad E = 12 \text{ V}$$

(i) when the resistance is 6 kΩ

$$I = \frac{12V}{6k\Omega} = \frac{12V}{6000\Omega} = 0.002 \text{ A or } 2 \text{ mA}$$

(ii) when the resistance is 6 MΩ

$$I = \frac{12V}{6M\Omega} = \frac{12V}{6000000\ \Omega} = \frac{12V}{6 \times 10^6\ \Omega}$$

$$= 2 \times 10^{-6} \text{ A or } 2\mu \text{ A}$$

To determine the voltage E when the current I and resistance R are known:

The formula applicable in this case is Eq. (2.3).

In the circuit shown in Fig. 2.9 if the current flowing is 2A and the resistance 6Ω the voltage is given by

$$E = 2\text{A} \times 6\,\Omega = 12\text{ V}$$

This voltage which appears across the two ends of the resistance is called the voltage drop and is denoted by IR.

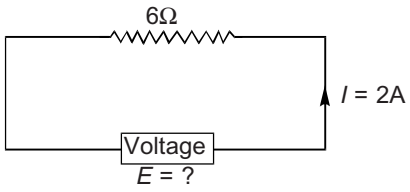


Fig. 2.9 To find *E* when *I* and *R* are Known

EXAMPLE: Three resistances of value 6Ω, 12Ω and 18Ω are connected across a voltage source as shown in Fig. 2.10. If the current flowing through the circuit is 0.5 A find the voltage drop across each resistor.

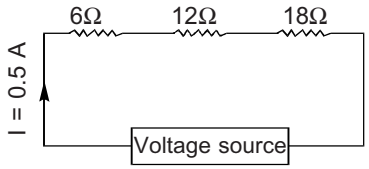


Fig. 2.10 See Example below

In this case:

I = 0.5 A; voltage drop = $I \times R$

Voltage drop across 6Ω = 0.5A × 6Ω = 3V

Voltage drop across 12Ω = 0.5A × 12Ω = 6V

Voltage drop across 18Ω = 0.5A × 18Ω = 9V

Total voltage drop across the three resistors is the sum of the individual voltage drops and is equal to the voltage of the voltage source:

Voltage of the voltage source $E = 3 + 6 + 9 = 18$ V

To determine the resistance R when the voltage E and the current I are known:

The formula to be used in this case is Eq. (2.4).

R will be in ohms, when E is in volts and I is in amperes. In the circuit in Fig. 2.11 E = 12 V, I = 2 A,

so
$$R = \frac{12V}{2A} = 6\Omega$$

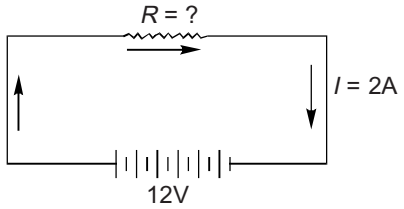


Fig. 2.11 To find *R* when *E* and *I* are Known

When the current is in milliamperes and microamperes, the resistance will be of a high value of the order of kilohms and megaohms.

EXAMPLE: Find the resistance of a circuit when the applied voltage is 12 V and the current flowing is (i) 2mA; (ii) 2μA.

$$R = \frac{E}{I}, \quad \text{here } E = 12 \text{ V}$$

(i) when $I = 2\text{mA}$

$$R = \frac{12V}{2\text{mA}} = \frac{12V}{2 \times 10^{-3}\ A}$$

$$= \frac{12}{2} \times 10^3\ \Omega = 6 \times 10^3\ \Omega$$

$$= 6\ \text{k}\Omega$$

(ii) when $I = 2\mu$ A or 2×10^{-6} A

$$R = \frac{12V}{2 \times 10^{-6}} = \frac{12}{2} \times 10^6\Omega = 6 \times 10^6\Omega$$

$$= 6\ \text{M}\Omega$$

A useful memory aid to Ohm's law is given in Fig. 2.12. In the diagram the quantity to be determined is covered with the thumb and the two exposed letters in their relative positions give the value of the covered quantity. Thus when I is covered the result is given by E/R, when E is covered the result is $I \times R$ and when R is covered its value is given by E/I. A practical form of the memory aid is given in Fig. 2.13.

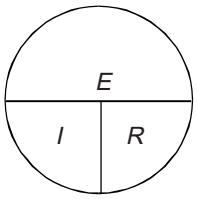


Fig. 2.12 Memory Aid to Ohm's Law

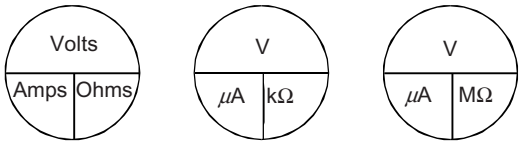


Fig. 2.13 Practical form of Memory Aid

Equation (2.1) can be expressed in two other forms by applying Ohm's law.

We know from Ohm's law that $E = IR$

$$\therefore \quad P = E \times I = I \cdot R \cdot I = I^2R \tag{2.5}$$

Similarly, from Ohm's law $I = \frac{E}{R}$ $\quad \therefore P = \frac{E^2}{R^2} \cdot R = \frac{E^2}{R}$ (2.6)

The power of an electrical appliance in watts is also known as *wattage*.

EXAMPLE: What is the wattage of an electric toaster which draws a current of 3A when used on a 230 V power supply?

$$\text{Power} = E \times I$$

where $E = 230$ V and $I = 3$ A

$$\text{Wattage} = 230 \times 3 = 690 \text{ W}$$

EXAMPLE: The resistance of an element of an electric room heater is 50Ω. How much power will it consume on a 230 V mains supply?

$$P = \frac{E^2}{R}$$

where $E = 230$ V, $R = 50\ \Omega$

$$P = \frac{(923)^2}{50} = 1058 \text{ W}$$

A watt is a small unit of electrical power. The bigger practical unit is the kilowatt (kW) which is equal to 1000 W. A still bigger unit is megawatt (MW) which is equal to 1000,000 W or 10^6 W. The units smaller than a watt are milliwatt (mW) and microwatt (μW).

$$1\text{W} = 1000 \text{ mW} \quad \text{or} \quad 10^3 \text{ mW}$$

$$1\text{W} = 1000{,}000\ \mu \text{ W} \quad \text{or} \quad 10^6\ \mu \text{ W}$$

The power of 1058 W in the above example can also be stated as 1.058 kW.

We use electrical energy for heating and lighting purposes. We have to pay the electric charges according to the amount of electrical energy consumed. Kilowatt hour (kWh) is the unit adopted for measuring the consumption of electricity. The number of kilowatt hours or units of electricity consumed can be calculated by the product of the power in kilowatts multiplied by the time in hours for which the electrical energy has been consumed.

EXAMPLE: What is the energy consumed per day in kilowatt hours in a household using 6 electrical lamps of 100 W each which are lighted for 5 hours everyday. What will the bill for energy consumption be in a month of 30 days if the rate of electric charge is 25 paise per unit?

Power of six 100 W lamps $= 6 \times 100 = 600$ W

$= 0.6$ kW

Energy consumed in 5 hours a day $= 0.6 \times 5 = 3$ kWh

Energy consumed in 30 days $= 3 \times 30 = 90$ kWh

$= 90$ units

Total charges for 90 units or 90 kWh at the rate of 25p per unit

$$= \frac{90 \times 25}{100} = \text{Rs. } 22.50$$

The power of an electrical motor is generally given in terms of a unit called horsepower (hp). 1 hp is equal to 746 W.

Thus the horsepower of a 220 V electric motor taking a current of 5 A will be

$$= \frac{220 \times 5}{746} = \frac{1100}{746} = 1.49 \text{ hp}$$

Voltage-Current-Resistance-Power Relations A practical representation of these quantities and their relationship to other quantities is given in Fig. 2.14. The four basic quantities *E, I, R* and *P* discussed earlier occupy each of the four segments in the inner circle. Adjacent to each of these inner segments are the segments of the outer circle which contains the three formulas representing that particular quantity in terms of two other basic quantities. Thus *P* in the inner segment is equal to E^2/R, I^2R and *IE* each of which occupies one of three outer segments adjacent to the inner segment containing *P*. This is another useful memory aid for the application of Ohm's Law.

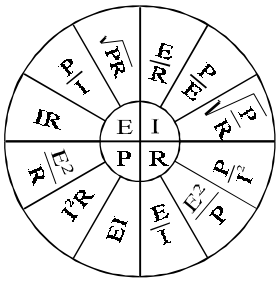


Fig. 2.14 *E, I, R* and *P* Relations

2.9 MAGNETISM

2.9.1 Properties of a Magnet

A magnet is a piece of iron or steel which has the property of attracting small pieces of iron. In its natural form, a magnet was discovered as an iron ore called magnetite. It was also called a loadstone or leading stone because it was originally used to steer ships in a proper direction while on the high seas.

Two important properties of a magnet are:

1. It attracts small pieces of iron or other magnetic materials. Nickel and cobalt are also magnetic materials.
2. When suspended freely in air by a silk thread, one end of the magnet always points towards the geographical north of the earth, also called the North pole, North-seeking pole or N-pole. The other end of the magent which points towards the geographical south of the earth is called the South-pole, South-seeking pole or S-pole.

In fact, the earth itself is a big magnet with one pole near the north geographical pole and the other near the south geographical pole. The magnetic pole of the earth located near the geographical north pole is actually the south pole of the earth magnet and it attracts towards itself the North seeking or N-pole of any magnet suspended freely.

Fig. 2.15 Magnetism is Strongest near the Poles

The ends of the magnet where attractive force is greatest, are called the poles of the magnet. In a magnet, the magnetism is concentrated near its ends or poles and its strength gradually decreases towards the centre. This can be shown by dipping a magnet in iron filings, when the thickest cluster of iron filing will be found to stick to the ends, as shown in Fig. 2.15. The force of attraction is, therefore, strongest near the poles of a magnet.

2.10 ATTRACTION AND REPULSION BETWEEN POLES

A piece of iron wire can be attracted by the north pole as well as by the south

pole of a magnet. If, however, the north pole of a magnet *A* is brought near the north pole of a freely suspended magnet *B* as in Fig. 2.16 (a), the north pole of the suspended magnet will move away and get repelled. Similarly, when the south pole of magnet *A* is brought near the south pole of the suspended magnet *B*, the result will again be repulsion of the south pole of the suspended magnet. This shows that like poles of two magnets repel each other.

Now hold magnet *A* near the north pole of the suspended magnet *B* as shown in Fig. 2.16 (b). The north pole of magnet *B* will get more attracted towards the south pole of magnet *A*; similarly, when the north pole of magnet *A* is brought near the south pole of magnet *B*, the result will again be attraction. This proves that unlike poles of two magnets attract each other. The result of the above experiments can be summed by saying:

"Like poles repel and unlike poles attract each other".

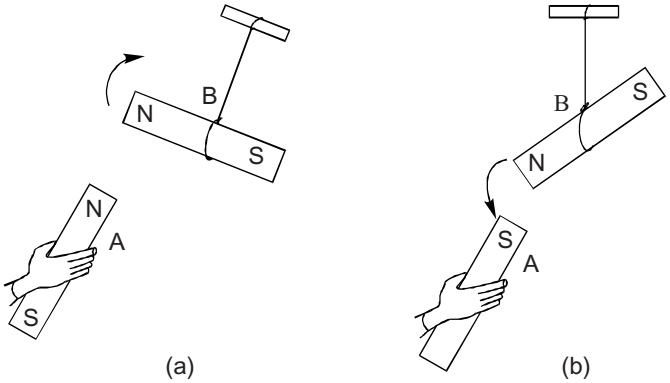


Fig. 2.16 Attraction (a) Repulsion and (b) between Poles

This force of attraction or repulsion between two magnetic poles depends on the strength of each magnetic pole and varies inversely as the square of the distance between the two poles.

2.11 MAGNETIC FIELD

The space round a magnet in which its influence can be felt is called the magnetic field. This magnetic field is the strongest near the poles of the magnet and gets weaker as we move away from the poles. A magnetic field is not visible but its existence can be shown by the effect it produces on magnetic materials like iron filings. Place a glass sheet on a bar magnet and sprinkle some iron filings on the glass sheet. Tap the glass sheet on a bar magnet and sprinkle some iron filings on the glass sheet. Tap the glass sheet gently with your fingers and the iron filings will arrange themselves in a regular pattern of lines from one pole to the other as shown in Fig. 2.17.

The lines along which the iron filings arrange themselves are called *the magnetic lines of force*. These lines of force are very crowded near the poles indicating that the magnetic field is very strong here. Each particle of iron

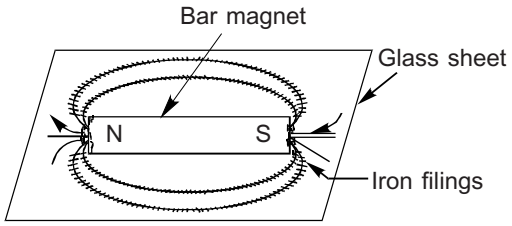


Fig. 2.17 Magnetic Lines of Force

filings becomes a tiny magnet under the influence of the magnetic field of the bar magnet and arranges itself along a certain magnetic line of force.

The magnetic field of any magnet can actually be plotted with the help of a compass needle which consists of a very small magnet pivoted inside a non-magnetic case with a glass top, (Fig. 2.18). The needle can move freely on its pivot. Its *north pole* is painted a permanent shade to distinguish it from the *south pole*.

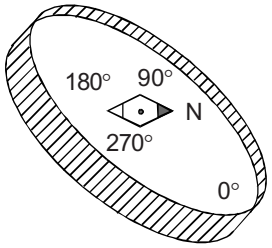


Fig. 2.18 Compass Needle

Take a sheet of paper and place a bar magnet on it. Mark the boundary of the bar magnet with a pencil. Place the compass needle near the north pole of the magnet and with a sharp pencil, mark the position of the north pole of the compass needle when it becomes steady. Move the compass needle slightly so that its centre now rests on the point marked earlier with pencil. The north pole of the compass needle will now point to a different direction. Mark with the pencil the new position of north pole. Go on shifting the compass needle from point to point, marking the position of its north pole every time till you reach the south pole of the bar magnet. Draw a smooth curve through all the points. This curve represents one magnetic line of force. Starting from another point near the north pole of the bar magnet, another magnetic line of force can be drawn. In this way a number of magnetic lines of force can be drawn on both the sides of the bar magnet and the entire magnetic field plotted as shown in Fig. 2.19.

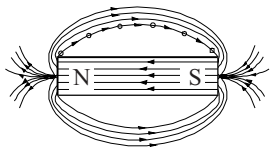


Fig. 2.19 Magnetic Field of a Bar Magnet

The magnetic field plotted in Fig. 2.19 is similar to the pattern indicated by the iron filings in Fig. 2.15. The iron filings actually arrange themselves along these magnetic lines of force. These magnetic lines of force will crowd together near the poles where the magnetic field is the strongest.

The magnetic lines of force emerge from the north pole of a magnet and passing through the surrounding medium these lines of force enter the magnet at the south pole and again emerge from the north pole. These lines are continuous curves and complete a circuit like the electron current which leaves the negative pole of a battery and passing through the conductor returns to the positive pole of the battery and through the battery back to the negative pole again.

2.12 MAGNETIC FLUX AND FLUX DENSITY

The lines of magnetic force that emerge from the north pole of a magnet comprise what is called *magnetic flux* and is generally represented by the Greek letter ϕ (phi). A strong magnet provides greater flux than a weak magnet.

Flux density B is the number of magnetic lines of force that pass through the unit area of a section perpendicular to the direction of the magnetic flux.

Mathematically, $B = \dfrac{\phi}{A}$

where A is the area through which a flux ϕ passes. Flux density B is measured in terms of a unit called gauss (G). A gauss is equal to one line (also called maxwell) per square centimeter.

EXAMPLE: A magnet produces a flux of 20,000 magnetic lines in a perpendicular area of 4 cm^2. Find the flux density in gauss.

$$\phi = 20{,}000 \text{ lines}$$

$$A = 4 \text{ cm}^2$$

$$B = \frac{20{,}000}{4} = 5000 \text{ G}$$

The earth's field strength produces a flux density B of over 0.2 G whereas a strong magnet can produce a B of 50,000 G.

Gauss is the unit of flux density in the CGS (centimeter, gram, second) system. In the practical system of units called the MKS (meter, kilogram, second) system a bigger unit of flux called weber (Wb) is used. 1 Wb is equal to 10^8 magnetic lines of force or maxwell. In the MKS system, the flux density B will be measured in webers per square meter or Wb/m^2.

2.13 MAGNETIC INDUCTION

A magnet has the property of imparting its magnetism to other magnetic materials like iron and steel. If a bar magnet is rubbed against a soft iron nail, the latter becomes a magnet temporarily but loses its magnetism after some time. Similarly, if a piece of magnetic material like soft iron is introduced in the magnetic field of a permanent bar magnet, the soft iron piece becomes a magnet without actually coming in physical contact with the bar magnet. This phenomenon of producing a magnetic effect in a magnetic material without any physical contact between the material and the magnet is known as *magnetic induction*.

The soft iron piece placed near a permanent magnet becomes a temporary magnet with one end becoming the south pole and the other end the north pole as in Fig. 2.20.

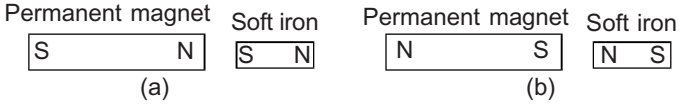


Fig. 2.20 Magnetic Induction

It may be seen that the north pole of the permanent magnet in Fig 2.20(a) induces a south pole of opposite polarity in the soft iron piece. These two opposite poles will attract each other and the smaller iron piece will be attracted by the permanent magnet. If the permanent magnet is reversed in polarity as in Fig. 2.20(b) its south pole will induce a north pole at the nearest end of the soft iron piece and these two opposite poles will again attract each other. This explains the fact that either pole of a permanent magnet attracts to itself a small piece of soft iron. In fact, a permanent magnet first converts a piece of soft iron into a temporary magnet by induction and then attracts it due to the opposite polarity induced in the closest or nearest end of the soft iron.

Magnetic induction can be explained by the Molecular Theory of Magnetism. According to this theory, each particle or molecule of a magnetic substance like iron is a small magnet with a north pole and a south pole. These molecular magnets lie at random in the unmagnetised state of the iron piece when the field of any molecular magnet is being neutralised by the opposite pole or field of the neighbouring molecular magnets. Thus, soft iron exhibits no magnetic properties in its unmagnetised state as in Fig. 2.21(a).

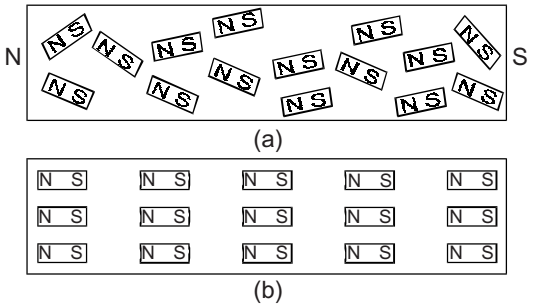


Fig. 2.21 Molecular Theory of Magnetism

When a soft iron piece is placed in the magnetic filed of a permanent magnet, the magnetic lines of force passing through the iron piece make the molecular magnets line up in a particular direction with all the south poles pointing in the direction of the north pole of the permanent magnet and all the north poles pointing in the opposite direction. Thus, the iron piece gets magnetised under the influence of the permanent magnet with the closest ends having opposite polarities. When the bar magnet is removed, the piece of soft iron will lose its magnetism because the molecular magnets will again try to lie in a random manner destroying each other's magnetism. The molecular theory gets support from the fact that a magnet loses its magnetism if it is struck with a hammer or heated over a flame. The molecular magnets get disarranged due to agitation produced by hammering or by heat.

It is also observed that soft iron loses its magnetism easily when the magnetising field is removed but steel retains magnetism for a long time even after the magnetising force is removed. This property of retaining magnetism

by a magnetic material is called retentivity. Steel has greater retentivity than soft iron and hence steel is used for making permanent magnets.

2.14 PERMEABILITY

If a piece of soft iron is placed in a magnetic field, the lines of force get concentrated in the soft iron piece as shown in Fig. 2.22. This property of a material to concentrate magnetic flux is called permeability. The flux density in a magnetic material is much more than if the same space is occupied by air or vacuum. Any material that has high permeability gets easily magnetised due to induction.

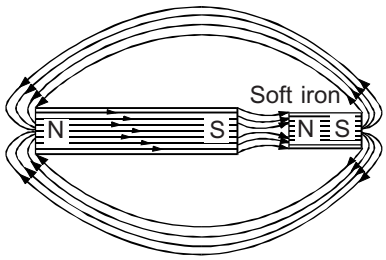


Fig. 2.22 Permeability of Soft Iron

Permeability of a material is measured relative to air or vacuum whose permeability is taken as 1 G. On this basis the relative permeability of iron and steel varies between 100 and 9000. Permeability is generally represented by the Greek letter μ (mu).

2.15 MAGNETIC MATERIALS

Magnetic materials are classified in three main groups according to the magnetic properties exhibited by them.

2.15.1 Ferromagnetic Materials

Magnetic materials that get highly magnetised in a magnetic field are called ferromagnetic materials. These materials posses high values of permeability varying from 50 to 5000. This category includes iron, steel, nickel, cobalt and certain alloys like alnico and permalloy. These are strongly attracted by a magnet.

2.15.2 Paramagnetic Materials

These materials get only weakly magnetised in the direction of the magnetic field and possess a permeability slightly greater than one. The list includes aluminium, platinum, manganese and chromium. These are weakly attracted by magnets.

2.15.3 Diamagnetic Materials

These include bismuth, antimony, copper, gold, silver, zinc and mercury. In this case the permeability is less than 1. They develop only weak magnetism but in a direction opposite to that of a magnetic field. These materials will, therefore, be actually repelled by a magnet and not attracted to it.

2.15.4 Ferrites

These are non-metallic materials which have a high permeability like iron. These are ceramic materials and unlike iron they are insulator and have a very high value of resistivity. Ferrites are used as core materials for high frequency coils and transformers because of the low losses they suffer at high frequencies.

2.16 ELECTRICAL CURRENT AND MAGNETIC FIELD

If a conductor carrying electric current is held over a magnetic needle, the north pole of the magnetic needle gets deflected at right angles to the current carrying conductors as in Fig. 2.23(a).

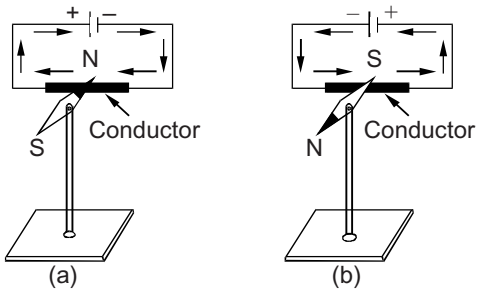


Fig. 2.23 Magnetic Effect of a Current carrying Conductor

2.16.1 Magnetic Effect of a Current Carrying Conductor

If the direction of the current is reversed, the north pole of the magnetic needle gets deflected in the opposite direction as in Fig. 2.23(b). This shows that a current carrying conductor develops around it a magnetic field whose direction depends on the direction of the flow of current. The existence of a magnetic field around a current carrying conductor can also be demonstrated by piercing a copper wire through a piece of cardboard and sprinkling some iron filings on the cardboard. If a current is passed through the wire from a battery and the cardboard tapped gently, the iron filings will arrange themselves in concentric circles around the wire as shown in Fig. 2.24.

If a small compass needle is taken around the wire and magnetic lines of force plotted, they will also form concentric circle around the wire like the iron filings. If the north pole of the compass needle is pointing in the clockwise

direction with the current flowing upwards as shown in Fig. 2.24, the compass needle will get deflected in the anti-clockwise direction when the direction of current is reversed.

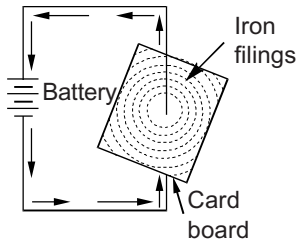


Fig. 2.24 Magnetic Field round a Current carrying Conductor

These experiments show that a conductor carrying current behaves like a temporary magnet having its magnetic field in concentric circles round the wire and in a plane at right angles to the current carrying conductor. The direction of this magnetic field depends upon the direction of the current through these conductors.

If the straight wire used in the above experiments is wound in the form of coil with a number of turns, the magnetic lines of force get concentrated inside the coil, the turns of the coil adding their own magnetic field to the total magnetic field which is quite strong inside the coil as shown in Fig. 2.25.

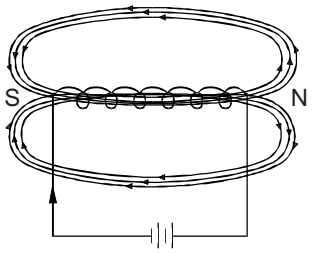


Fig. 2.25 Magnetic Field in a Coil

The coil with the current flowing in its turns behaves like a magnet with one end as north pole and the other end as south pole. Such a coil whose length is much greater than its diameter is called a *solenoid*. The direction of the magnetic field inside the coil is determined by the direction of the current flowing through the turns of the coil. The strength of the field inside the coil, however, depends on the amount of current I and the total number of turns N in the coil and is proportional to the product NI. The product NI is known as *ampere-turns*.

2.17 MAGNETIC POLARITY OF A COIL

A coil carrying current behaves like a bar magnet with the north pole at one end and south pole at the other end. Whether a particular end becomes a north pole or a south pole can be determined by a number of methods. Two of these rules are given below:

2.17.1 End Rule

Look at the end of the coil. If the electron flow (current) is clockwise, the end will have north polarity, if anti-clockwise, the end will have south polarity as in Fig. 2.26.

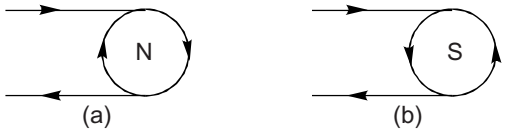


Fig. 2.26 End Polarity in a Current carrying Coil

2.17.2 Left Hand Rule

Grip the coil or the solenoid with the left hand, wrapping the fingers round the coil in the direction of the electron flow. The thumb will point in the direction of the north pole as in Fig. 2.27. Remember the electron flow is from the negative side of the voltage source through the coil and back to the positive terminal.

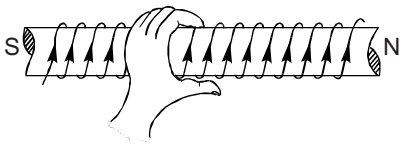


Fig. 2.27 Left Hand Rule for the Polarity of a Current carrying Coil

2.18 MAGNETIC CIRCUIT

The magnetic field developed inside a solenoid depends on the ampere-turns *NI*. The intensity of the magnetic field generally denoted by *H* will also be determined by the length of the solenoid. A longer coil will have less intensity at a point inside the coil or solenoid than a shorter solenoid having the same ampere-turns *NI*. It can be proved that if the length of the solenoid is *I*, then

$$H = \frac{4\pi}{10.l} \cdot NI = \frac{1.256}{l} \cdot NI \tag{2.7}$$

If the coil is wound on a material with permeability μ, the relation between the flux density and the field strength *H* is as follows

$$\mu = \frac{B}{H} \tag{2.8}$$

With a cross-section area *A* cm^2, the total flux through the coil will be

$$\text{Total flux } \phi = B \cdot A = \mu H \cdot A$$

Substituting the value of *H* from Eq. (2.7) above,

$$\phi = \mu \cdot \frac{1.256}{l} \cdot NI \cdot A$$

or
$$1.256 \cdot NI = \phi \cdot \frac{1}{\mu A} \tag{2.9}$$

Compare this equation with the Ohm's law equation.

$$E = IR$$

where 1.256 *NI* corresponds to the electromotive force (emf) *E* and is known as the magnetomotive force (mmf). The unit of mmf in CGS units is gilbert (Gb).

The quantity $l/\mu A$ which corresponds to the resistance *R* is Ohm's law is called *reluctance*. ϕ corresponds to current *I* of Ohm's law. Thus Ohm's law when applied to a magnetic circuit will be

$$\text{mmf} = \text{flux} \times \text{reluctance}$$

The formula for reluctance, viz, $I/\mu A$ shows that reluctance is directly proportional to length, inversely proportional to area of cross-section and inversely proportional to the permeability. Thus, for the same magnetomotive force or

for the ampere-turns *NI*, iron which has greater μ than air or vacuum, will have less reluctance and will produce much greater flux than air or vacuum.

2.18.1 Air Gap in a Magnetic Circuit

When it is desired to concentrate the magnetic lines within the magnetic material itself an annular ring of iron is used as a core material. The magnetic lines of force form closed rings within the core itself and no flux leaks out of the magnetic material which forms a close magnetic circuit. This type of coil is known as Toroid or Torroidal coil. Since iron has high μ, the reluctance of the closed circuit is very small and only few turns of coil are able to produce high flux inside the core material. If, however, a small air gap is allowed in the core as in Fig. 2.28 (b), the reluctance of the core will now increase and greater mmf will be needed to produce the same flux. In this case the magnetic lines of force will spread into air space near the air gap and any magnetic material near the air gap will have magnetism induced in it. This principle is used in magnetic tape recording where a strong magnetic field produced by audio currents in the air gap of the recording head induces magnetism into the magnetic material coated on the plastic tape which moves across the air gap of the recording head.

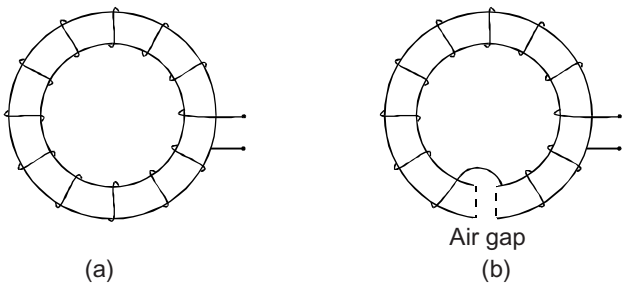


Fig. 2.28 Effect of Air Gap in a Magnetic Circuit

2.19 ELECTROMAGNETIC INDUCTION

It has already been shown how a current carrying conductor or coil has a magnetic field associated with it. This magnetic field is produced by the motion of electrons in the current carrying conductor. Similarly, if a conductor or a coil is moved in a magnetic field, the free electrons in the conductor are set into motion and a current flows into the conductor or the coil when the circuit is closed. Electricity and magnetism are two interlinked phenomena and one cannot be separated from the other. The production of electricity from magnetism and vice versa is termed *electromagnetic induction*, which has very wide practical applications in the operation of electric motors and generators.

Electromagnetic induction can be demonstrated by the following simple experiment.

Bring a bar magnet near a coil to which a sensitive microammeter (or a galvanometer) is connected as in Fig. 2.29(a). The needle of the microammeter shows deflection as the bar magnet is brought near the coil and the magnetic

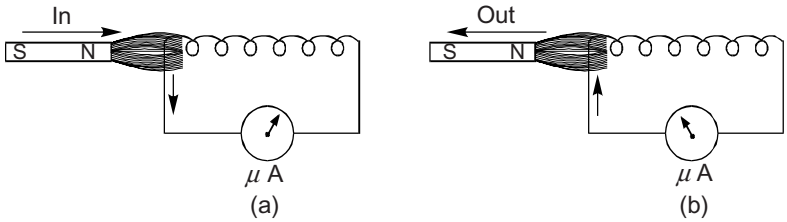


Fig. 2.29 How Electromagnetic Induction can be Demonstrated

lines of flux cut the coil. When the magnet is held stationary near the coil the meter shows no deflection and the current flow stops. Now move the magnet away from the coil as in Fig. 2.29 (b). The needle of the current meter again shows deflection but in the opposite direction. The current flow stops again if the magnet is held stationary. The effect produced is the same if the magnet is kept stationary and the coil is moved towards or away from the magnet.

Furthermore, if the magnet is moved in and out quickly, the current produced is stronger than when the magnet is moved slowly. Also, with a stronger magnet having greater flux density the current produced in the same coil is stronger than when a weak magnet is used.

All the facts stated above are governed by a set of laws known as Laws of Electromagnetic Induction. There are two important laws of Electromagnetic Induction:

1. Faraday's law
2. Lenz's law

2.19.1 Faraday's Law of Electromagnetic Induction

Faraday's law is actually composed of two laws:

Law I

An emf is induced in a closed circuit whenever the number of magnetic lines of force or the magnetic flux cutting the closed circuit changes.

The production of induced emf or current in a coil by moving a magnet near it has already been explained. The same effect can also be produced by placing two coils near each other as shown in Fig. 2.30(a). When the current flowing in the first coil called primary is changed, the magnetic flux linking the second coil called 'secondary' changes and induced emf is produced in it. The primary current can be changed either by a switching arrangement or by connecting some AC voltage across it. To enhance the effect of flux linkages the two coils are sometimes wound on a common magnetic material called the *core*. This, in fact, is the principle of transformers.

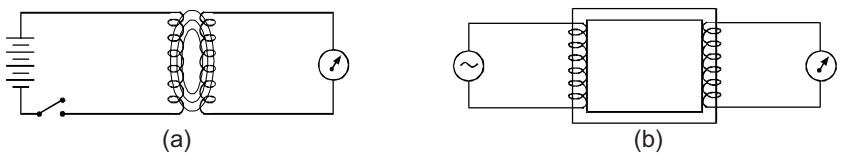


Fig. 2.30 Induced Voltages indicating Transformer Action

Law II

The magnitude of the induced emf depends on the rate of change of magnetic lines of force or magnetic flux cutting the closed circuit.

A higher emf is produced if *the rate of change*, i.e. the change of magnetic lines of force *per second* is higher. It can be proved that an emf of 1 V is produced if the number of lines of force cutting a coil changes by magnetic linkage of 10^8 lines per second. If the same change takes place in 1/10 of a second, the voltage induced will be 10 V. The rate of change of flux can also be increased by increasing the number of turns in the coil.

Summing up, we can say that the induced voltage will depend on the total flux cutting the coil, the number of turns in the coil and the rate of change of magnetic lines of force cutting the coil.

2.19.2 Lenz's Law

Lenz's law determines the polarity or direction of the induced emf or induced current. The law states that the direction of the induced emf or current is such that its own magnetic field will oppose the change that produced the induced emf or current. In Fig. 2.31 when the north pole of a magnet is brought near the coil, the end of the coil facing the north pole also becomes a north pole due to induced current and tries to repel the north pole of the magnet that is responsible for inducing the current in the coil. When the magnet is withdrawn from the coil, the induced polarity is a south pole which tries to attract or bring back the north pole that has induced this polarity. By application of the end rule or the left-hand rule already described, the direction of the induced current in the above two cases can be found.

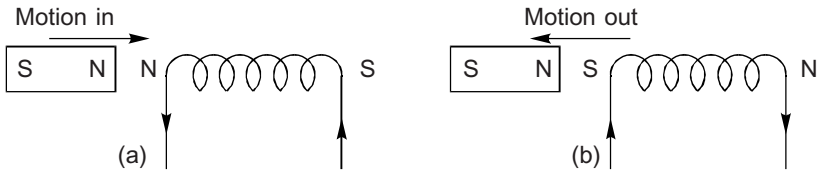


Fig. 2.31 Lenz's Law determines the Polarity of the Induced Voltage

It will be seen that in the two cases mentioned above there is opposition to the change that induces the emf and work has to be done to overcome this opposition. It is this work that changes into electrical energy. This is the Law of Conservation of Energy.

2.20 MAGNETISATION—HYSTERESIS

A ferromagnetic material like iron or steel can be magnetised by winding a coil round it and passing a current through this coil. As the magnetising force which is proportional to the ampere-turns is increased, the magnetism or flux produced in the magnetic material also increases but the magnetism or the flux always lags behind the magnetising force and this phenomenon of lagging

behind is known as *hysteresis*. This phenomenon arises from the fact that the small molecular magnets, of which the magnetic material is composed, resist any realignment produced by the magnetising force.

To overcome this resistance or internal friction of the molecular magnets, work has to be done by the magnetising force and energy is expanded which appears in the form of heat. This is called the *hysteresis loss*. When the magnetising force is removed, the molecular magnets do not come back to their original alignment but work has to be done again in demagnetising the material by reversing the magnetising force. Heat is produced again. Any magnetic material subjected to a cycle of reversing magnetising field results in the production of heat called hysteresis loss. Steel and other hard magnetic materials are subject to higher hysteresis loss than soft iron.

2.21 HYSTERESIS CURVE

The phenomenon of hysteresis can be studied by plotting a curve between the magnetising force *H* which is given by the ampere-turns and the flux density *B* which can be measured by special instruments. A *B*-*H* curve is shown in Fig. 2.32.

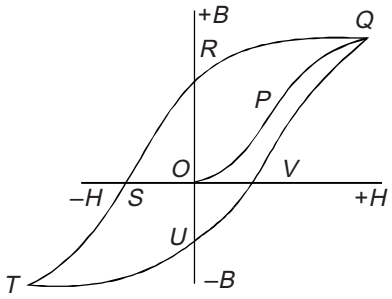


Fig. 2.32 Hysteresis Loop

As the magnetising force *H* is increased from *O* to the positive direction, the value of *B* increases till a saturation point is reached at *Q*. If the current or *H* is gradually reduced, the flux density *B* does not retrace the original curve but falls along the curve *QRST* as *H* is reduced through *O* to its maximum value on the negative side. It will be seen that when *H* is reduced to *O*, the magnetism *B* is not reduced to *O* but has a value equal to *OR*. This magnetism which is left in the magnetic material after the magnetising force *H* is reduced to *O* is called *residual magnetism* and the property of retaining magnetism by a material is called *retentivity*. Steel has greater retentivity than soft iron. Hence permanent magnets are made of steel rather than of soft iron.

Again to bring the value of flux to *O* as at *S*, a demagnetising force equal to *OS* in the reverse direction has to be applied. This is called the *coercive force*. Steel requires greater coercive force than soft iron and so does not lose its magnetism easily when subjected to external forces.

By increasing the value of H in the positive direction, the curve $TUVQ$ can also be traced. The magnetising force has completed a cycle of changes from zero to positive, back to zero and negative value and through zero again to positive value. The curve so completed is the curve $QRSTUVQ$ and this is called the *hysteresis loop*. It can be proved that the area of the hysteresis loop represents the energy loss or the hysteresis loss in the material and is a definite factor to be considered in all equipment where changing currents produce changing states of magnetisation.

2.22 TYPES OF MAGNETS AND THEIR USES

There are two types of commonly used magnets—*permanent magnets* and the *electromagnets*.

2.22.1 Permanent Magnets

Permanent magnets are made of hard magnetic materials like steel. The magnetic material is magnetised during its manufacture by subjecting the material to a strong inductive field. Cobalt steel, which is generally used for permanent magnets, is able to retain its magnetism for a long time after magnetisation because of its greater retentivity. Soft iron will only make a temporary magnet and will lose its magnetism easily. Certain alloys of steel particularly almico (alloy of aluminium, nickel, iron and cobalt with a percentage of copper and titanium) is used for making good grade commercial magnets.

Permanent magnets are normally made in the form of bar magnets, (Fig. 2.33(a)) or as horse-shoe magnets (Fig. 2.33 (b)). These can also be made in other shapes to suit special purposes.

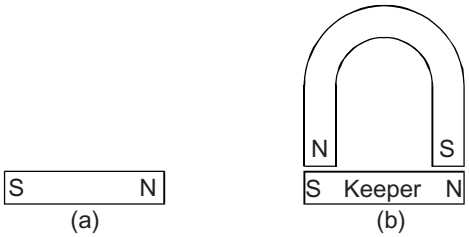


Fig. 2.33 Types of Magnets (a) Bar Magnet (b) Horseshoe Magnet with Keeper

When in storage a piece of soft iron called a keeper is put across the two poles as in Fig. 2.33 (b). A keeper forms a closed magnetic circuit with the magnet and helps retain the strength of the magnet over a long period by preventing any external magnetic fields from inducing opposite polarities in the permanent magnet and thereby weakening its strength.

Permanent magnets find extensive use in permanent magnet loudspeakers, dynamic microphones, pickups, head phones and measuring instruments like multimeters.

2.22.2 Electromagnets

An *electromagnet* is made by winding a coil or a solenoid round a magnetic material and passing current through the coil from some external source or battery as in Fig. 2.34. As soon as the switch is made, the iron coil becomes a magnet and can attract other pieces of iron like nails. The electromagnet loses its magnetism when the switch is opened and current stops flowing round the coil. Soft iron forms a good material for the core of the electromagnet because it gets easily magnetised and demagnetised. The strength of the magnet depends on the amper-turns (*NI*) of the coil and the permeability of the core material. Very powerful electromagnets which can lift tons of iron can be made in this way.

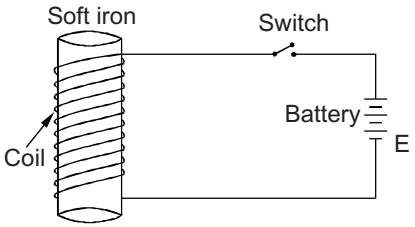


Fig. 2.34 Principle of an Electromagnet

Electromagnets are used in lifting magnets and in the separation of iron ore and iron scrap from other non-magnetic materials, manufacture of relays, electric bells and buzzers. The record and playback heads in tape recorders are nothing but electromagnets.

2.22.3 Relay

It is an electromagnetic device which can be used to switch on or switch off one or more circuits by a remote control operation. It consists of a coil of wire wound on a soft iron core as in Fig. 2.35. When a current is made to flow through the coil from an external battery, the soft iron becomes a magnet and attracts the iron piece called the *armature*. An armature is a piece of iron that moves under the influence of a magnetic field. The iron is balanced on a pivot so that when it is attracted by the electromagnet, the other end of it pushes together or pushes apart some switch contacts thereby closing or opening some other circuits. The number and types of these secondary contacts varies considerably. As many as eight make and break contacts can be operated by one relay. Relays find extensive use in broadcasting and television studios and also in telephone exchanges.

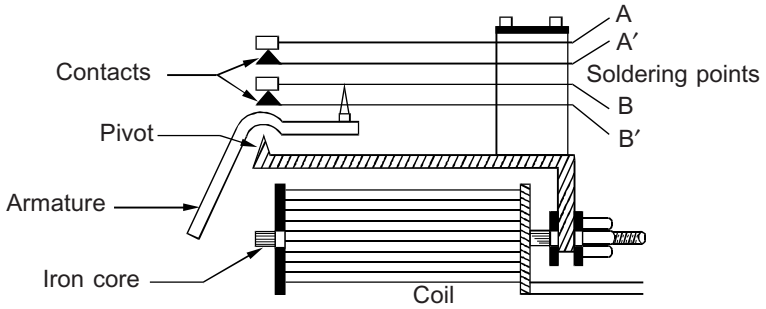


Fig. 2.35 Electromagnetic Relay

2.22.4 Electric Bell or Buzzer

An electric bell or a buzzer is also an electromagnetic device. An electromagnet is used to produce the rapid back-and-forth motion of an armature which has a hammer or clapper at the end that keeps striking a gong to produce the continuous ringing sound (Fig. 2.36).

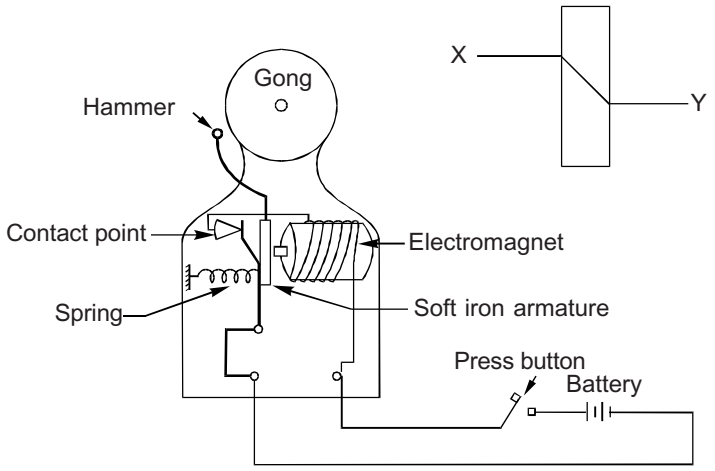


Fig. 2.36 An Electric Door Bell

When a door bell is pressed, the circuit is completed and the electromagnet gets energised due to current flowing through its coil. The electromagnet attracts the soft iron armature and the hammer attached strikes the gong thereby producing a sound. When the armature is pulled away by the electromagnet the contact strip is also pulled away from the contact point and this breaks the circuit. The electromagnet gets de-energised. The spring pulls the armature back and contact is made again. This action is repeated very rapidly so long as the press button is kept pressed and a continuous ringing sound is produced by the hammer repeatedly striking the gong.

In the case of a buzzer, however, there is no gong or hammer and the buzzing sound is produced by the armature striking repeatedly against one end of the electromagnet core.

Summary

Electricity that is produced by rubbing substances with each other is called static electricity. A body can be charged by friction either with positive electricity or with negative electricity. A positive charge on a body results from the deficiency of electrons which are tiny particles charged with the unit of negative electricity. Protons are particles charged with a unit of positive electricity. A negatively charged body contains excess of electrons or a deficiency of protons. A neutral body contains an equal number of electrons and protons.

Atoms are the building blocks for all matter which may exist in solid, liquid or gaseous form. An atom consists of a nucleus at the centre which contains protons and neutrons with electrons revolving round the nucleus in different orbits. The number of protons in the nucleus is equal to the number of electrons in the orbit. Each element has a fixed number of electrons in the orbit round the nucleus. The physical and chemical properties are determined by the atomic structure of the element. When the electrons in the outermost orbit of an element are free to move about, the element can conduct electricity easily, and is called a conductor. When electrons in the outermost orbit are not free to move about the element is known as an insulator. A semiconductor is neither a good conductor nor a good insulator.

Current is the rate of flow of charge in a conductor. The unit of current is ampere. Voltage is the potential difference (pd) and electromotive force (emf), the electrical pressure that drives a current through a conductor. The unit of pd is volt. The opposition that exists in the flow of current is called resistance and is denoted by the symbol R. The three quantities current. voltage and resistance are connected by the formula $I = E/R$ which is Ohm's law. The other forms of Ohm's law are $E = IR$ and $R = E/I$. Power is the rate at which work is done and is given by the formula $P = E \times I$. The unit of power is watt. Kilowatt hour is the unit for electrical energy.

Magnetism A magnet is a piece of iron or steel which attracts small pieces of iron. When suspended freely, one end of magnet which always points to the north is called the north pole and the other end the south pole. In the case of two magnets, like poles repel and unlike poles attract each other. The space around a magnet in which its influence can be felt is called a magnetic field which is represented by magnetic lines of force starting from the north pole and ending at the south pole. The total number of lines of force leaving the north pole is called flux and flux density is the number of lines of force in a unit area perpendicular to the direction of the lines of force. A piece of iron placed in a strong magnetic field gets magnetised due to induction. The magnetic lines of force get concentrated in the soft iron due to permeability.

A conductor or a coil carrying current behaves like a magnet, the polarity depending on the direction of current flow. Similarly, a magnet moved near a conductor or a coil induces an emf in the coil. The magnetism produced by the current and vice versa is called electromagnetic induction and forms the basis for the operation of electric motors and generators. Electromagnetic induction is governed by Faraday's Laws and Lenz's Law.

There are two main types of magnets—permanent magnets and electromagnets. Permanent magnets are extensively used in permanent magnet speakers, microphones, pickups and multimeters, etc. Electromagnets find extensive use as lift magnets in relays and electric bulbs.

Review Questions

1. What is static electricity? What type of charge is acquired by a glass rod when rubbed with silk?
2. How will you show experimentally that similar charges repel and opposite charges attract each other?
3. What is an electron? How does it differ from a proton?
4. Define element and compound. Separate the elements from the compounds in the following: copper, water, common salt, oxygen, aluminium, neon, hydrochloric acid.
5. What is an atom? What are the particles present in an atom? Where do these particles normally reside in an atom?
6. Describe the atomic structure of an element with atomic numbers 8 and 12.
7. Define conductors, insulators and semiconductors. Give two examples of each.
8. Separate the following into insulators, conductors and semiconductors: carbon, copper, silver, iron, germanium, glass, rubber, silicon, mercury, mica, dry air.
9. Define current, voltage and resistance. What is the practical unit for each of these?
10. State Ohm's law in all its three forms. What is the resistance of the element in an electric press which draws a current of 2.5 A at 220 V. (*Ans:* 88 Ω)
11. What will the electric consumption bill be for the month of January in a household using the following electrical appliances: (i) 1 kW electric room heater for 4 hr a day, (ii) $\frac{1}{2}$ kW electric toaster for 2 hr daily; (iii) 2-100 W lamps and 3-60 W lamps for 5 hr a day. Electricity is charged at 20 paise per unit. (*Ans:* Rs. 42.70)
12. What is power? What is the relation between horsepower and watt. Show that 1kW is equal to 1.34 hp.
13. What is the resistance in each of the following cases:
    (i) 1200 W dissipated with 230 V.
    (ii) 10 μA with 12V
    (iii) 5 hp across 120V
    (*Ans*: (i) 40.33Ω, (ii) 1.2 MΩ (iii) 3.86Ω)
14. What is a magnet? Describe the two properties of a magnet.
15. Given a magnet and a magnetic needle, how would you proceed to determine the polarity of the magnet?
16. What is a magnetic field? How can the magnetic field of a bar magnet be plotted experimentally.
17. What is flux density? A flux density B of 5000 G/cm^2 is produced in a cross-sectional area of 4 cm^2. What is the total flux ϕ produced by the magnet? (*Ans.* 20,000 G)

18. The distance between two magnetic poles is halved. Will the force of attraction or repulsion be (i) halved, (ii) doubled, (iii) reduced to one fourth, or (iv) increased four times.
19. What is magnetic induction? How can this phenomenon be explained on the basis of the molecular theory of magnetism?
20. Define electromagnetic induction. A current of 200 mA flows through a coil which has 2000 turns and is 0.2 m long. Find the field intensity *H* in Gb/cm. (*Ans*: 25.12 Gb/cm)
21. State Faraday's Laws of Electromagnetic Induction. Give two examples of the practical application of these laws.
22. Separate into (i) ferromagnetics, (ii) paramagnetics, and (ii) diamagnetics the following metals: nickel, steel, aluminium, bismuth, alnico, chromium, antimony,
23. State whether True or False with brief reasons.
    1. A neutron is much heavier than a proton.
    2. Mica is a much better conductor than copper.
    3. Wattage of two resistors increases when connected in series.
    4. Steel has greater retentivity then soft iron.
    5. Hysteresis loss is the production of heat due to a magnetising cycle, (*Ans*: 1 False 2. False 3. True 4. True 5. True)

chapter 3

DC and AC Voltages and Currents

3.1 DIRECT CURRENT AND ALTERNATING CURRENT

We know how current flows through a circuit consisting of a voltage source, conductor and load. If the voltage source is a battery with positive and negative poles, the electron current flows from the negative pole, through the conductors and the load back to the positive pole of the battery as shown in Fig. 3.1.

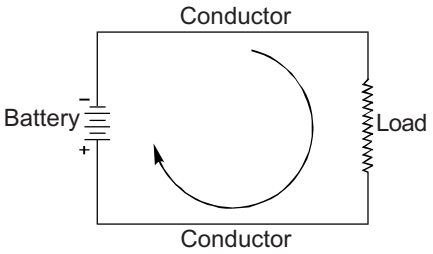


Fig. 3.1 Direct Current (DC)

So long as a steady potential difference exists between the terminals of a battery, a steady current will flow through the circuit in the same direction. A current that flows always in the same direction without changing or reversing the direction of the flow is called a *direct current* or DC. A direct current needs a DC source of voltage like the chemical voltage source, or a battery. Direct current is the earliest form of the current and laws like Ohm's law mainly pertain to direct current and voltages.

If, however, the voltage source is an alternator in which the polarity at the two terminals of the source changes from positive to negative and vice versa at regular intervals and will alternately flow in one direction and then in the reverse direction depending on the polarity of the terminals of the voltage source as shown in Fig. 3.2. This

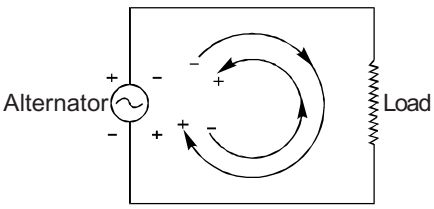


Fig. 3.2 Alternating Current (AC)

type of current which reverses direction at regular intervals is called *alternating current* or AC. An alternating current requires an AC voltage source or an alternator.

It might appear that because of the property of reversing direction continuously, AC may not be of much practical use because the useful effects produced by it while flowing in one direction might be undone when it flows in the opposite direction. This is, however, not the case. The effects produced by current like heat and light do not depend on the direction of the flow of electrons. In fact, this reversal of direction by AC is its biggest advantage. It is this reversal of direction that makes transformer action possible and lends to AC much more flexibility than possessed by DC. The advantages of AC so much out-weigh the advantages of DC that AC is rapidly replacing DC in most practical applications so that more than 90 per cent of the electric power consumption in the world today is from AC voltage sources.

3.2 VOLTAGE SOURCES

A voltage source is a device that produces electricity by converting some other form of energy into electrical energy. By its action the voltage source merely creates a potential difference (pd) or electromotive force (emf) between two terminals by building up opposite charges on these terminals. When an electrical appliance like an electric bulb or a heater is connected across the two terminals of the voltage source, a current flows in the circuit and electrical energy is consumed to produce different types of effects. The forms of energy that are generally converted into electrical energy are chemical energy, heat energy, light energy and magnetic energy. Various types of voltage sources are, therefore, based on one or the other type of energy conversion mentioned above. Some of the common types of voltage sources or power sources are described in the following paragraphs.

3.3 DC VOLTAGE SOURCES

3.3.1 Chemical Voltage Sources—Batteries

A chemical voltage source is one of the most important sources of electrical energy. It is a self-contained voltage source and does not need any outside energy like heat, light or mechanical energy to produce electricity. All the electrical energy supplied by a chemical source of voltage is produced by chemical action within the source itself. Chemical voltage sources normally exist in the form of batteries and cells of various types. These batteries or cells are extensively used as portable sources of power in portable radio and TV sets, flash lights, photoflash lamps, hearing aids, electronic watches and clocks and in electronic measuring instruments. Batteries are also used as sources of electrical power in automobiles, trams, aeroplanes and ships. In fact, a battery is the most versatile voltage source in use today.

3.3.2 Types of Batteries

A battery is a combination or group of cells connected to each other to supply a voltage or current higher than that supplied by a single cell. A cell is the basic unit of a battery. However, the terms 'battery' and 'cell' are more or less used interchangeably. A cell or a battery is classified as a 'primary' or 'secondary' depending upon the manner in which the chemical energy is converted into electrical energy. In a primary cell Fig. 3.3(a) the chemical energy is converted into electrical energy directly with the help of the chemical materials contained in the cell. A secondary cell must be charged with electrical energy first, to enable it to convert the chemical energy into electrical energy. Because of its action of storing energy supplied to it, a battery consisting of secondary cells is often called a *storage battery* (Fig. 3.3(b)).

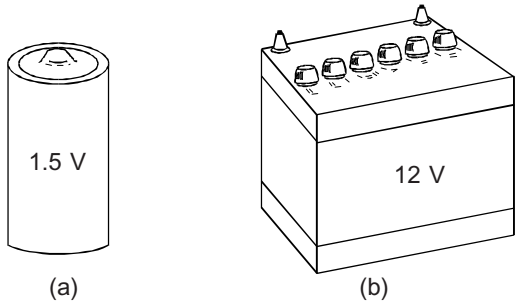


Fig. 3.3 Primary and Secondary Cells (a) Dry Cell (b) Storage Battery

Cells are further classified as wet cells or dry cells. A wet cell uses liquid chemicals whereas a dry cell uses these chemicals in the form of a paste. It is not dry in the real sense of the word. Primary batteries generally consist of dry cells and are used in transistor radios and other electronic devices where a limited current is required. The secondary batteries or storage batteries generally use secondary cells and are employed where the current consumption is heavy as in automobiles and aeroplanes.

3.3.3 Primary Cells-Voltaic Cells

When two different metals are immersed in a solution, the chemical action between the metals and the solution results in the production of electricity. Such an arrangement forms the basic primary wet cell and is called a *voltaic cell* after its discoverer, Alessandro Volta. The metals in the cell are called *electrodes* and the chemical solution is called the *electrolyte* (Fig. 3.4(a)). The electrolyte reacts oppositely to the two different electrodes. It causes one electrode to lose electrons and build up a positive charge and the other electrode to gain an excess of electrons and develop a negative charge. The cell voltage is the difference of potential between the two oppositely charged electrodes.

One form of basic primary wet cell is a cell in which zinc (Zn) and copper (Cu) are used as the two electrodes and a solution of sulphuric acid (H_2SO_4) in water, serves as the electrolyte as shown in Fig. 3.4(b). In this cell the electro-

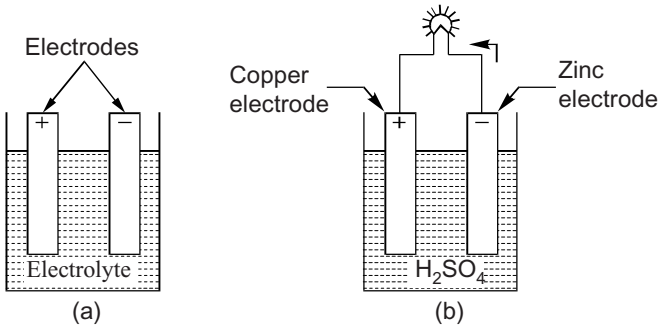


Fig. 3.4 Primary Cell (a) Simple Voltaic Cell (b) Construction of a Voltaic Cell and Direction of Flow of Current

lyte breaks up into hydrogen (H^+) and sulphate (SO_4^-) ions. The sulphate ions which are negatively charged attack zinc and form zinc sulphate ($ZnSO_4$) and release electrons from zinc atoms. The zinc plate develops an excess of electrons and becomes negatively charged. The positively charged hydrogen ions (H^+) move towards the copper plate and attract a few electrons from the copper plate thereby creating a deficiency of electrons which results in a positive charge on the copper plate. Thus, a difference of potential is created between the positive copper plate and the negative zinc plate which is the potential difference or the emf of the cell. The value of this emf depends on the materials used for the electrodes and the electrolyte. In the case of the zinc-copper-sulphuric acid cell the emf developed is about 1.08 V.

When a load like a small flashlight bulb is connected across the two terminals of the cell as in Fig. 34(b), the electron current flows from the zinc terminal to the copper terminal through the lamp load and the lamp lights up. On reaching the copper plate, the electrons get neutralised by the positive hydrogen ions but more and more electrons are released at the zinc plate due to chemical action. This cycle will continue so long as current is flowing through the lamp and will stop only when the current is switched off or the zinc plate gets completely dissolved in the solution.

A voltage cell suffers from two basic defects—(i) local action and (ii) polarization.

(i) Local Action Due to the presence of iron and carbon as impurities in the zinc plate, small local voltaic cells are formed in the zinc plate itself and local current flows even when no current is being supplied by the main cell. This affects both the output and the life of the cell. Local action is prevented by coating the zinc plate with mercury—a process known as *amalgamation.*

(ii) Polarization Neutral hydrogen gas produced at the copper electrode forms a non-conducting layer round this electrode which interferes with the normal working of the cell. This phenomenon is called polarization. This difficulty is generally overcome by adding a depolariser like manganese dioxide to the electrolyte which reacts with hydrogen to form water.

A convenient form of voltaic cell is the *Leclanche cell*, in which the electrolyte is ammonium chloride, zinc is the negative electrode and carbon the positive electrode. The so-called dry cell is actually a Leclanche cell.

In this cell the electrolyte is not in a liquid form but is a paste of ammonium chloride or sal ammoniac with powdered manganese dioxide and granular carbon. This is filled in zinc container which serves as the negative electrode. A carbon rod is embedded in the centre of the cell and is the positive electrode. Complete construction details of the dry cell are shown in Fig. 3.5.

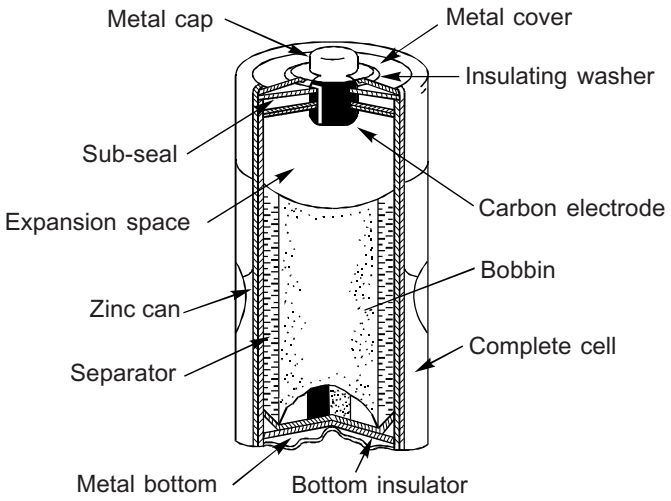


Fig. 3.5 Construction of a Dry Cell

A dry cell develops a voltage of 1.5 V. However, current supplied is dependent on the size of the cell. A cell which is big will deliver more current than a cell of smaller size but both will have the same voltage of 1.5V.

The electrolyte often leaks through the paper insulating covers and damages the cell container. To avoid this leakage, leakproof cells are being manufactured. These leakproof cells are encased in a steel jacket with a paper tube between the steel jacket and zinc container for insulating purposes. The local action that takes place even when no current is being drawn from a dry cell, imparts it a limited shelf-life. Cells which have been stored for a long time should be carefully checked for voltage and current before they are put to actual use.

3.3.4 Combination of Cells (Battery)

Although the words, 'battery' and 'cells' are often used interchangeably, technically speaking a battery is a combination of two or more unit cells. A combination of cells to form batteries is used when the voltage, or current, or both, supplied by a single cell is not enough. When a higher voltage than available from a single cell is needed, cells are connected in series. To increase the current

rating, cells are connected in parallel. When both voltage and current rating are required to be increased, then a series-parallel combination of cells is used.

3.3.4.1 Series Combination

In the series combination of cells (Fig. 3.6(a)) the positive terminal of one cell is connected to the negative terminal of the other cell, and so on. In this case the voltages add up and the series combination is also called the series-aiding combination. Thus, three 1.5 V cells will have to be connected in series for a 4.5 V dry battery (Figs 3.6 a and b). Since the same current flows through all the cells in the series combination, the current rating of the combination does not increase. In fact, the current rating of the combination will be the current rating of the weakest cell.

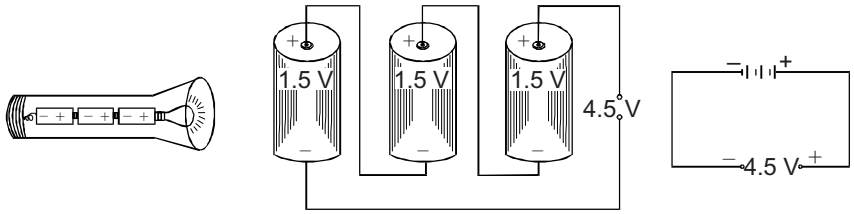


Fig. 3.6 Series Combination of Cells (a) 3 Cell Torch—Voltage $3 \times 1.5 = 4.5$ V (b) Actual Arrangement (c) Symbolic Representation

3.3.4.2 Parallel Combination

In a parallel combination of cells all the positive terminals are connected to each other and all the negative terminals are also connected, as shown in Fig. 3.7. The parallel combination is equivalent to a single cell with increased size or electrodes and the electrolyte. Consequently, the current rating increases but the voltage rating remains the same as for one cell. All the cells connected in parallel should have the same voltage otherwise the cells with higher voltage will supply current to the cells with lower voltage.

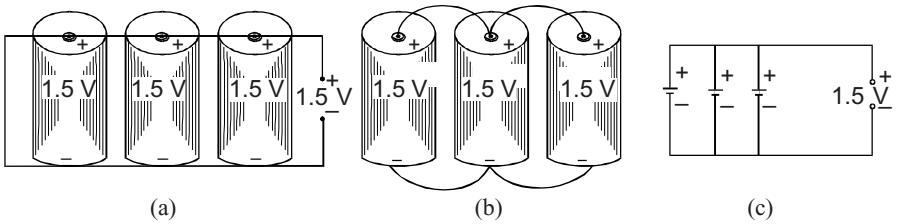


Fig. 3.7 Parallel Combination of Cells

3.3.4.3 Series-Parallel Combination

A series-parallel combination is shown in Fig. 3.8. In this arrangement the voltage as well as the current rating of the combination is three times the voltage and the current rating of each individual cell.

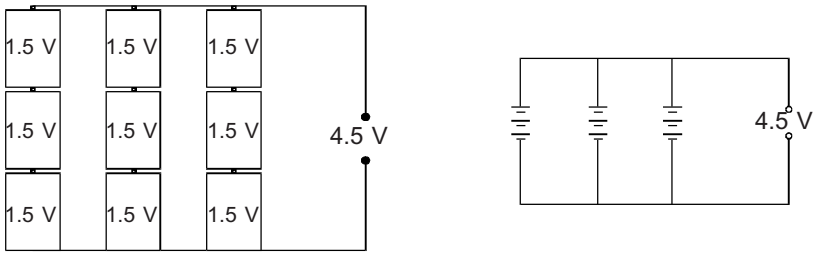


Fig. 3.8 Series-parallel Combination of Cells

For this arrangement the voltage and current rating of individual cells should be the same.

3.3.5 Secondary Cells—Storage Batteries

A secondary cell is different from the primary cell in that its chemical action is reversible, whereas in a primary cell it is not. In other words, a primary cell converts the chemical energy built into it into electrical energy and in the process destroys itself. It cannot be recharged or reactivated whereas a secondary cell just receives electrical energy from an external source and stores it in the form of chemical energy during the process called charging. This stored chemical energy is then released by the secondary cell as electrical energy during the process of discharging, when the secondary supplies current to a load. This cycle of charging and discharging can be repeated and the secondary cell gives a much longer life than a primary cell.

Because of its property of storing electrical energy in the form of potential chemical energy, the secondary cell is also called a *storage cell*. A battery made from a combination of secondary cells is known as a *storage battery.* Storage batteries or secondary batteries are used where high values of load current are required as in the case of automobile batteries. The starting load current for an automobile can be as high as 200 to 300 A. A secondary cell has an output voltage of 2 to 2.2 V. As such, a 6 V automobile battery uses three secondary cells in series combination and a 12 V automobile battery will use six secondary cells in series. The rating of the secondary battery will depend on the physical size of each secondary cell forming the battery.

3.3.5.1 Lead-acid Cell

One of the most popular forms of secondary cells used for storage batteries is the *lead-acid cell*. It is formed by two electrodes both of lead sulphate ($PbSO_4$) immersed in an electrolyte which is pure distilled water as in Fig. 3.9. This arrangement will not behave like a cell till the two electrodes are made dissimilar and the electrolyte changed into one that will react with the electrodes. This is done by a process called *charging*.

Charging For charging a lead-acid cell, a current is passed through the cell by connecting the two electrodes to the positive and negative terminals of an external source of dc voltage (Fig. 3.10) called a *battery charger*. A few drops

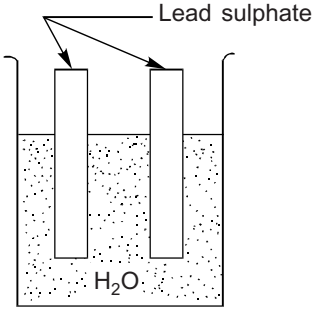


Fig. 3.9 Lead-acid Cell (uncharged)

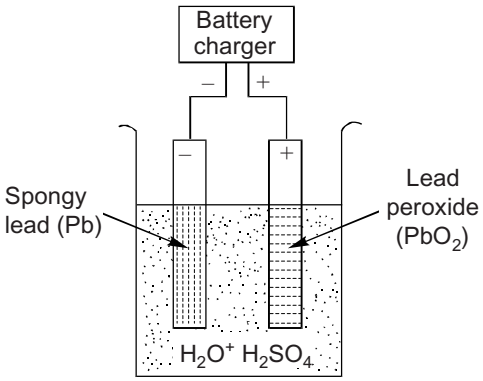


Fig. 3.10 Changing of Lead-acid Cell

of sulphuric acid added to distilled water facilitates electrolysis and produces positive H^+ions and negative O^{2-} ions from water (H_2O). At the electrode which is connected to the negative terminal of the battery charger, the positive hydrogen ions (H^+) from water attack the negative SO_4^- ions of the $PbSO_4$ plate to form H_2SO_4. At the same time the positive lead (Pb_2^+) ions are neutralised by the electrons from the battery charger and produce spongy lead (Pb). When the process of charging is completed, the negative electrodes change from $PbSO_4$ to Pb and the concentration of sulphuric acid in the electrolyte increases considerably.

At the positive electrode which is the electrode connected to the positive terminal of the battery charger, the positive hydrogen ions (H^+) again combine with negative SO_4^- ions of $PbSO_4$ to form (H_2SO_4). The electrons from the positive electrode have to leave the cell for flow of current and this deficiency of electrons is made up by the negative oxygen (O^{2-}) ions of water which combine with the positive lead ions (PB^{2+}) of $PbSO_4$ to form neutral lead peroxide PbO_2. When the process of charging is completed the positive electrode changes from $PbSO_4$ to lead peroxide PBO_2 and the concentration of sulphuric acid in the electrolyte increases considerably. In other words in a fully charged cell the negative electrode consists of spongy lead (Pb) and the positive plate is lead peroxide (PBO_2) and the electrolyte consists of a highly concentrated solution of sulphuric acid H_2SO_4.

A fully charged lead-acid cell has an emf of about 2.2 V.

Discharging When a fully charged lead-acid cell is connected to a lamp load as in Fig. 3.11, the electrical energy which was stored in the cell as chemical energy during charging is released as electrical current flowing through the lamp load which glows. The chemical processes in the cell are now reversed when the electrons leave the negative electrode and pass through the load to enter the cell again at the positive electrode. At the negative electrode the spongy lead again reacts with SO_4^- ions from the electrolyte and changes into $PbSO_4$ which is deposited on the negative electrode. The positive hydro-

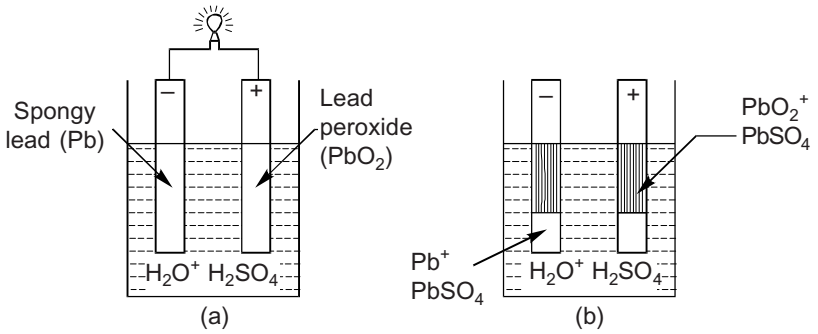


Fig. 3.11 Chemical Processes involved in Charging and Discharging a Lead-acid Cell (a) Charged Cell (b) Discharged Cell

gen (H^+) ions combine with oxygen ions to produce water which reduces the concentration of sulphuric acid in the electrolyte. At the positive electrode also the lead ions from PbO_2 react with SO_4^- ions to produce $PbSO_4$ and H^+ ions combine with oxygen ions to produce water. As more and more current is drawn from the cell, each electrode gets a deposit of $PbSO_4$ and the concentration of sulphuric acid in the electrolyte decreases. A stage is reached when the cell cannot generate sufficient emf to supply the current to the load. The cell is then discharged. Its emf falls to 1.75 V. However, the discharged cell can again be charged with the help of a battery charger as already described and this cycle of charging and discharging and recharging can be continued over long periods of time till the electrodes crumble or become too thin to store any further charge of electricity.

This reversible process in a lead-acid cell can be represented by the following reversible chemical equation:

$$Pb + PbO_2 + 2H_2SO_4 \quad \rightleftharpoons \quad 2\ PbSO_4 + 2H_2O$$

3.3.5.2 Construction of a Lead-acid Cell

The electrodes used in the actual construction of lead-acid cells are "formed" outside by a special electrolytic process. Each electrode consists of a number of elements called "plates" made of lead-antimony alloy on which the active material which is lead oxide is pasted. A forming charge is given to these plates coated with lead oxide to produce the positive and the negative poles. In the forming process the active material changes to lead peroxide which forms the positive (+) plate and spongy lead forms the negative (–) plate. In order to give the electrodes a larger surface area for higher current rating, each electrode is made of a group of plates. The plates of the negative electrode and those of positive electrode are interleaved and separated by thin sheets of porous non-conducting material called *separators* to prevent shorting of the positive and negative plates (Fig. 3.12). The group of plates for each electrode is connected to a *lead strap* that is attached to the associated lead terminal. The entire assembly is enclosed in an acid resistant container which has a top cover with a vent hole. The top is sealed with a sealing compound and the sulphuric

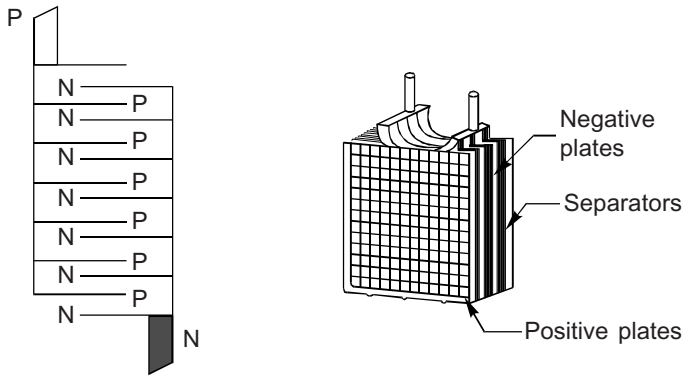


Fig. 3.12 Construction of a Lead-acid Cell

acid solution of known specific gravity is poured into the cell through the filler plug up to a level which keeps the plate assembly just submerged in the acid solution. The cell then can be charged with a battery charger at the prescribed rate.

3.3.5.3 Specific Gravity

The state of charge or discharge in a lead-acid cell is indicated by the concentration of the sulphuric acid (H_2SO_4) in the electrolyte. The measure of concentration of sulphuric acid is its specific gravity which is defined as the ratio of the weight of one cubic centimetre of the solution to the weight of one cubic centimetre of water.

$$\text{Specific gravity} = \frac{\text{Weight of 1 cc of sulphuric acid solution}}{\text{Weight of 1 cc of water}}$$

Pure concentrated sulphuric acid has a specific gravity of 1.835 which means that it is 1.835 times as heavy as an equal volume of water. When mixed with water this specific gravity of the solution of sulphuric acid decreases. The electrolyte of a fully charged lead acid cell has a specific gravity varying between 1.200 to 1.250 depending upon the manufacture of the cell. A half charged cell generally has a specific gravity of 1.150 which falls down to as low a value as 1.110 in the case of completely discharged cell.

3.3.5.4 Hydrometer

The device or instrument used for measuring the specific gravity of a solution is called a hydrometer. A battery hydrometer as shown in Fig. 3.13 consists of a glass syringe containing a calibrated float which dips lower or

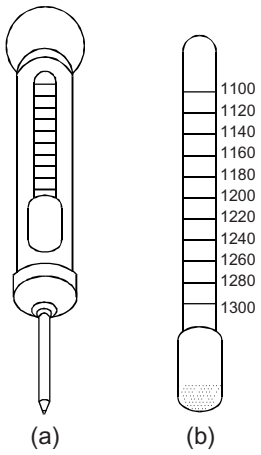
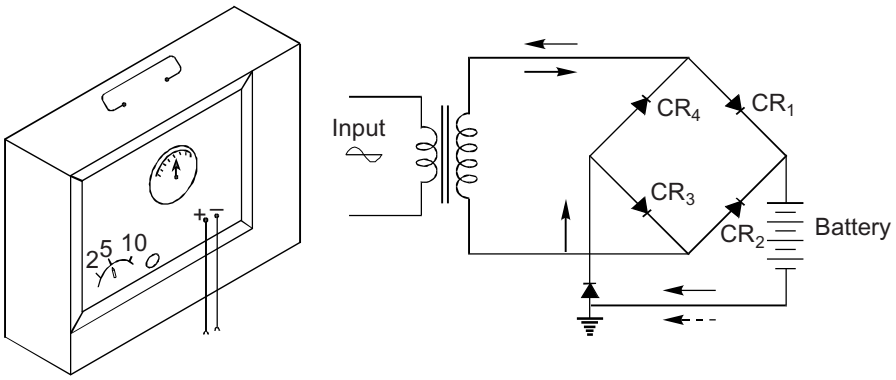


Fig. 3.13 Battery Hydrometer
(a) Hydrometer
(b) Float

higher into the electrolyte sucked into the glass tube of the syringe from the lead-acid cell. The reading on the calibrated float against the level of the electrolyte indicates the specific gravity of the electrolyte. In reading the specific gravity the decimal points are generally omitted and the specific gravity of a fully charged cell which is 1.250 will be simply read as "twelve fifty" and that of a half-charged cell which is 1.150 s "eleven fifty", etc.

3.3.5.5 Battery Charger

The source of DC voltage for charging a storage battery is provided by a battery charger. It consists of a rectifier unit which converts a 220 V AC supply into a DC supply voltage suitable for charging a storage battery which is also sometimes called an *accumulator*. The DC voltage required is slightly higher than the voltage of the battery to be charged. The rate of charging can be controlled and is indicated by an ammeter fitted into the battery charger. One such battery charger is shown in Figs. 3.14(a), and 3.14(b) shows the circuit diagram of a battery charger.



(a) Battery charger (b) Circuit diagram of a battery charger

Fig. 3.14

Trickle Charger A storage battery must be kept fully charged and should not be allowed to remain in a discharged or partly charged condition for long to avoid damage to the battery. For this purpose a small battery charger called a *trickle charger* is connected across the battery to provide a small charging current of a few milliamperes to the battery whether it is in use or not. This keeps the battery fully charged. A trickle charger is particularly useful for batteries which have to be kept out of use temporarily or for standby batteries which are used only in emergencies. A trickle charger, however, cannot be used for charging a new battery or for charging completely discharged batteries where high rates of charging are required.

3.3.6 Current Rating—Capacity—Efficiency—Precautions

A battery is rated in terms of the discharge current it can supply continuously for a specified interval of time. The output voltage should remain constant at a

minimum level of 1.5 to 1.8 V per cell over the entire period of time. A common rating is based on an 8 hour continuous discharge and is measured in terms of *ampere-hours* (A h). Thus a 100 A h battery can supply a current of 100/8 = 12.5 A based on an 8 h discharge. The same battery can supply more current for a shorter time and less current for a longer time. The current rating depends on the size and surface area of the electrodes. Current ratings vary in the case of a lead-acid battery from 100 to 300 A h. A battery cell with a high ampere-hour capacity is called a *heavy duty battery.* Efficiency is the Ratio of capacity output to capacity input.

The "amphere-hour" efficiency is from 80-90 per cent and the "watt-hour" efficiency is from 60-75 per cent.

Thus if a cell is charged at 10 amps for 16 hours, the input is 160 amps hours.

The output would be about $160 \times \frac{80}{100} = 128$ amp. hours at 80% efficiency.

This would give a current of 9 amps for about 14 hours at an efficiency of $\frac{128}{160} \times 100 = 80\%$.

Precautions:

1. Give battery proper initial charge.
2. Give a new battery plenty of work and liberal charging.
3. Do not charge too much or too little or at very high or low rate.
4. Do not run batteries for low in voltage or specific gravity.
5. Do not let batteries stand too long completely discharged.
6. Charge once a week, if possible.
7. Keep plates covered with electrolyte, making up evaporation losses with distilled water.
8. Test acid strength periodically.
9. Keep terminals and top of cell clean and dry and terminals coated with vaseline.

3.3.7 Alkaline Cell—Nickle/(NiFe) Cell

This is also known as Edison cell. The electrolyte is a potassium hydroxide (KOH) solution, the positive plate being formed of an oxide or hydroxide of nickel held in a nickel-steel frame, negative plate consisting of pure iron (Fe), in a container of welded steel.

No acid enters into working and the chemical action has been described as

$$Ni\ (OH)_4 + Fe \rightleftharpoons Ni\ (OH)_2 + Fe\ (OH)_2$$

During discharge the iron is oxidised to ferrous hydroxide $Fe\ (OH)_2$; during charging the changes are reversed.

The Edison cell may be made of moderate size and have found considerable use for traction work in small vehicle and aircraft, since they are much lighter than the ordinary heavy lead acid accumulates.

The electrolyte is 21 per cent solution of potassium hydroxide of specific gravity 1.21, to which a little lithium hydroxide is added. The e.m.f. of such a cell is 1.33 to 1.35 volts and is slightly dependent on the strength of potassium hydroxide. The efficiency is low.

The advantage of iron cell is its indifference to violent **, to overcharging and discharging ** rate.

3.4. ELECTROMAGNETIC VOLTAGE SOURCES—GENERATORS

One of the most important sources of electrical power are the electromagnetic sources called generators. No other source of electrical power can produce the large amounts of electrical power that the generators can. Generators are rotating machines which convert mechanical energy into electrical energy with the help of magnetism. A conductor or a set of conductors is rotated in a magnetic field and voltages are developed in the rotating conductor due to magnetic induction. The energy for the rotation of the conductors can be supplied by a petrol or diesel engine, by a steam turbine, waterfalls or even by an atomic reactor. Accordingly, we have diesel generators, hydro generators, thermal generators and atomic generators.

3.4.1 Basic Principles of a Generator

If a conductor *AB* is moved in a magnetic field as shown in Fig. 3.15(a), the free electrons in the conductor are set into motion by the magnetic field and a current is induced in the conductor in accordance with Faraday's laws of electromagnetic induction. An excess of electrons is produced at one end of the wire which becomes the negative pole and a deficit of electrons is produced at the other end which becomes a positive pole. A meter connected across the two poles *A* and *B* will show a deflection. The induced current flows in one direction when the conductor moves up, and in the opposite direction when the conductor moves down. The direction of the current flow is given by the Left Hand Rule for generators as shown in Fig. 3.15(b).

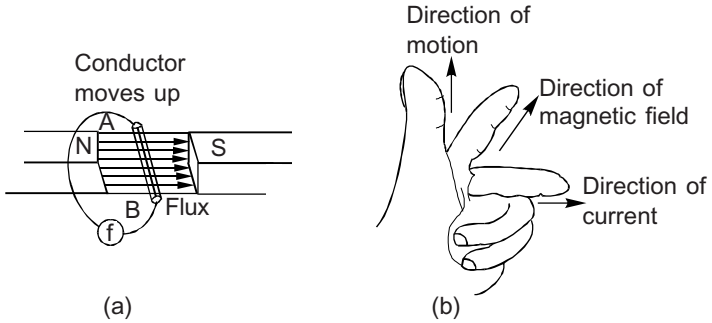


Fig. 3.15 Basic Principle of a Generator

If the straight conductor is bent in the form of a loop *ABCDEF*, and the same rotated in a magnetic field in the counter clockwise direction as in Fig. 3.16 emf is induced in each side of the loop when it cuts the magnetic flux. While the side *AB* of the loop moves up, the side FE moves down. In accordance with the left hand rule for generators given above, equal but opposite emfs are induced in the opposite sides as the loop rotates. However, the directions of the two emfs are such that they are in series with respect to the open ends of the loop and the effective voltage across the two ends of the loop is twice the voltage induced in either side of the loop. As the loop rotates in the magnetic field, it cuts the magnetic flux at varying angles and the rate of change of flux cutting the loop is different in different positions resulting in varying amplitudes of the induced emf. Figure 3.17 shows the loop in different positions during one complete rotation or revolution of the loop. When the plane of the loop is perpendicular to the magnetic field as in positions 1 and 3 (0° and 180°) the sides of the loop are passing between the flux lines and the rate of change of flux is minimum resulting in zero induced voltage. When the plane of the loop is parallel to the magnetic field as in position 2 and 4 (90° and 270°), the sides of the loop are cutting straight across the flux lines and the rate of change of flux is maximum resulting in maximum induced emf. However, the direction of the induced emf is opposite in position 2 and 4 as per the left hand rule. In all other positions of the loop the flux is cut at an angle and the induced emf varies between the maximum and minimum.

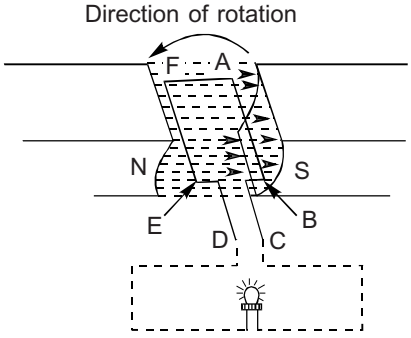


Fig. 3.16 Production of emf by Rotating a Loop in the Magnetic Field

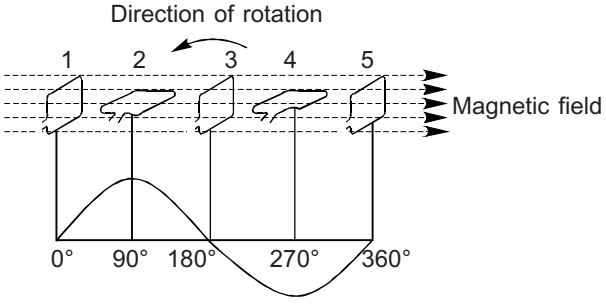


Fig. 3.17 Voltage Generated by a Loop Rotating in a Magnetic Field

3.4.2 Sine Wave

If the voltage produced by the loop at different angles during one complete revolution or rotation is plotted in the form of a graph, it takes the waveform shown in Fig. 3.17. The voltage starts at zero, increases to a maximum value (90°) and then decreases till it reaches zero value again (180°). At this point the voltage reverses polarity and increases again till it reaches its maximum again in the opposite direction (270°). It then decreases till it reaches zero value (360°). This cycle is then repeated at regular intervals of time. Such a waveform as described above is called a *sine wave.* This name is derived from the fact that voltage generated at any point is proportional to the sine of the angle between the magnetic field and the direction of motion of the rotating loop.

3.5 TYPES OF GENERATORS

There are two types of generators (i) direct current generators or DC generators, and (ii) alternating current generators or AC generators.

3.5.1 DC Generators

Basically all generators are AC generators, but the AC waveform or the sine wave produced by an AC generator can be converted into DC form by the *commutator and brushes* as in Fig. 3.18. A commutator consists of two semicylindrical pieces of conducting material separated by an insulating material. The brushes are made of soft conducting material that can easily slide on the commutator surface. Each half of the commutator is connected permanently to one end of the loop and the commutator rotates with the loop. Each brush presses against one segment of the commutator. The brushes remain stationary while the commutator rotates. The brushes press against opposite segments of the commutator and every time the voltage reverses polarity, the brushes also, switch from one segment to the other. This means that one brush always develops positive polarity and the other, negative polarity with respect to each other. Although the voltage developed across the brushes will be fluctuating as in Fig. 3.19, it will always have the same polarity and is a DC fluctuating voltage.

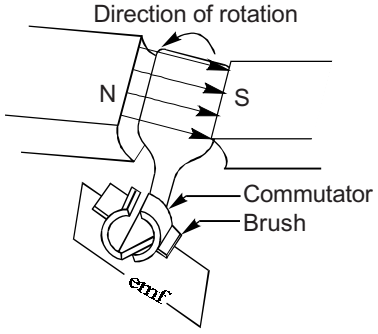


Fig. 3.18 DC Generator

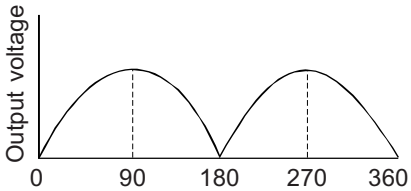


Fig. 3.19 Output Voltage of a DC

The variation in the voltage produced by a single rotating loop is called a *ripple* which makes the output unsuitable for any practical applications. The ripple can be reduced either by the use of a filter network or by the use of two rotating loops arranged at right angles to each other. The two ends of each loop are connected to two separate segments on the commutator which now has four segments. Since the two loops are positioned at right angles to each other, when the voltage in one loop is decreasing the voltage in the other loop is increasing, and vice versa. This makes the output voltage across the two brushes more steady with less variation and less ripple than in the case of a single loop. The output voltage waveform is shown in Fig. 3.20. By using more and more loops instead of one, the ripple in the generator output can be further reduced and the output voltage is very nearly a steady DC voltage as in Fig. 3.21. The combination of loops and commutator is called an *armature*. This is the rotating part of the generator and is also sometimes known as a *rotor*. The magnetic field is produced by a permanent magnet or by means of an electromagnet produced by field windings.

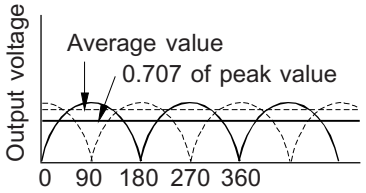


Fig. 3.20 Output Voltage with two Loops at Right Angles

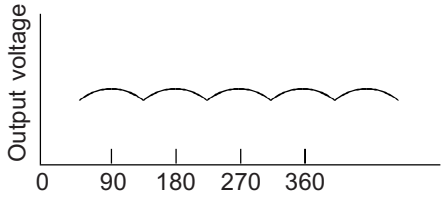


Fig. 3.21 Steady DC Voltage with more Loops then one

Dynamos A DC generator is also known as a dynamo. Small dynamos find extensive use in automobiles for charging the storage batteries fitted into the automobile. Since the speed of the dynamo varies with the speed of the automobile, a regulating device called a voltage regulator with a reverse current cut-out relay is fitted with the dynamo to keep the voltage and the current within safe limits at high speed and to disconnect the dynamo from the battery at low speed to avoid the battery discharging through the dynamo and thereby burning it out.

Small dynamos are also used with bicycles, scooters and motor-cycles mainly for lighting purposes.

3.5.2 AC Generators

As explained earlier a conductor in the form of a loop when rotated in a magnetic field produces a voltage which has the form of a sine wave as shown in Fig. 3.17. This is the basic principle of an AC generator. An AC generator is also called an *alternator*, because of the alternating voltage produced by it. Unlike a DC generator, an AC generator does not use a commutator. A pair of slip rings and a pair of graphite brushes enable the AC voltage produced to be connected to the external load as shown in Fig. 3.22. One slip ring is connected to each end of the rotating loop. The brushes remain stationary and one brush

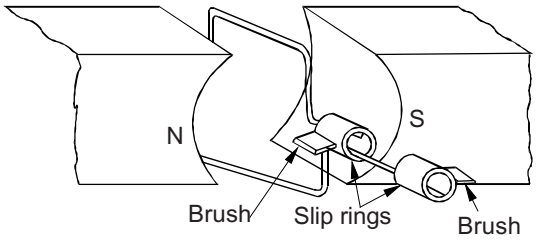


Fig. 3.22 AC Generator

presses firmly against each of the two rotating slip rings. The magnetic field is either produced by a permanent magnet or by an electromagnet. This arrangement of slip rings and brushes creates problems of insulation and sparking when large output powers are involved. Therefore, in most practical generators the field is rotated and the armature is kept stationary. In such generators armature coils are fixed permanently around the inner circumference of the housing of the generator while the field coils and the pole pieces rotate on a shaft within the stationary armature. The number of poles used for producing the magnetic field can also be more than two. The stationary armature has also the advantage that it enables the generator to increase its speed for obtaining higher output voltages. An armature consisting of a single loop produces very small voltage. In actual practice the armature consists of a number of coils each having more than one loop. If the armature coils are all connected in series as in Fig. 3.23 the output of the generator at any instant is the sum of the

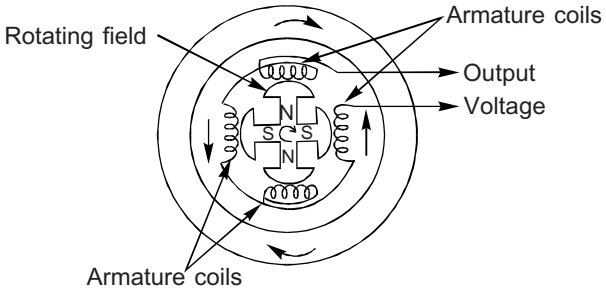


Fig. 3.23 Generator with a Stationary Armature (4 pole)

voltages induced in individual coils. A generator having its armature wound in this way is called a single phase generator. Greater efficiency can, however, be obtained by connecting the armature coils in other ways than in a single phase generator.

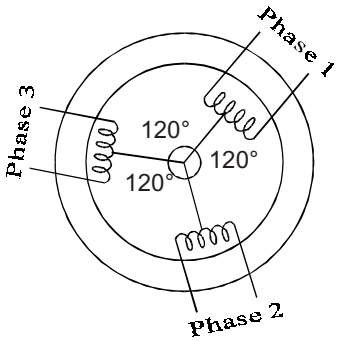


Fig. 3.24 Three Phase AC Generator

If the armature coils are so wired and spaced that the generator has two separate outputs that differ in phase by 90°, the generator is called a two phase generator. Each of the coils will have its own pair of slip rings and brushes. In a three phase generator (Fig. 3.24) there are three equally spaced winding and the three output voltages are all 120° out of phase with each other. Each winding is connected to its own pair of slip rings. Three phase generators have great practical importance when large AC output powers are obtained from the generator.

3.6 OTHER VOLTAGE SOURCES

3.6.1 Heat Generated Voltages—Thermoelectricity

Heat and electricity are both forms of energy. Heat energy can, therefore, be converted into electricity. If a junction of two dissimilar metals like copper and zinc is heated then a voltage appears across the colder ends of the metals. Copper develops a positive charge and zinc develops a negative charge as shown in Fig. 3.25. Electricity produced by heat is called *thermoelectricity* and the combination of metals used for the production of thermoelectricity is known as a *thermocouple*. *RF* meters used for the measurement of high frequency currents make use of a thermocouple. Heat produced by the *RF* current passing through a wire is used to heat the thermocouple. DC current so produced is measured by a DC meter as in Fig. 3.26.

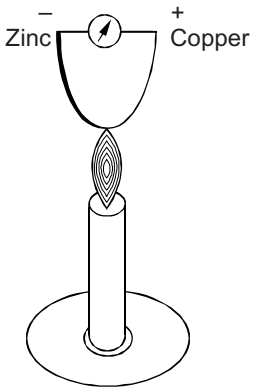


Fig. 3.25 Production of Electricity by Heating a Thermocouple

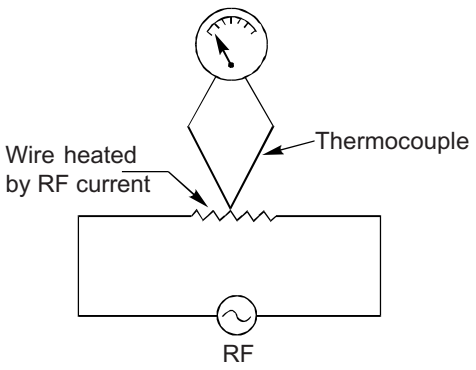


Fig. 3.26 RF Meter

3.6.2 Light Generated Voltages—Photoelectricity

Light is a form of energy and is supposed to consist of packets of energy called *photons*. When a light beam containing photons strikes certain metals and their alloys, electrons are emitted from the surface of the metal and are collected by another electrode placed near the metal surface. The metal surface that loses electrons develops a positive charge and the electrode that collects the electrons becomes negatively charged. The electricity produced by light is called *photoelectricity* and the electron emission produced by the action of light is called *photoemission.* Materials like cesium, lithium, selenium, sodium, potassium, cadmium and germanium that produce photoemission by the action of light are known as photosensitive materials.

If the two electrodes, one consisting of a photosensitive material and another for collecting the electrons is enclosed in an evacuated glass envelope, it constitutes what is known as a photoelectric cell or PE cell (Fig. 3.27). Photoelectric cells are used in control devices, such as exposure meters in photography, for sound reproduction in talkies and also in the manufacture of TV cameras.

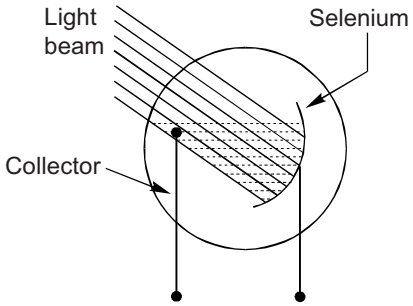


Fig. 3.27 Photoelectric Cell

3.6.3 Solar Cells

A solar cell makes use of the sun's light energy to develop small voltages in a photocell. A solar cell uses selenium as photosensitive material and iron as backing for collecting the electrons emitted by selenium. Selenium becomes the positive electrode and iron the negative electrode. Silicon is also used in solar cells as photosensitive material. A solar cell can develop only a small voltage of about 0.26V. To get sufficient output power from solar cells, a large number of solar cells are connected together to form a *solar battery* which is used as a power source for charging batteries in satellites.

3.6.4 Pressure Generated Voltages

When certain crystals like quartz, tourmaline and Rochelle salt are subjected to external pressure, positive and negative charges are built upon the opposite sides of the crystal as shown in Fig. 3.28. These charges are reversed when the direction of force or pressure is reversed. Electricity so produced by pressure is called *piezoelectricity* and the phenomenon is

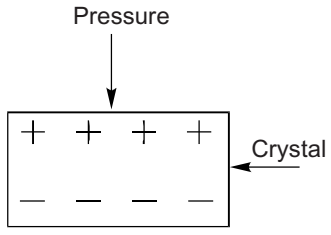


Fig. 3.28 Pressure Generated Voltages-Piezoelectricity

called the piezoelectric effect. A variable pressure like the one produced by sound waves will produce variable voltages on the two sides of the crystal. The voltages produced by the piezoelectric effect are extremely small but these can be amplified to produce certain desired effects. Piezoelectric crystals are used in the manufacture of crystal microphones, phono-pickups, loudspeakers and headphones.

3.7 A COMPLETE ELECTRIC CIRCUIT

A complete electric circuit normally consists of the following:

1. *Voltage Source:* Various types of voltage sources have already been described. The most common type of voltage or power sources are the storage batteries and the generators.
2. *Conductors or the Connecting Wires:* Conductors are necessary to provide a path for the current, or flow of electrons from the negative terminal of the source through the connecting wires and the load, back to the positive terminal of the source. No current can flow unless the path is provided by the conductors.
3. *Load:* This is the device that uses electrical energy to produce some useful effects like lighting a lamp, heating an electric iron, rotating a ceiling fan or driving an electric motor. The term load is used for the device that consumes electrical energy and also for the amount of current or power drawn by the load from the source.
4. *Switch:* When a load is connected to the source by some connecting wires and current flows from the negative terminal of the source through the conductor and the load and back to the positive terminal of the source, the circuit is said to be a closed circuit. If the complete path for the current does not exist due to a break in the conducting wire or otherwise, the circuit is called an open circuit. A device which is used to 'Open' or 'Close' an electric circuit is called a switch. Various types of switches such as toggle switches, knife switches, push button switches, and wafer switches are used in electrical and electronic circuits.

A complete electric circuit consisting of a source, conductor, load and switch is shown in Fig. 3.29. In drawing the circuit diagram of electric or electronic circuits, schematic symbols are used to represent the various parts of the circuit. A symbolic representation of an electric circuit is given in Fig. 3.30.

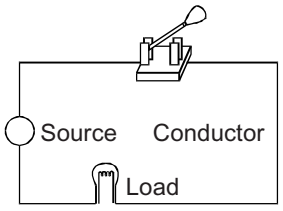


Fig. 3.29 An Electric Circuit

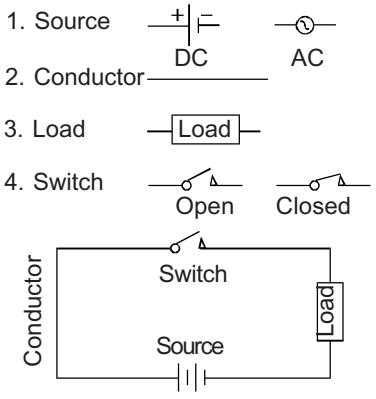


Fig. 3.30 Symbolic Representation of an Electric Circuit (1) Source (2) Conductor (3) Load (4) Switch

3.7.1 Earth Return Circuits

Earths surface is a very good conductor. This fact is often utilised to treat "earth return" as a part of the circuit. This not only simplifies a circuit but also results in a lot of saving in wiring. Metal chassis in the case of electronic equipment and hull of a ship can be used for this purpose.

For example in Fig. 3.31(a), the battery shown is being used to ring a bell.

The positive terminal of the battery is joined to the bell through a switch but the connection from the negative battery terminal is taken by using an earth return between it and the bell.

Similarly, in Fig. 3.31(b), the dynamo is connected to a number of lamps in parallel through a single line wire but the return connection is made by earthing the negative dynamo brush and one side of the switch. However, the insulation of the single wire should be reliable to avoid earthing of the dynamo.

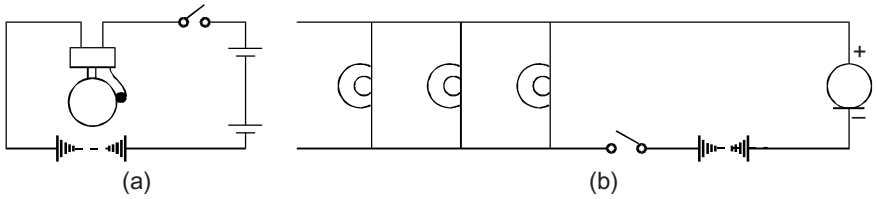


Fig. 3.31 Earth Return

3.8 ALTERNATING CURRENT (AC)

Alternating current has already been defined as current which reverses direction at regular intervals.

Advantages of Alternating Current

3.8.1 Reduction in Transmission Power Losses

The wires connecting the load to the voltage source carry the same current I

that flows through the load. If these connecting wires have a resistance R, the heat produced in these wires is $I^2 R$ and represents wasted energy and is called $I^2 R$ loss. This should be kept to the minimum.

For carrying electrical power from the generating station to the actual users, thousands of miles of transmission lines have to be used and the I^2R losses on these lines can be considerably reduced if the current carried by these transmission lines is relatively small. In the case of AC power, the voltage E at the generating end can be stepped up and the current considerably stepped down for a fixed amount of power ($E \times I$) and the I^2R losses brought down to the minimum. At the receiving end, the voltage E is stepped down and the current stepped up again to the values actually required by the users of power. Thus, the same amount of energy ($E \times I$) is transmitted over long distances, without much waste of energy on the transmission lines. This cannot be done in the case of DC power where the stepping up and stepping down of voltages and current is not possible.

3.8.2 Conversion of Sound Waves into Electrical Voltages

AC not only reverses direction at regular intervals but its value or magnitude also varies evert instant like the pressure of sound waves varies in air. This makes it easy for sound waves to be converted into corresponding electrical waves by means of a device called a microphone. This conversion of sound waves into AC electrical voltages has made electrical communication possible.

3.8.3 Radiation of Energy for Radio Communication

Rapidly changing AC currents flowing through a conductor can radiate energy in the form of electromagnetic waves which form the basis for radio and wireless communication.

3.8.4 Running a Radio or TV Set

An AC source is necessary to produce resonance in electrical circuits containing inductance and capacitance for tuning a radio or TV set and other electronic circuits.

It would thus appear that AC not only can do all that DC can do but also posses certain characteristics which DC does not. This does not, however, mean that DC should be completely replaced by AC in all fields. DC is essential for the operation of vacuum tubes and transistors which form vital components in all electronic equipments. Often we need to convert AC into DC for the successful operation of such equipments.

3.9 AC VOLTAGE SOURCES

By far the best and the most common source of AC voltage is the AC generator or alternator. The bulk of AC power consumed in the world is produced by alternators. In an alternator, a loop or armature is made to rotate in a magnetic

field. The voltage induced in the rotating loop or armature not only varies with the angle at which the conductors of the loop cut the magnetic field but also completely reverses polarity at the slip rings and brushes at the end of each half revolution or twice during one complete revolution. By definition, the voltage developed at the terminals of the loop is an AC voltage. The voltage induced in the loop at different positions of the loop as it goes through one complete rotation or revolution is shown in Fig. 3.32. AC voltages of low power can also be produced by electronic devices called oscillators.

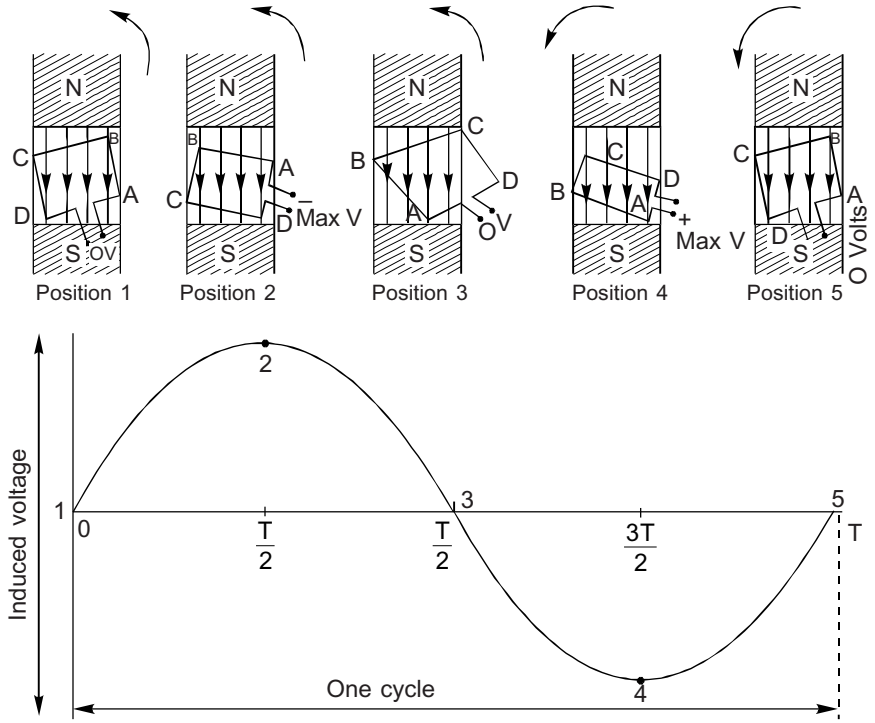


Fig. 3.32 Production of AC Voltage

3.9.1 AC Waveforms

A waveform is the representation or picture of any voltage or current that varies with time. Such a representation is generally constructed on a graph paper where time intervals are represented along the *x*-axis and the value of the current or voltage at particular instants of time are represented along the *y*-axis on a suitable scale. A smooth curve passing through the points plotted on the graph paper at different instants represents the waveform of the voltage or the current as the case may be. An AC waveform of current or voltage with time along the *x*-axis is shown in Fig. 3.33(a).

Since AC voltages are produced as a result of rotation of a loop in magnetic field, the position of the loop at different times in the course of its rotation is shown by angles in degrees or radian represented along the *x*-axis, and the

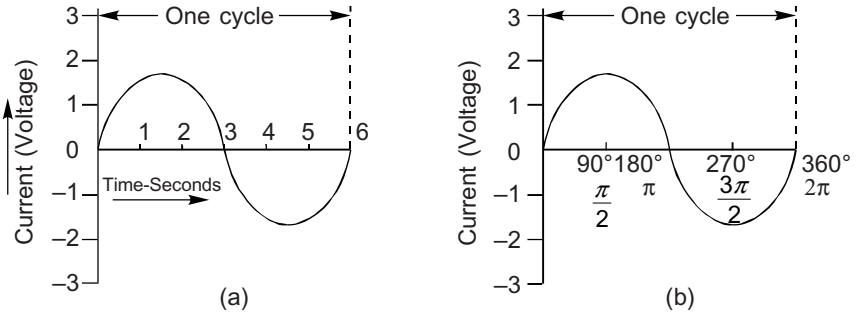


Fig. 3.33 (a) AC Waveform with time along x-axis (b) AC Waveform with Angular Rotation along x-axis

value of the current or voltage is again shown along the y-axis as before. Such a waveform is shown in Fig. 3.33(b). A complete rotation is represented by 360° or 2π radians. A radian is the angle subtended at the centre of the circle by an arc whose length is equal to the radius of the circle as in Fig. 3.34. One radian equals 57.3° if the value of π (pi) is taken as 3.1416.

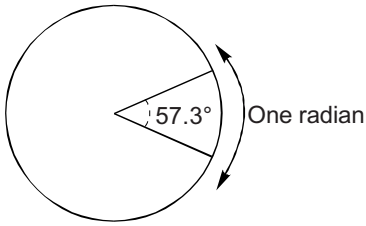


Fig. 3.34 A Radian

3.9.2 Sine Wave

For the production of AC voltage by the rotation of a loop in a magnetic field the rotating loop can be represented by a rod OA of fixed length equal to the radius of the circle and which rotates in a counter-clockwise direction about the end O representing the centre of the circle. The heights A_1, B, A_2O and A_3B of the point A above the x-axis through O, represent the voltage or the current generated in different positions of the rotating rod OA. This height is plotted on the y-axis in different positions represented by angles shown along the x-axis. The curve drawn through the various points representing the voltages at different positions of the rotating rod is the AC waveform generated by the rotating rod OA as shown in Fig. 3.35. This is the characteristic waveform that represents an AC voltage or AC current.

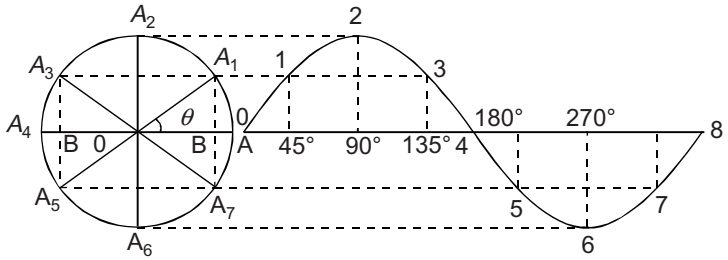


Fig. 3.35 AC Waveform produced by a Rotating Vector

In any position of OA, the angle through which the rod has rotated from its initial position OA is θ and

$$\sin\theta = \frac{A_1 B}{OA_1}$$

The value of OA_1, OA_2, OA_3, etc. is fixed and hence the voltage produced at any time is proportional to sin θ, θ being the angle between the magnetic field and the direction of motion of the loop. Hence an AC voltage waveform so produced is known as the *sine wave.*

The characteristics of a sine wave are:

1. It starts from zero value and increases to maximum value in one direction. It then starts decreasing till it reaches zero value again. At this point the voltage reverses polarity and increases till it reaches the same maximum value in the opposite direction. It then decreases until it reaches zero value again. This completes one cycle which is repeated again and again.
2. A sine wave is symmetrical about the x-axis. The position of the wave above the x-axis is exactly of the same shape, height and width as the position below the x-axis. If a wave is not symmetrical about the x-axis, it is not a pure sine wave and will not represent pure AC voltage or current.

When a resistive load is connected across a sine wave or AC voltage Fig. 3.36(a), the current that flows in the circuit is also a sine wave. Hence, a sine wave can represent a pure AC voltage or a pure AC current. The wave shape is the same in the two cases but the magnitude of the voltage or current may be different as in Fig. 3.36 (b).

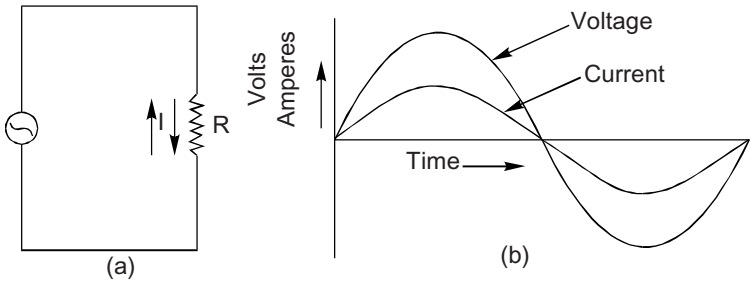


Fig. 3.36 Sine Waves representing AC Voltage and AC Current (a) AC Current (b) Voltage and Current Waveforms

3.9.3 Sine and Cosine Waves

A sine wave is the basic form of AC voltage or current. This waveform is produced by the rotation of the armature in a generator and the amplitude of voltage generated at any instant is proportional to the sine of the angle of rotation of the armature. In trigonometric values, the cosine of an angle varies in the same way as the sine of the angle except for a shift of 90° in values. In Fig. 3.37, a cosine wave is shown in dotted lines. It will be seen that the

variations of the sine wave and the cosine wave are the same but the maximum of the sine wave occurs 90° after the maximum of the cosine wave. In other words, the two wave forms are displaced in time by 90°.

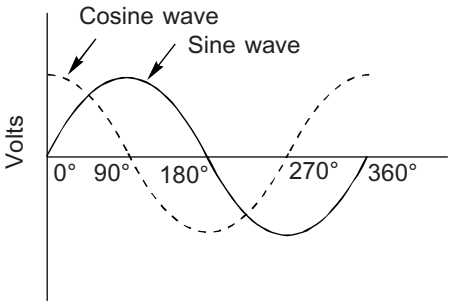


Fig. 3.37 Sine Wave and Cosine Wave

Both the cosine wave and sine wave are known as *sinusoidal waves* or simple sinusoids. There are other AC waveforms which are not sinusoidal and these are called *non-sinusoidal* waveforms. Two important forms of non-sinusoidal waveforms that are used in electronic work in general, but TV work in particular, are the *square waveform* and the *sawtooth wave-form* and these will be briefly described here.

3.9.4 Square Wave

A square waveform is shown in Fig. 3.38. In this type of non-sinusoidal waveform, the current or voltage does not vary in magnitude continuously but rises instantly from zero value to some maximum value at point 1. It then stays steady at this maximum value for a period of time as from point 1 to 2. At the end of this period of time the current or the voltage suddenly drops down to zero value at point 3, reverses direction and suddenly or instantly rises to its maximum value in the opposite direction at point 4. It stays at this negative value for a period of time and on reaching point 5, again drops down instantly to zero value at point 6. This waveform is repeated again and again at regular intervals.

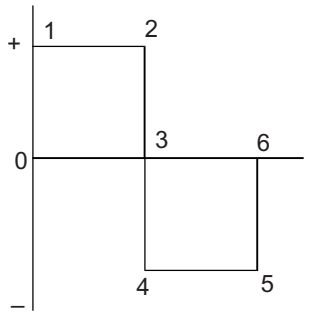


Fig. 3.38 A Square Wave

Square waveforms are produced in electronics by switching circuits. The voltage or current rises suddenly from the zero value to maximum value when the switch is closed and stays at this value so long as the switch remains closed. When the switch is opened the current or voltage falls to zero value. When the switch is closed again with reversed polarity, the current runs to the negative maximum and again stays at this value till the switch is opened again when the current or voltage drops down to zero value.

3.9.5 Sawtooth Wave

This is another form of non-sinusoidal wave which is shown in Fig. 3.39. The sawtooth waveform derives its name from the fact that it resembles the tooth of a common wood saw. In this type of waveform the rise of voltage from zero to maximum value

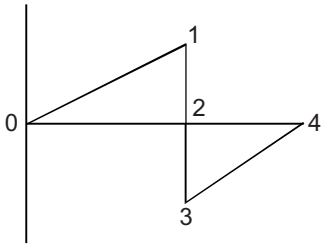


Fig. 3.39 A Sawtooth Wave

is linear and the drop from maximum to zero value is sharp. A linear change of voltage or current is one in which the value of voltage or current changes by equal amounts in equal intervals of time. Such a linear change is represented by a sloping straight line.

In the sawtooth wave the voltage starts from zero and linearly increases to its maximum value at point 1. It then instantly drops down to zero value at point 2, reverses direction and again increases to its maximum value in the reverse direction at point 3. From its maximum negative value it rises linearly till it reaches zero value again at point 4. This cycle will be repeated at regular intervals.

Sawtooth waves can be generated by electronic circuits and find many practical applications such as sweep voltages in cathode-ray oscilloscopes and for the deflection of electron beams in television picture tubes.

3.9.6 DC Wave

We have seen that in an AC voltage or current both the magnitude and direction change and the direction is reversed periodically with the magnitude varying between zero and a certain maximum value in both directions. Such an AC voltage or current is represented by a sine curve. In the case of DC, however, both the magnitude and the direction remain constant without undergoing any changes. A DC waveform would therefore be a straight line parallel to x-axis and at a distance above the x-axis (positive side only) representing the value of the current or voltage as shown in Fig. 3.40.

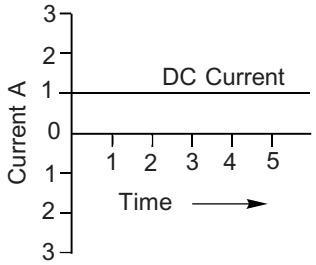


Fig. 3.40 DC Waveform

In certain other types of current waveforms, the magnitude or value of current varies but the direction never changes. The entire waveform remains completely above the x-axis and never becomes negative. Such a current is called a *fluctuating DC*. Sometimes the shape of a fluctuating DC resembles an AC waveform and its behaviour is also more like an AC than DC. Two common forms of fluctuating DC are depicted in Figs 3.41(a) and (b).

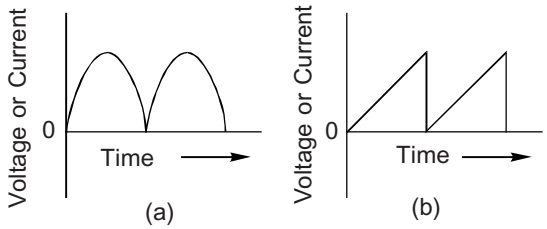


Fig. 3.41 Fluctuating DC

A fluctuating DC sometimes appears as a combination of AC and DC as shown in Fig. 3.42. Here the AC component of the combination varies with reference to a DC level. In this fluctuating DC the waveform is similar to the

AC waveform except that it is entirely on the positive side of the *x*-axis, though this is not the reference level for the AC variations. Such combinations of AC and DC waveforms occur in electronic circuits and the two can be separated by devices known as capacitors and transformers as will be described in the chapters to follow.

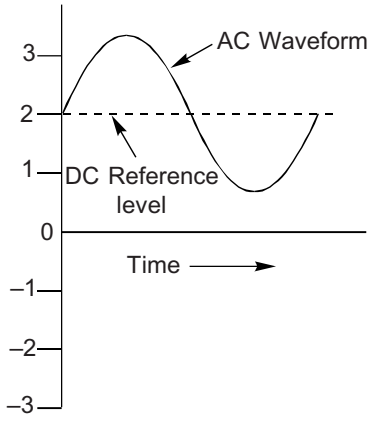


Fig. 3.42 Combination of AC and DC

3.9.7 Voltage and Current Values in a Sine Wave

The value of voltage or current remains constant in DC, but in AC which is represented by a sine wave the value of voltage or current changes every instant and this value at any particular instant is called *instantaneous value.* The maximum instantaneous value reached by the voltage or current is called the *peak value* or the *amplitude* of the voltage or current. Figure 3.43 shows two maximum or peak values in each cycle–a positive peak value and a negative peak value which occur at 90° and 270° of the rotation of the loop generating the sine wave. A value known as peak-to-peak value is twice the peak value on a sine wave. The peak-to-peak value is the distance from the maximum positive value to the maximum negative value.

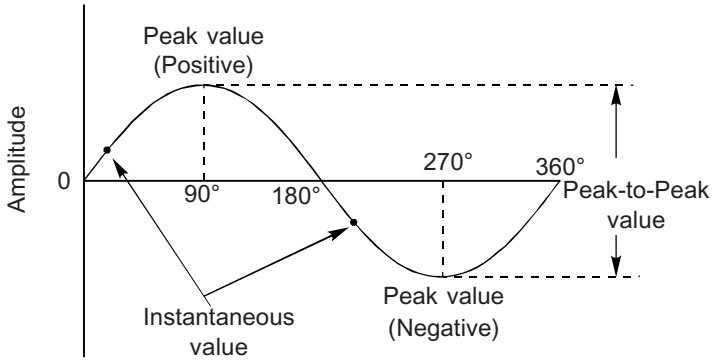


Fig. 3.43 Instantaneous, Peak and Peak-to-peak Values in a Sine Wave

Although the three values of voltage or current (instantaneous, peak and peak-to-peak values) mentioned above are generally useful in defining a sine wave, there are two other values of AC voltage or current which are more commonly used in connection with AC sine waves. These values are known as *average value* and *effective value* or *root mean square (rms) value.*

Average Value The arithmetical average of the various instantaneous values of voltage or current in a sine wave is obtained by adding together the values at different angles in a half cycle and dividing the sum of the values by the number of values taken. The half cycle over which this average is taken is also

known as alternation. The average is always taken over one alternation or half cycle and not over a full cycle. If taken over a full cycle the positive and negative values in each half cycle would cancel out and the result will be zero. It can be proved mathematically that the average value of a half cycle of a sine wave is 0.637 of the peak value. Expressed as an equation

$$\text{Average value } (E_{av}) = 0.637 \times \text{Peak value } (E_{pk}).$$

For example, the average of an AC voltage having a peak value of 120 V will be $120 \times 0.637 = 76.44$ V.

Effective Value or Root Mean Square (rms) Value The effective value of an AC voltage or current is that value which will produce the same amount of heat in a resistance as would be produced by DC voltage or current of the same value. As the heat produced is given by the relation $P = I^2R$ or $P = E^2/R$, in calculating the effective value we get the average or mean of the squares of instantaneous values by adding all the squares of the instantaneous values over a half cycle and dividing by the number of values taken and then taking the square root of this average value. The effective value so obtained is called the root mean square or rms value.

It can be shown mathematically that the

Root Mean Square (rms) value $(E_{rms}) = 0.707 \times$ Peak value (E_{pk})

Thus, the rms voltage for an AC with a peak voltage of 100 V will be

$$E_{rms} = 0.707 \times 100 = 70.7\text{V}$$

When we speak of a line voltage of 230 V in our household electric supply, we are actually referring to the rms voltage.

If we know the rms voltage or current in any circuit, the peak voltage can be calculated from the following formula:

$$\text{Peak voltage } (E_{pk}) = \frac{1}{0.707} \times \text{rms voltage } (E_{rms})$$

or

$$E_{pk} = 1.414 E_{rms}$$

The peak voltage in an AC supply line with an rms voltage of 230 will be

$$E_{pk} = 1.414 \times 230 = 325.2 \text{ V}$$

Any of the components or materials used in a 230 V AC circuit should be able to withstand a peak voltage of 325 V or more.

Figure 3.44 shows the relationship between peak value, average value and root mean square value.

EXAMPLE: Peak-to-peak voltage of a sine wave is 400 V. Calculate the average value and the rms value of the voltage.

$$\text{Peak-to-peak voltage} = 400 \text{ V}$$

$$\text{Peak voltage } \frac{400}{2} = 200 \ V$$

$$\text{Average voltage} = 0.637 \times \text{peak voltage}$$

$$E_{av} = 0.637 \times 200 = 127.4 \text{ V}$$

$$\text{rms voltage} = 0.707 \times \text{peak voltage}$$

$$E_{rms} = 0.707 \times 200 = 141.4\text{V}$$

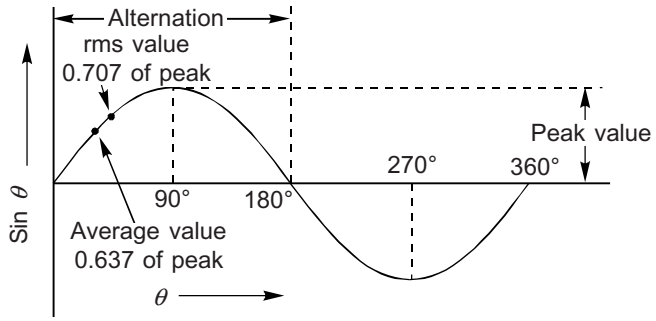


Fig. 3.44 Relationship between Peak, Average and rms Values in a Sine Wave

3.10 OTHER IMPORTANT TERMS CONNECTED WITH A SINE WAVE

Frequency A sine wave completes one *cycle* when its value varies from zero to maximum and back to zero in the positive direction and again from zero to maximum and back to zero in the negative direction.

The number of such cycles completed in one second is known as the *frequency* of the sine wave. AC power supply in India has a frequency of 50 cycles per second which means the AC generator supplying power for household use and for industrial purposes completes 50 revolutions in one second. In the U.S. the frequency of power supply is 60 cycles per second. Electronic equipment like tape recorders and TV sets designed for use at 60 cycles per second will not work satisfactorily, when used in India where the power supply is 50 cycles per second or 50 c/s.

The unit of frequency is called hertz (Hz), which is equal to one cycle per second. Radio waves have a much higher frequency which is expressed in kilohertz (kHz) and megahertz (MHz).

Audio Frequency and Radio Frequency Alternating currents which have lower frequencies ranging from 20 to 20,000 Hz are called audio frequencies (AF) because they correspond to the frequencies of sound waves which can be heard by the human ear. However, the actual range of audio frequencies seldom goes beyond 16,000 Hz. *Radio frequencies* (RF) are frequencies much above the audible range of frequencies and generally include frequencies from 3 to 30 MHz. Radio frequencies are also sometimes termed as *high frequencies* (HF).

The time taken by a sine wave to complete one cycle is called its period. If T is the period of a sine wave, then

$$T = \frac{1}{f} \text{ where } f \text{ is the frequency in hertz.}$$

The period of the AC power supply which has a frequency of 50 Hz is 1/50 seconds (s) or 0.02 seconds. Radio waves which have a much higher frequency have a very low period generally expressed in microseconds (μs). A radio wave with a frequency of 1 MHz has a period equal to

$$\frac{1}{1\text{ MHz}} = \frac{1}{10^6\text{ Hz}}$$

or 10^{-6}s which is 1μs.

Wavelength All radio waves travel with a fixed velocity which is the same as the velocity of light (3×10^{10} cm/s). The distance travelled by a wave in one cycle is called *wavelength* (Fig. 3.45) and is usually expressed in metres or centimetres. Since one cycle is completed in a time equal to the period of a sine wave, it would seem that there is a definite relationship between the frequency and the wavelength. The higher the frequency, the shorter the wavelength and vice versa. Moreover, the frequency f, the wavelength λ and the velocity c of the wave are interrelated according to the following formulas:

$$f \times \lambda = c$$

or $$\lambda = \frac{c}{f}$$

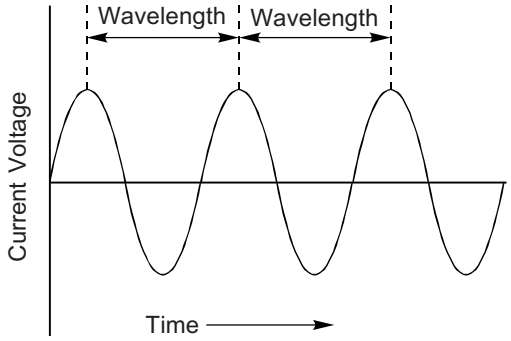


Fig. 3.45 Wavelength

In the case of radio waves or electromagnetic waves, c is equal to 3×10^{10} cm/s. Therefore

$$\lambda = \frac{3 \times 10^{10}}{f}\text{ cm} \quad \text{or} \quad \frac{3 \times 10^8}{f}\text{ m}$$

Sound waves are also sinusoidal waves but the velocity with which sound waves travel is not the same as the velocity of radio waves. Sound waves travel with a velocity of about 330 m/s (1100 ft/s) in air under average conditions and as such the wavelength of sound waves of a particular frequency will be shorter compared to the wavelength of an electromagnetic wave of the same frequency. In the case of sound waves,

$$\lambda(\text{sound}) = \frac{\text{velocity (sound)}}{\text{frequency}}$$

$$\lambda(\text{sound}) = \frac{330}{f}\text{ m}$$

EXAMPLE: A sound wave and an electromagnetic wave both have the same frequency of 1kHz. Calculate the wavelength in each case.

In the case of an electromagnetic wave

$$\lambda = \frac{3 \times 10^8 \text{ m/s}}{1 \text{ kHz}} = \frac{3 \times 10^8}{10^3} \text{ m}$$

$$= 3 \times 10^5 \text{ m} \quad \text{or} \quad 300000 \text{ m}$$

In the case of a sound wave

$$\lambda \text{ (sound)} = \frac{330 \text{ m/s}}{1 \text{ kHz}} = \frac{330}{10^3} = 0.33 \text{ m}$$

Phase If two generators *A* and *B* start rotating at the same time and with the same speed, they will complete one rotation in the same time and each will generate a complete sine wave. The two sine waves will not only begin simultaneously but will also pass through the maximum values and zero values at the same time as in Fig. 3.46. Two such waveforms are said to be in step or in phase.

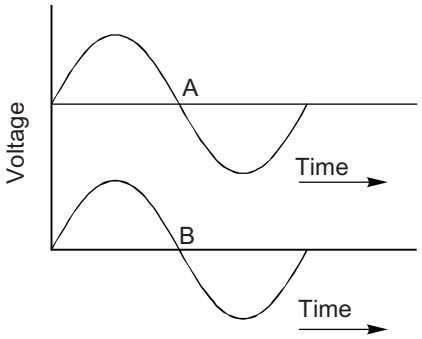


Fig. 3.46 In-phase Waveforms

Phase Angle In a sine wave the instantaneous value of voltage or current depends on the angle of rotation of the rotating loop or armature. This angle of rotation is generally represented by the Greek letter θ (theta) and is called the phase angle of rotation. The relation between the peak value of voltage and its instantaneous value at any phase angle θ is represented by the formula.

$$E_{mst} = E_{pk} \sin \theta$$

Thus, the value of voltage at phase angles of 90° and 270° will be maximum positive and maximum negative respectively as sin θ is 1 for 90° and –1 for 270°. The value will be zero at 0°, 180° and 360° as the value of sin θ is zero at all these three angles. In this way any point on the sine wave can be referred to by the particular angle.

Phase Difference If the two generators *A* and *B* referred to earlier do not start rotating simultaneously but *B* starts a little later than *A*, the maximum and minimum of the two sine waves will not coincide and the maximum and minimum of *B* will occur after the maximum and minimum of *A*. The outputs from the two generators are said to be out-of-step or out-of-phase. The phase difference is the angle through which *A* has already rotated before *B* starts rotating. If the two generators are moving with the same speed, the phase difference is maintained throughout the complete cycle and in all successive cycles.

The phase difference is expressed as a fraction of the cycle like half cycle, quarter cycle or one-eighth cycle but a better way of expressing the phase

difference is in terms of the angle of rotation in degrees. A complete rotation being 360°, a phase angle of a half cycle is 180°, a quarter cycle of 90° and one eighth of a cycle is 45°. In the case of two out-of-phase voltages or currents the one that reaches its maximum and minimum values ahead of time is said to lead and the one that is behind in time is said to lag. The angle of lead or lag is the phase angle between the two out-of-phase voltages or currents.

One of the important cases of out-of-phase sine waves is the case where the phase difference between the two sine waves is 90° as shown in Fig. 3.47. Here the sine wave *A* that leads has reached its maximum value when the sine wave *B* starts from zero. At any instant of time, sine wave *A* will have a value that sine wave *B* will have at 90° or a quarter of a cycle later. Two such sine waves with a phase difference of 90° are said to be in quadrature.

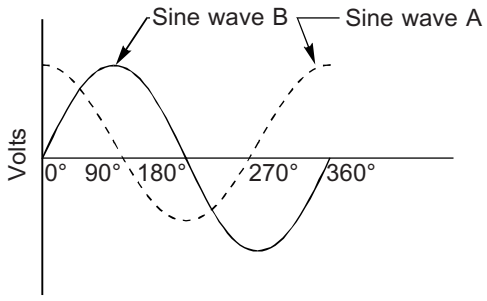


Fig. 3.47 Two Sine Waves in Quadrature

Two sine waves whose phase angles are compared must have the same frequency. The amplitudes of the two sine waves can, however, be different. We can compare the phase difference of the two AC voltages, two currents or a voltage and a current.

3.11 GRAPHICAL REPRESENTATION OF AC VOLTAGES AND CURRENTS—VECTOR DIAGRAMS

There are two types of physical quantities—those that have magnitude only, like rupees, apples and chairs and can be completely described by a number such as Rs. 10, 9 dozen apples or 5 chairs. Physical quantities that have magnitude only are called *scalar quantities* or simply scalars. Scalar quantities can be added or subtracted arithmetically, Rs. 20 and Rs. 10 when added will always make Rs. 30. Similarly, 12 apples less 8 apples will always make 4 apples.

The other type of physical quantities are those that have both magnitude and direction and are called vector *quantities*. Force, velocity, AC voltage and current are all examples of vector quantities. A velocity of 30 m/h has no meaning unless the direction of motion is also mentioned. Take the case of two trains, one moving with a velocity of 50 m/h and the other with a velocity of 40 m/h. If we want to know how far distant the two trains will be after half an hour of starting from the same point, we must know the direction in which each

train is moving besides knowing the magnitude of the velocity of each train. Thus, if both the trains are moving in the same direction, they will be 5 m from each other (50-40) after 1/2 hour. If the two trains are moving in opposite directions, they will be 45 m (50 + 40)/2 from each other after 1/2 hour. If the trains are moving in any other direction, the distance between them after half an hour will depend on the angle between the two directions in which they are actually moving. Similarly, if we want to study the effect of two forces acting on a body, merely knowing the magnitude of the two forces will not provide the correct answer. It will also depend on the direction in which each force is acting. If two equal forces act on a body in the same direction, the effect will be equivalent to a force of double the strength, but if the forces are acting in opposite directions the effect will be nil, as the two equal and opposite forces merely cancel each other out.

In an AC voltage, not only is the magnitude of the voltage different at different instants in a cycle but its direction also changes from time to time. AC voltage is, therefore, a vector quantity and so is AC current whose behaviour is similar to that of voltage.

Vector Representation Since a vector has got both magnitude and direction, it can be represented by a straight line whose length, on a convenient scale, indicates the magnitude of the vector and an arrow placed at the end of the straight line indicates the direction of the vector. For example, a velocity of 30 m/hr in the west to east direction will be represented by a straight line 6 cm (5 m to a cm) long drawn in the west to east direction with an arrow at the end pointing in the direction of motion as in Fig. 3.48.

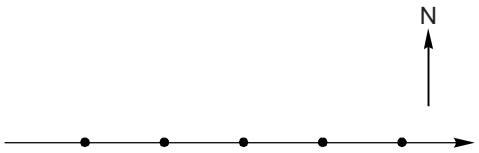


Fig. 3.48 Vector representation of a Velocity of 30 miles per hour

In vector representation, the *x*-axis is generally taken as the reference line and straight lines of appropriate length with arrow heads are drawn, making the required angles with the reference line, to represent vectors in different directions. In Fig. 3.49(a) the vector is in the direction making an angle of 30° with

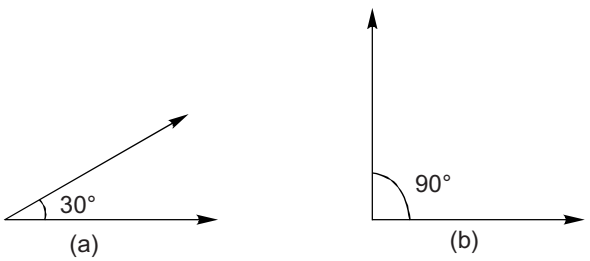


Fig. 3.49 Vectors in different Directions

the *x*-axis and, in Fig. 3.49(b), the two vectors are at right angles to each other. So far AC voltages and currents have been represented by sine waves but it is simpler to represent these quantities by vectors.

Addition and Subtraction of Vectors Like all other physical quantities, vectors can also be added, subtracted, multiplied and divided. In electrical and electronic work, we have to deal mainly with addition and subtraction of vectors and, as such, only methods for the addition and subtraction of vectors will be described here, the actual method depending on the relative direction of the vectors concerned. Since the vectors have both magnitude and direction, the methods for the addition and subtraction of vectors are geometrical methods and not mere mathematical or arithmatical additions or subtractions as in the case of scalars. A vector obtained by the addition or subtraction of vectors is called the *resultant*.

Addition of Vectors

Addition of Vectors having the Same Direction When vectors having the same direction are to be added, the magnitude of the resultant vector is obtained by adding the magnitudes of the individual vectors. The direction of the resultant is the same as the direction of the individual vectors as in Fig. 3.50(a).

Figure 3.50(b) represents the vectorial addition of the two batteries which are connected in series and their voltages are in the same direction.

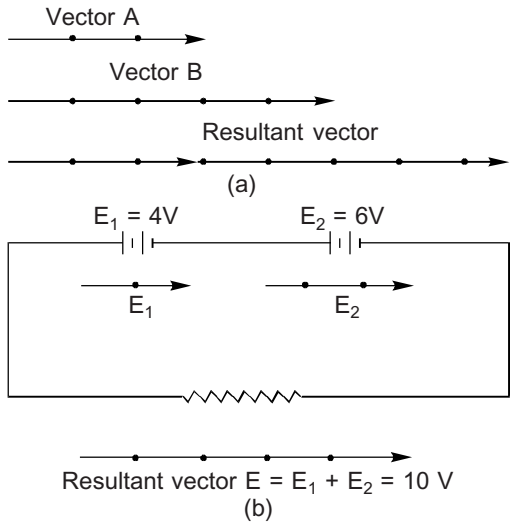


Fig. 3.50 (a) Addition of Vectors having the same direction (b) Vectorial Addition of Two Batteries connected in Series-Aiding

Addition of Vectors having Opposite Direction When vectors having opposite directions are to be added, the magnitude of the resultant vector is obtained by subtracting the magnitude of the smaller vector from the magnitude of the larger vector. The direction of the resultant vector is the same as the

direction of the larger vector as in Fig. 3.51. An example of the addition of the two vectors having opposite direction is two batteries E_1 and E_2 connected in series opposing as shown in Fig. 3.52.

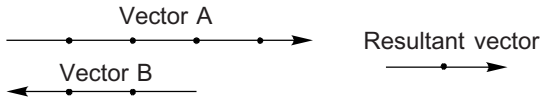


Fig. 3.51 Addition of Vectors having Opposite Directions

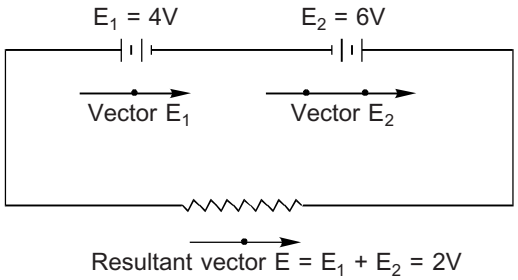


Fig. 3.52 Vector Addition of Two Battery Voltages in Series opposing

Addition of Vectors Inclined to Each Other To add vectors which have directions making an angle with each other, the resultant is found by a method known as the parallelogram method of vectorial addition. A parallelogram is a geometrical figure whose opposite sides are parallel and equal in length.

The two vectors are made to form the two sides of a parallelogram which is completed by adding the two remaining sides as in Fig. 3.53. The resultant vector is represented by the diagonal of the parallelogram both in magnitude and direction. The length of the diagonal gives the magnitude of the resultant on the same scale as used for individual vectors and the direction of the resultant is defined by angle θ made by the diagonal with the vector B.

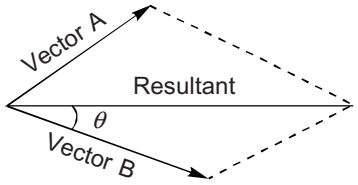


Fig. 3.53 Parallelogram method of Vector Addition

In case of more than two vectors, the resultant of any two vectors is found first and the resultant is combined with another vector and the process continued till the final resultant is obtained.

EXAMPLE: Find the resultant of two AC voltages E_1 and E_2 at an instant when their magnitudes are 5V and 3V respectively with E_2 leading E_1 by a phase angle of 60°.

Draw a horizontal straight line AB, 5 cm long to represent E_1 in magnitude and direction. Draw another line AD, 3 cm long making angle of 60° with AB in the counterclockwise direction to represent E_2. Complete the parallelogram $ABCD$ and join AC as in Fig. 3.54. Measure the length of AC in centimeters and convert it into volts on the basis of 1 cm = 1 V. This will give the

magnitude of the resultant of E_1 and E_2. For the direction of the resultant, measure the angle θ in degrees which will give the direction of the resultant referred to the direction of AB. The magnitude and direction of the resultant is given by the vector AC.

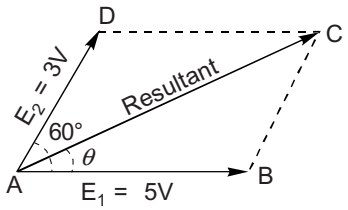


Fig. 3.54 Vector addition of AC Voltage

Addition of Vectors at Right Angles to Each Other. A very special case of addition of two vectors is the one when the vectors are at right angles to each other or when they are 90° apart. The resultant can be found by the parallelogram method as explained above but a simpler solution is possible in this case because the parallelogram becomes a rectangle and the diagonal of the parallelogram which is the resultant becomes the hypotenuse of right angled triangle whose two sides represent the vectors to be added as shown in Fig. 3.55(a).

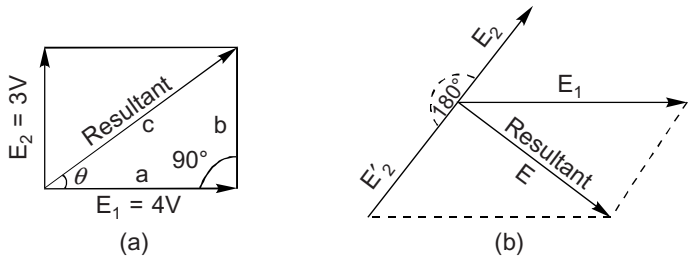


Fig. 3.55 (a) Addition of Two Vectors making an Angle of 90° with each other (b) Subtraction of Two Vectors

A mathematical solution is possible by the use of the well known Pythagorean theorem which states that in right angled triangle.
$c^2 = a^2 + b^2$ where c is the hypotenuse and a and b are the two sides of the triangle.
Thus, if $E_1 = 4$ V and $E_2 = 3$ V, and if E is the resultant,

$$E^2 = E_1^2 + E_2^2$$

Resultant $$E = \sqrt{E_1^2 + E_2^2} = \sqrt{4^2 + 3^2} = \sqrt{16 + 9} = \sqrt{25} = 5V$$

The direction of the resultant 5 V is given by the angle θ, whose value can be calculated by using the relation

$\sin \theta = \frac{b}{c} = \frac{3}{5} = 0.6$. This is the sine of an angle whose approximate value from the trigonometric tables is 37°.

As will be explained later such cases of two vectors having a phase difference of 90° are quite important in electrical and electronic circuits using coils and capacitors because a phase difference of 90° exists between the voltage and current in these circuits.

Subtraction of Vectors The subtraction of two vectors is similar to the addition of vectors except that the vector to be subtracted is reversed in direction by rotating it through 180° but keeping the same magnitude. The two vectors are then added using one of the appropriate methods of vector addition already explained. In Fig. 3.41(b) if vector E_2 is to be subtracted from vector E_1, rotate E_2 through 180°, so that E'_2 becomes equal but opposite to E'_2. Find the resultant of E_1 and E^1 by the parallelogram method of vector addition. The diagonal E of the parallelogram will represent the resultant vector both in magnitude and direction.

3.11.1 Ohm's Law in AC Circuits

When an AC voltage is applied to a circuit containing only a resistance, the voltage and current are in phase and Ohm's law is applicable to the circuit in the same way as in a DC circuit. However, in a DC circuit, there is only one value of voltage or current but in an AC circuit the voltage and current values can be *peak, average, effective* or *instantaneous* value. Hence while applying Ohm's law to a resistive AC circuit, care must be taken to use the same type of value for the voltage and current. Ohm's law can, therefore, have any of the following forms in a resistive AC circuit.

$$\text{DC values } \frac{E}{I} = R$$

$$\text{Instantaneous values } \frac{E_{inst}}{I_{inst}} = R$$

$$\text{Peak value } \frac{E_{pk}}{I_{pk}} = R$$

$$\text{rms values } \frac{E_{rms}}{I_{rms}} = R$$

$$\text{Average values } \frac{E_{av}}{I_{av}} = \text{R}$$

If voltage and current do not have the same type of values, they must first be converted to the same type before applying Ohm's law.

EXAMPLE: Find the value of the rms current flowing through a resistance of 10 Ω when an AC voltage having peak value of 100 V is applied across it.

Since rms current is required, the peak voltage must be converted to rms voltage by the formula as

$$E_{\text{rms}} = 0.707 \times E_{\text{pk}};\ E_{\text{rms}} \text{ for 100 V peak is } 100 \times 0.707 = 70.7 \text{ V}$$

$$I_{\text{rms}} = \frac{70.7}{10} = 7.07\text{A}$$

If the circuit is not purely resistive but contains inductive and capacitive components also, the voltages and currents are not in phase and the application of Ohm's law involves the use of vectors.

3.11.2 Power in AC Circuits

The power P dissipated in a DC circuit is given by the product of the voltage E and the current I or

$$P = E \times I$$

This formula holds true for AC circuits provided the voltage and current are in phase which will happen in a resistive circuit. However, the value of power will be instantaneous, effective, or peak depending on the type of value used for voltage and current. Thus

$$P_{\text{inst}} = E_{\text{inst}} \times I_{\text{inst}}$$
$$P_{\text{eff}} = E_{\text{eff}} \times I_{\text{eff}}$$
$$P_{\text{pk}} = E_{\text{pk}} \times I_{\text{pk}}$$

If the AC circuit contains certain reactive components like coils and capacitors, the voltage and the current are not in phase. There are times during the cycle when the voltage is negative and the current is positive and vice versa. During an interval when the voltage and current have opposite signs power is fed back into the source, meaning thereby that there is less power consumed in the external circuit than is indicated by the apparent power. The actual power consumed in the external circuit is called the *real power or true power* and is the product of voltage and current when these are in phase. The apparent power is the power supplied by the source, which is less than the true power because a part of this power is returned to the source.

3.12 POWER FACTOR

The ratio of the *true power* to the *apparent power* is called the power factor.

$$\text{Power factor} = \frac{\text{True power}}{\text{Apparent power}}$$

This is always less than 1 in circuits which contain reactive elements and the voltage and current are not in phase. Power factor is expressed as a factor such as 1/2, or as a percentage such as 50%. In a purely resistive circuit, the voltage and current are in phase and apparent power is equal to the true power. The power factor is equal to unity or 100% which is the maximum value it can attain. We can also write the above formula as:

$$\text{True power} = \text{Apparent power} \times \text{Power factor}$$

If the phase difference between the voltage and current is θ, the power factor is $\cos\theta$ and the above formula is also sometimes written as

$$\text{True power} = \text{Apparent power} \times \cos\theta$$

In purely reactive (no resistance) circuits phase difference between voltage and current is 90° and $\cos 90° = 0$ or the power factor is zero.

Such circuits do not consume any power and are *wattless* circuits. In many industrial concerns the use of electric motors considerably lowers the power factor resulting in overloading of the generators due to the fact that the AC power consumed is much less than the power supplied by the generators. To

avoid this, improvement in the power factor is effected by connecting a suitable capacitor across the power line. This is called correcting the power factor.

Summary

A current which always flows in the same direction without reversing its direction of flow is called direct current or DC but a current that reverses direction at regular intervals is called alternating current or AC. Alternating current has many advantages over direct current and the bulk of electric power consumed in the world is AC power. The main source of AC voltage is an AC generator or an alternator which produces by its rotation a waveform called a sine wave. In a sine wave, the value of the voltage or current changes from instant to instant and the maximum instantaneous value reached during a cycle is called the amplitude or peak value. Other values like average value and effective value can be derived from the peak value by known formulas.

One complete rotation of the loop producing a sine wave is called a cycle and the number of cycles in one second is the frequency of the sine wave. The distance travelled by a wave during one cycle is its wavelength. A voltage source is a device for producing electricity by converting some other form of energy into electrical energy. One important voltage source is a cell which converts chemical energy into electrical energy. A primary cell called a voltage cell converts chemical energy directly into electrical energy and cannot be charged again. A secondary cell has to be charged first with electrical energy which is stored in the cell as chemical energy and then released as electrical current doing discharge by a reversible chemical action. A secondary cell is also called a storage cell or accumulator. A battery is a combination of cells so connected as to increase the voltage, or current, or both, beyond what is obtainable from a single cell. A dry battery is used when high voltage and low current are required and an accumulator is used for low voltage but high current rating. A storage battery can be charged and recharged any number of times by a battery charger.

One of the most important voltage sources is the electromagnetic voltage source called a generator which is the biggest source of electrical power produced in the world. A generator converts mechanical energy into electrical energy with the help of magnetism. Generators are either DC generators called dynamos or AC generators also known as alternators.

Other voltage sources are thermoelectricity, photoelectricity and piezoelectricity. All these sources produce extremely low voltages and require amplification by other electronic means.

Review Questions

1. What is a battery? How many dry cells are required to construct a 45 V dry battery having three times the current rating of a single dry cell? (*Ans.* 90)
2. How many secondary cells are used in a 12 V storage battery having a current rating of 120 A. How long would the above battery take to be fully charged if the charging rate is 5 A. (*Ans.* 6, 24)

3. What is a generator? What is the difference between a thermal generator and diesel generator?
4. What is a dynamo? Describe briefly its construction. Give an example of a common type of dynamo.
5. What is the difference between thermoelectricty and piezoelectricity? Give examples of the practical use of each of these types of electricity.
6. What is the difference between DC and AC? Give their relative advantages.
7. What are the characteristics of a sine wave? Indicate which of the following are sine waves:
   1. Sound waves 2. Square waves 3. Sawtooth waves 4. Fluctuating DC 5. Light waves.
8. Explain what is meant by peak, E_{rms} and average value of an AC voltage? How are these related to each other? Peak-to-peak value of an AC voltage is 100 V. Calculate the effective value and average value of the voltage.
   (*Ans.* E_{rms} = 35.35 V, Average = 31.85 V)
9. What is the peak value of the current flowing through a 100 W electric bulb connected across a 230 V 50 Hz power line. (*Ans.* 0.615 A)
10. What is the wavelength of a broadcast station radiating at 1370 kHz? How is this frequency useful for regional and local broadcasts?
    (*Ans.* 219 m)
11. Define vector and scalar quantities. Which of the following are vector quantities?
    1. Acceleration 2. Kilograms 3. Dynes 4. Seconds.
    Find the resultant of two voltages of magnitudes 8 V and 6 V having a phase difference of 90°. (*Ans.* 10 V)
12. Define true power, apparent power and power factor. What is the power factor in the following cases?
    1. When the voltage and current are in phase.
    2. When the voltage leads the current by 60°.
    3. When the voltage leads the current by 90°.
    (*Ans.* (1) 1, (2) 0.5, (3) 0)
13. Indicate by means of sine wave curves the relation between an AC voltage leading its current by 90°. Show the same relationship by a vectorial representation.
14. Correct the following statements with full justification:
    1. Sound waves and radio waves are both electromagnetic waves and travel with the same velocity.
    2. Power factor can be made more than 1 by connecting a suitable capacitor across the power line.
    3. The higher the frequency of a sine wave the longer is its wavelength.
15. What is average power dissipation in an AC circuit where the peak voltage is 100 V and the peak current in 20 A? (*Ans.* P_{AV} = 811.5 W)

chapter 4

Passive Electronic Components

I. RESISTORS

4.1 RESISTANCE

Resistance has been defined as the opposition to the flow of current. This opposition comes from the electrons present in the atom of a material. In materials like copper, the electrons are more free to move about than the electrons in a material like rubber. In other words copper has less resistance than rubber. Resistance is denoted by the symbol *R* and the unit of resistance is *ohm*.

4.2 FACTORS DETERMINING THE RESISTANCE OF A MATERIAL

The resistance of a material depends on a number of factors. These factors are (i) length, (ii) area of cross-section, (iii) nature of the material, and (iv) temperature of the material.

(i) Length The greater the length of a material, the greater the resistance offered by it to the flow of current. Thus, if a copper wire of length *l* has a resistance *R*, the resistance of length 3*l* of the same copper wire will be 3*R* as shown in Fig. 4.1. Resistance is directly proportional to the length *l* of the material.

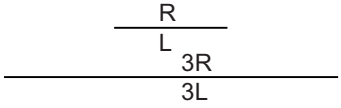


Fig. 4.1 The Greater the Length, the Greater the Resistance

(ii) Area of Cross-section A thicker material has greater area of cross-section than a thinner material. The thicker a material the more plentiful is the supply of electrons in it. Hence, it allows more current to flow and offers less resistance than a thinner material with fewer electrons and more resistance.

Resistance is inversely proportional to the area of cross-section A of the material.

The area of cross-section is generally expressed in square centimetres or circular mils. A mil is equal to 0.001 inch or one thousandth of an inch. One circular mil or 1 c mil is the cross-sectional area of a conductor with a diameter of 1 mil. A round conductor with a diameter of 0.002 inch will have a diameter of 2 mil and its area in c mil will be $(2)^2$ or 4 c mil. A mil is a unit of length whereas a circular mil is a unit of area. A c mil is quite convenient for expressing the area of cross-section of round wires and is often used in Standard Wire Gauge tables where the diameter and area of cross-section of different gauges of wire are given in mils and c mils respectively. The higher the gauge number of a wire, the smaller is its cross-sectional area and greater the resistance of a wire, the smaller is its cross-sectional area and greater the resistance of the wire for a given length.

(iii) Nature of Material *Specific resistance:* Resistance also depends on the nature of a material. A copper wire will not have the same resistance as an iron wire of equal dimensions. The nature of a material is indicated by a constant ρ (rho) known as the *specific resistance* or *resistivity* of the material. Thus, the resistance R of a material l cm long and having an area of cross-section A is expressed by the formula:

$$R = \rho \frac{l}{A}$$

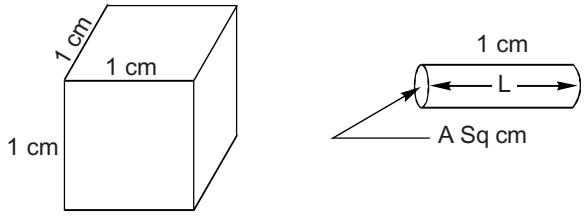


Fig. 4.2 One Cubic Centimetre

where ρ is the specific resistance of the material. In the above formula if l = 1 cm and A = 1 cm^2 then $R = \rho\ 1/1 = \rho$. This gives us a definition of specific resistance ρ. Specific resistance is, in effect, the resistance of 1 cm^3 of a material at a known temperature.

Specific resistance or resistivity of some important materials used in radio and electronic work is given in Table 4.1

The specific resistance of metals is also expressed in terms of standard wire size with length equal to 1 ft with a cross-sectional area of 1 c mil. This specific resistance is expressed as ohms per circular mil ft as indicated in Table 4.1.

Table 4.1 Specific Resistance of Some Important Materials

Material	*Resistivity at 20°C*	
	Ohms per C mil ft	*Micro-ohms per cm³*
Aluminium	17	2.83
Brass	45	7.5
Carbon (graphite)	400-1100	33-18
Copper	10.37	1.72
Iron	59	9.8
Molybdenum	34	5.7
Nickel	60	10
Phosphor bronze	70	11.5
Silver	9.8	1.63
Tungsten	33	5.51
Nierhrome	650	108
Manganin	290	48
German silver	185	31

Metals like silver, copper and aluminium have less value of specific resistance and are used as conductors for various types of wires and cables. Tungsten and iron have high specific resistance and are used as resistance wires. Certain alloys like nichrome and manganin also have high specific resistance values and are used as elements for heaters, toasters, electric irons and also for winding high value resistances.

EXAMPLE: Calculate the resistance of 50 ft of copper wire having a cross-sectional area of 509.5 c mil.

$$R = \rho \frac{1}{A}$$

$\rho = 10.37\Omega$ per c mil/ft; $l = 50$ft

$A = 509.5$ c mil

$$R = 10.37\ \Omega\ \frac{\text{c mil}}{\text{ft}}\ \frac{50\text{ft}}{509.5\text{ c mil}} = \frac{10.37 \times 50}{509.5}$$

$= 1\Omega$ (approx.)

(iv) Temperature The resistance of a conductor also depends on the temperature of the conductor. This dependence of resistance on temperature is indicated by α (alpha) which is known as the temperature coefficient of resistance. It indicates how much the resistance changes with the temperature.

α can have positive, negative or zero value. When α is positive, the resistance of the material increases with temperature. Most metals like copper, silver, tungsten have positive α. It therefore, becomes necessary to state the temperature at which the material will have the specified value of resistance.

A negative value of α means that resistance will decrease with an increase of temperature. Materials with negative α will have lower resistance at higher temperatures and higher resistance at lower temperatures. Some conductors like carbon, germanium and silicon have negative α. This has important practical

applications in the form of thermistors which are carbon resistors with a negative temperature coefficient of resistance. These are used to compensate for any changes in the resistance of wire conductors that take place with an increase of temperature.

Certain alloys of metals like manganin and constantan have zero α which means that the resistance does not change with temperature. These are used in precision wire wound resistors whose resistance does not change with a change of temperature.

4.3 RESISTORS

A resistor is an electronic component which has a known value of resistance. A resistor is specially designed to introduce a desired amount of resistance in a circuit. A resistor is used either to control the flow of current or to produce a desired voltage drop. Resistors are, perhaps, the most common components used in electronic and electric circuits.

4.4 TYPES OF RESISTORS

Resistors are of two types–carbon resistors and wire-wound resistors.

Carbon Resistors These can be either carbon composition resistors or carbon film resistors.

(a) Carbon Composition Resistors are made by mixing granules of carbon with a binding material and moulded in the form of rods as shown in Fig. 4.3. Wire leads are inserted at the two ends and the package is sealed with a non-conducting coating.

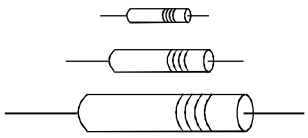


Fig. 4.3 Carbon Resistors

(b) Carbon Film Resistors are types of carbon resistors made by depositing a carbon film on a ceramic rod. The value of the resistance is set by cutting a spiral groove through the film. The groove adjusts the length and the width of the ribbon so that the desired value of the resistance is reached.

Carbon resistors are available in values varying from 1Ω to 20 MΩ.

Wire-wound Resistors These are made by wrapping a known length of wire of a nickel-chrome alloy called nichrome over a form made of ceramic (Fig. 4.4). Advance and manganin wires are also used for the construction of wire-wound resistors. All these materials have a much higher resistivity than copper. After taking out leads the entire winding is coated with a protective coating. Since the length of wire required increases with the ohmic value of the resistance, the range of these resistors is from less than 1 Ω to several thousand ohms. These resistors are used in circuits carrying high

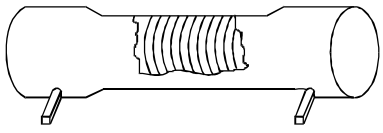


Fig. 4.4 Wire-wound Resistor

current when relatively high amounts of power are dissipated. Wire-wound resistors are used in precision instruments where stable and accurate resistance values are required.

4.5 RESISTOR RATINGS

There are three factors that determine the rating of a resistor. These are (i) resistance value, (ii) tolerance, and (iii) power rating or wattage.

(i) Resistance Value This is the value of the resistance expressed in ohms, e.g. 10 Ω, 1 kΩ or 10 MΩ. This resistance value is either written or stamped on the body of the resistor as in the case of wire-wound resistors. In the case of carbon resistors (film), the value of the resistance is indicated by a colour code which is described below.

(ii) Tolerance This is the variation in the value of the resistance that is expected from the exact value indicated. 10% tolerance in the case of a 1000 Ω resistor will mean that its value can vary from 900 Ω to 1100 Ω. Tolerances of 5%, 10% and 20% are common for carbon resistors. Wire-wound resistors are more exact in value and their tolerances are low.

(iii) Wattage This is the maximum amount of heat in watts that can be dissipated by a resistor without damage to it. The physical size of a resistor gives an indication of its wattage. The larger the size of a resistor, the higher will be its wattage rating.

Carbon resistors generally have a low wattage of value $\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1$ or 2W. Wire-wound resistors are available in higher wattage ratings from 5W to several hundred watts.

Wattage is a very important rating of a resistor for use in a particular circuit. Low wattage resistors can be used in transistor circuits but comparatively higher wattage ratings are required for valve or tube circuits.

4.6 COLOUR CODE FOR CARBON RESISTORS

The ohmic value and the tolerance of a carbon resistor is indicated by a colour code. Ten colours are used to represent each of the digits from 0 through 9 as indicated below in Table 4.2.

Table 4.2

Colour	*Digit*	*Colour*	*Digit*
Black	0	Green	5
Brown	1	Blue	6
Red	2	Violet	7
Orange	3	Gray	8
Yellow	4	White	9

In the case of carbon resistors having axial leads, the colour bands are printed near one edge of the resistor as in Fig. 4.5(a). Reading from left to right, the first colour band indicates the first significant figure; the second band the second significant figure and the third band indicates the multiplier (×10) or the number of zeroes to be added after the two significant figures. The tolerance is indicated by the fourth band or by its absence. A gold band indicates 5% tolerance, silver band 10% tolerance and no band 20% tolerance. Thus a resistor of 47,000 Ω with a tolerance of ± 5% will have the colour code shown in Fig. 4.5(b).

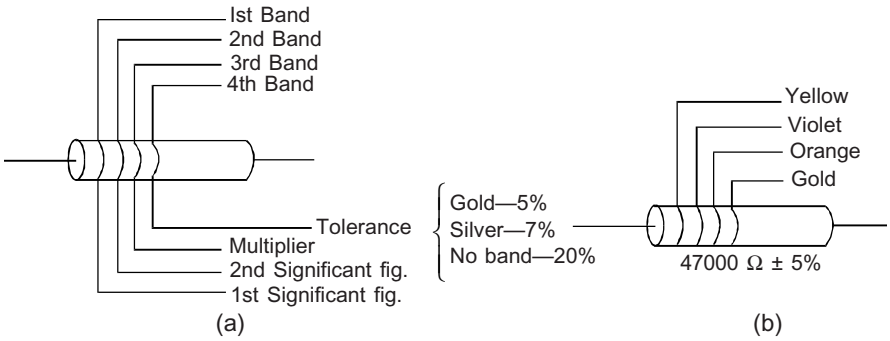


Fig. 4.5 (a) Colour Code for Carbon Resistors (b) 47000 Ω + 5%

EXAMPLE: What is the value and tolerance of a carbon resistor which has the following bands starting from the left: blue, black, green and silver.

1st band is *blue*, so *first* significant figure is 6

2nd band is *black*, so *second* significant figure is 0

3rd band is *green*, so *number* of zeroes is 5

4th band is *silver,* so tolerance is ± 10%

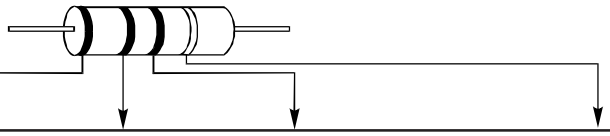
Therefore, the resistor value is

$$6000000 \pm 10\%$$

which is $5M\ \Omega \pm 10\%$

For resistance values under 10, use is made of the fractional multipliers 0.1 (gold) and 0.01 (silver). When the third band is gold, the resistance value indicated by the first two bands is multiplied by 0.1. Thus, if the first and second bands are green and violet respectively and the third band is gold, the resistance value is $56 \times 0.1 = 5.6\ \Omega$. If the third band is sliver, the resistance value would be $56 \times 0.01 = 0.56\ \Omega$. A complete colour code scheme for carbon resistors is given in the Table 4.3. Carbon as well as wire-wound resistors can be either fixed resistors or variable resistors.

Table 4.3 Colour Code Chart for Resistor



Colour	*1st Band 1st Significant fig*	*2nd Band 2nd Significant fig*	*3rd Band Multiplier*	*No. of zeros to be added*	*4th Band Tolerance*
Black	—	0	XI	None	
Brown	1	1	X10	0	
Red	2	2	X100	00	
Orange	3	3	X 1000	000	
Yellow	4	4	X 10000	0000	
Green	5	5	X 100000	00000	
Blue	6	6	X 1000000	000000	
Violet	7	7	X 10000000	0000000	
Grey	8	8	X 100000000	00000000	
White	9	9	X 1000000000	000000000	
Gold	—	—	X 0.1	—	± 5%
Silver	—	—	X 0.01	—	± 10%
No band	—	—	—	—	± 20%

4.7 VARIABLE RESISTORS

The most common type of variable resistor is known as the potentiometer or simply POT as shown in Fig. 4.6(a). This is the type used as a volume control in radio and TV sets. The resistance value can be varied by the rotation of a shaft fixed at the end. A volume control is generally combined with an ON-OFF switch for power supply which can be operated from the common shaft. Figure 4.6(a) shows the construction of a potentiometer. In this potentiometer a

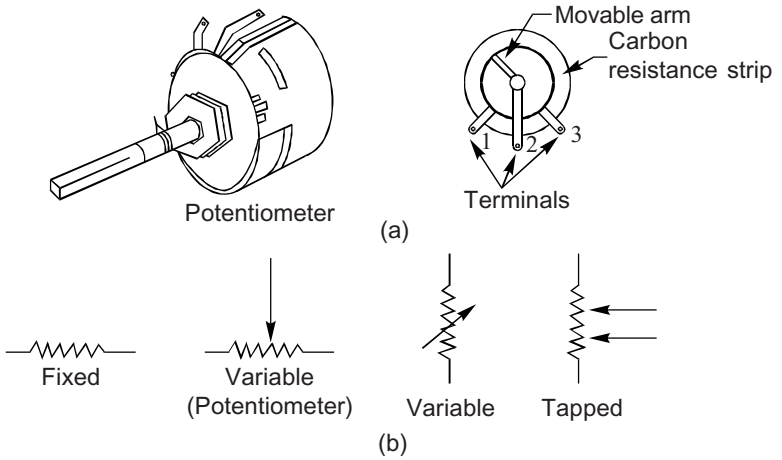


Fig. 4.6 Variable Resistors (a) Potentiometer (b) Symbols for Variable Resistors

movable contact slides over a strip of carbon. The resistance between terminals 1 and 3 is fixed and is the maximum value marked on the body of the potentiometer.

When the movable arm slides from left to right or clockwise the resistance between points 1 and 2 increases and that between points 2 and 3 decreases, see Fig. 4.6(a). The variation of resistance is not uniform in the case of carbon potentiometers. For uniform variation of resistance and for precision work, wire-wound potentiometers are used. These have wire-wound elements in place of carbon strips.

Figure 4.7 shows the use of a potentiometer as a voltage divider. The voltage to be divided is applied between points 1 and 3 and the variable voltage will appear between 3 and 2. Point 3 becomes common to input and output voltages. A potentiometer can be used as a Rheostat when the purpose is to vary only the resistance value in the circuit. In this case only the end terminals, i.e. 1 and 2 or 2 and 3 are used. Carbon preset potentiometers are trimming potentiometers for preset resistance controls with provision for further adjustment. These are used for radio and television appliances.

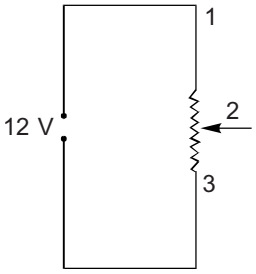


Fig. 4.7 Potentiometer as a Voltage Divider

When continuous variation of the resistance is not required, a tapped resistance is used. Fixed taps are provided at a number of points on a wire-wound resistance. A sliding contact on a bare wire resistance will vary the resistance in any of the convenient steps.

Symbols used for various types of resistors are shown in Fig. 4.6(b).

4.8 RESISTORS IN COMBINATIONS

Resistors can be connected in combinations in a number of ways to obtain some desired effective value of resistance and wattage. Some of the important methods of combining two or more resistors are (i) series combination, (ii) parallel combination, and (iii) series-parallel combination. The effective resistance of the combination in any particular case can be found out by the application of Ohm's law.

4.8.1 Series Combination

Resistors are said to be connected in series when they are connected end-to-end in successive order as shown in Fig. 4.8(a). When such a series combination is connected to a battery with voltage E, Fig. 4.8(b), there is only one path for the electrons to flow from the negative to the positive pole of the battery and this is through each of the resistors. The same number of electrons flow through each of the resistors and hence the current flowing through each resistor is the same. The voltage drop will, however, depend on the value of each resistor and the amount of current flow.

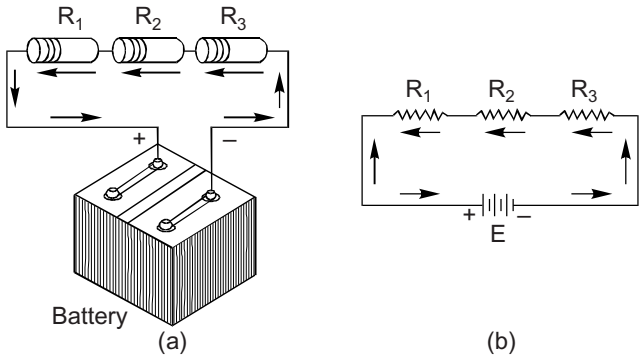


Fig. 4.8 Resistor in Series (a) Actual (b) Schematic

If E_1, E_2 and E_3 are the voltage drops across R_1, R_2 and R_3 respectively, then by Ohm's law:

$$E_1 = IR_1,\ E_2 = IR_2,\ E_3 = IR_3$$

The total voltage drop is equal to voltage E of the battery. In other words,

$$\begin{aligned} E = E_1 + E_2 + E_3 &= IR_1 + IR_2 + IR_3 \\ &= I\,(R_1 + R_2 + R_3) \end{aligned} \tag{4.1}$$

If R_T is the effective value of the resistors R_1, R_2, and R_3 in series, we have

$$E = IR_T \tag{4.2}$$

Combining equations (4.1) and (4.2)

$$R_T = R_1 + R_2 + R_3$$

The total effective value of resistors in series is equal to the sum of the individual resistance values. Thus, three resistance of value 5Ω each, connected in-series in a circuit can be replaced by a resistance of 15Ω.

The order in which the resistors are connected in series arrangement does not alter the current through the circuit.

EXAMPLE: Three resistors of value 5 Ω, 10 Ω and 15 Ω are connected in series across a 6 V battery. Find the current flowing through the circuit and voltage drop across each resistors.

$$R_T = 5 + 10 + 15 = 30\Omega$$

$$I = \frac{E}{R_T} = \frac{6V}{30\Omega} = 0.2A$$

The voltage drop across 5 Ω is $IR = 5 \times 0.2$ V = 1 V. Similarly the voltage drop across 10 Ω is $10 \times 0.2 = 2$ V and that across 15 Ω is $15 \times 0.2 = 3$ V.

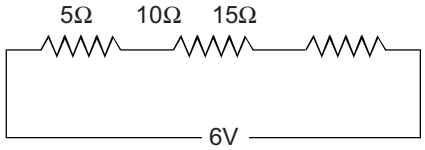


Fig. 4.9 Series Arrangement of Heaters in an AC/DC Radio Receiver

4.8.2 Parallel Combination of Resistors

Two or more resistors are said to be connected in parallel when one end of each resistor is connected to the positive pole and the other end to the negative pole of the common battery or voltage source. Figure 4.10(a) shows the actual connection and Fig. 4.10(b) is the schematic diagram of a parallel connection of resistors. The main electron stream leaving the negative terminal of the battery divides itself into three parallel streams of electrons or current flowing through each of the three resistances. These are called the branch currents, the value of each current depending on the resistance in the branch. The higher the resistance in the branch, the smaller the value of the branch current. All these branch currents again add up to form the main stream of electrons or the total current that returns to the positive pole of the battery and back to the negative pole through the battery itself. This cycle continues so long as the voltage is applied to the combination. In other words, the total current I_T flowing through the circuit is the sum of the branch current I_1, I_2, l_3.

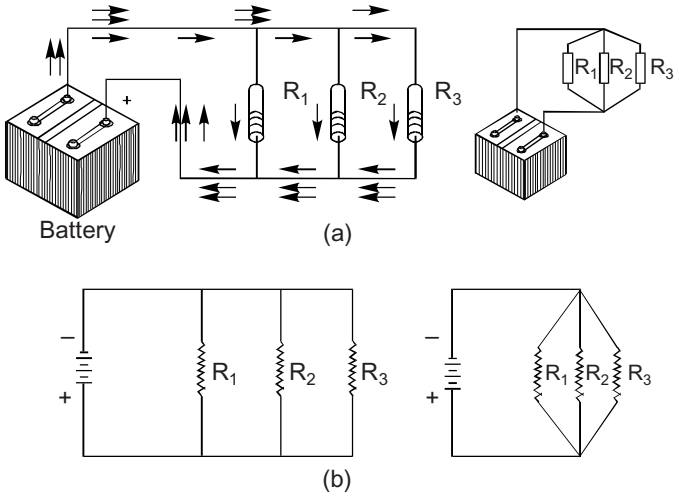


Fig. 4.10 Resistors in Parallel

$$I_T = I_1 + I_2 + I_3 \tag{4.3}$$

Applying Ohm's law we have:

$I_T = \frac{E}{R_T}$ where R_T is the total or effective resistance of the parallel combination

also $$I_1 = \frac{E}{R_T}, I_2 = \frac{F}{R_2} \text{ and } I_3 = \frac{E}{R_3}$$

Substituting in Eq. 4.3

$$\frac{E}{R_T} = \frac{E}{R_1} + \frac{E}{R_2} + \frac{E}{R_3}$$

or $$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \quad (4.4)$$

In other words this only means that in a parallel combination of resistors the reciprocal of the total effective resistance ($1/R_T$) in the circuit is equal to the sum of the reciprocals ($1/R_1 + 1/R_2 + 1/R_3$) of individual resistances in parallel.

EXAMPLE: A parallel combination of three resistors of value 10 Ω, 15 Ω and 30 Ω is connected across a 30 V battery as in Fig. 4.11. Find the (i) total effective resistance of the combination, (ii) total current flowing in the circuit, and (iii) current flowing through each branch of the circuit.

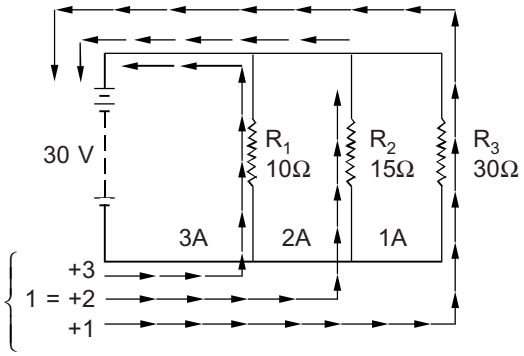


Fig. 4.11 Example 3 Resistors in Parallel

(i) $$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

$$\frac{I}{R_T} = \frac{1}{10} + \frac{1}{15} + \frac{1}{30} = \frac{6}{30}$$

$$R_T = 5\Omega$$

(ii) Total current $I_T = \dfrac{30\ V}{R_T}$

$$= \frac{30}{5\Omega} = 6A$$

(iii) Current in R_1 branch $I_2 = \dfrac{30\ V}{10\Omega} = 3\text{A}$

Current in R_2 branch $I_2 = \dfrac{30\ V}{15\Omega} = 2\text{A}$

Current in R_3 branch $I_3 = \dfrac{30\ V}{30\Omega} = 1\text{A}$

$I_T = I_1 + I_2 + I_3 = 3\text{A} + 2\text{A} + 1\text{A} = 6\text{A}$

Case of two resistors in parallel

(i) When there are only two unequal resistors in parallel, Eq. 4.4 becomes

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_2 + R_1}{R_1 \times R_2}$$

or
$$R_T = \frac{R_1 \times R_2}{R_1 + R_2} \tag{4.5}$$

(ii) When the two resistors are equal, Eq. 4.5 becomes

$$R_T = \frac{R \cdot R}{R + R} = \frac{R \cdot R}{2R} = \frac{R}{2} \tag{4.6}$$

The total resistance of two equal resistors in parallel is equal to one-half the resistance of each.

4.9 CONDUCTANCE

In parallel circuits we have to deal with reciprocals of resistances. The reciprocal of a resistance is called *conductance* and is denoted by the symbol G. Thus,

$$G = \frac{1}{R}$$

The unit of conductance is mho (℧). This unit is also called Siemen.
Equation 4.4 in terms of conductance becomes

$$G_T = G_1 + G_2 + G_3 + ...$$

Working with G is more convenient than working with R in parallel circuits. This avoids the use of reciprocals.

For three resistors of value 5 Ω, 10 Ω and 20 Ω connected in parallel as in Fig. 4.12, the values of G_1, G_2 and G_3 work out to be 0.2℧, 0.1℧ and 0.05℧, respectively. Current is proportional to conductance G for any applied voltage say 20 V,

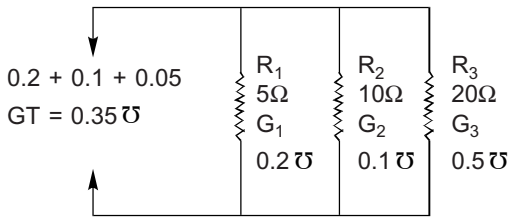


Fig. 4.12 Solution of Parallel Circuits by Using G in place of R

$$I_T = E \times G = 20 \times 0.35 = 7\text{A}$$
$$I_1 = E \times G_1 = 20 \times 0.2 = 4\text{A}$$
$$I_2 = E \times G_2 = 20 \times 0.1 = 2\text{A}$$
$$I_3 = E \times G_3 = 20\text{V} \times 0.05\Omega = 1\text{A}$$
$$I_1 + I_2 + I_3 = 7\text{A}$$

Common practical applications of the parallel circuit are the various household electrical appliances like electric lights, electric fans, electric irons etc. All the appliances are connected in parallel with the mains supply as shown in Fig. 4.13. Each of these appliances draws its own current from the mains depending on its resistance, but the total current is supplied by the common mains supply line. If due to any reason, the total current exceeds a certain predetermined value a *fuse* in the main supply is blown off and the mains line is permanently disconnected to avoid any damage to the building or to any of the electrical appliances. A fuse is only a thin wire made of either aluminium, tin coated copper or tin wire which melts as soon as the current exceeds the carrying capacity of the fuse wire. Fuses of various ratings starting from a fraction of a milliampere to hundreds of amperes are available. Sometimes fuses are also introduced in individual parallel circuits, in addition to the main fuse, so that any defect in one circuit will blow off its individual fuse and will not affect the working of other circuits connected in-parallel.

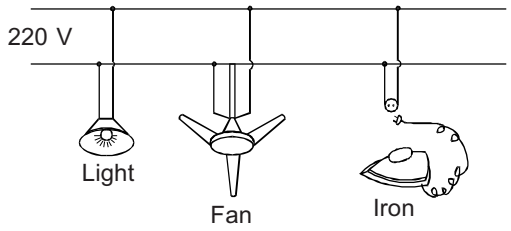


Fig. 4.13 Household Electric Appliances as Parallel Circuits

Principle of Shunts Two resistors in parallel are also said to be shunt-connected. The smaller resistor is said to be a shunt across the bigger resistor. Thus the principle of a shunt is useful in the construction of an ammeter with different ranges. By putting a shunt of suitable value in parallel with the internal resistance of the meter, only the current required for full-scale deflection is allowed to pass through the meter and the rest of the current passes through the shunt. The scale is calibrated in terms of the total current and not the current passing through the meter.

4.10 SERIES-PARALLEL COMBINATION OF RESISTORS

A series-parallel combination of resistors contains a combination of series connected and parallel connected resistances as shown in Fig. 4.14. To find the total effective resistance R_T of the circuit, the two parallel resistors R_2 and R_3 are reduced to an equivalent resistor by use of the formula $R = R_2 \times R_3/R_2 + R_3$. This equivalent resistance R is added to the series resistor R_1 to get the total effective resistance in the circuit. The current can be found by the application of Ohm's law.

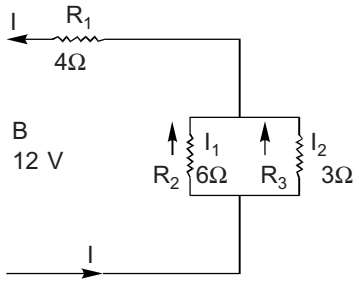


Fig. 4.14 Series-parallel Combination of Resistors

In the above case:

$$R = \frac{R_2 \times R_3}{R_2 + R_3} = \frac{6 \times 3}{6 + 3} = \frac{18}{9} = 2\Omega$$

$$R_T = R_1 + R = 4 + 2 = 6\Omega$$

Current $$I_T = \frac{E}{R_T} = \frac{12\ V}{6\Omega} = 2\text{A}$$

A slightly more complicated series-parallel arrangement is shown in Fig. 4.15.

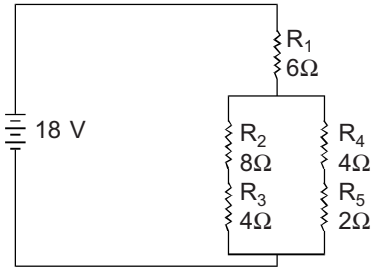


Fig. 4.15 Another Series-parallel Arrangement

In this case the equivalent resistance of $R_2 + R_3$ in series is 8 + 4 = 12 Ω and R_4 and R_5 in-series is 4 + 2 = 6 Ω. These two resistances of 6 Ω and 12 Ω are in parallel and are equivalent to a resistance of 6 × 12/6 + 12 = 72/18 = 4 Ω. This equivalent resistance of 4Ω is in series with R_1 which is 6 Ω.
The total effective resistance of the series parallel combination is

$$6\ \Omega + 4\ \Omega = 10\ \Omega$$

$$\text{Current } I_T = \frac{18\ V}{10\ \Omega} = 1.8\ A$$

A circuit of the above type is also called a *network*.

Wattage of Resistors in Combination The total resistance of a combination of resistances depends on whether the resistances are connected in series or in parallel but the total wattage of a combination of resistances is always equal to the sum of the individual wattages of the resistors irrespective of the fact whether the combination is a series combination or a parallel combination

or a series-parallel combination. This is because the total physical size of the combination increases with the addition of each resistor. Two resistors of 100 Ω 1 W each will have a total resistance R_T of 200 Ω and a total wattage of 2 W when connected in-series as in Fig. 4.17(a). The same resistors will have R_T of 50 Ω when connected in parallel as in Fig. 4.16(b) but the total wattage again will be 2 W. This fact can be used to obtain a desired value of resistance with a higher wattage rating by combining resistors in series or in parallel combinations.

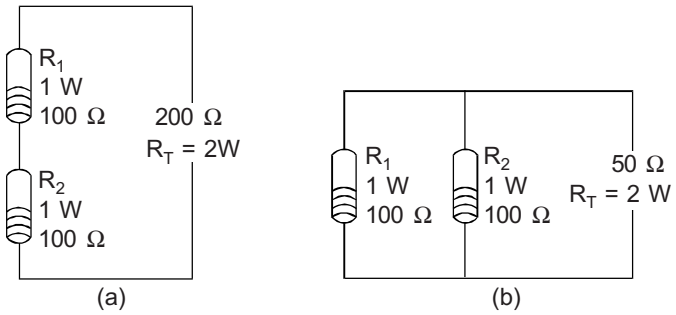


Fig. 4.16 Total Wattage of a Combination of Resistors

4.11 KIRCHHOFF'S LAWS

Ohm's law is applicable to relatively simple circuits, containing a single voltage source and a few resistances arranged in series, parallel, or series-parallel arrangement. In the case of complex circuits containing more than one voltage source and the components connected at random in many branches, it becomes difficult to analyse such circuits by the application of Ohm's law alone. The laws that help solve such complex circuits are known as Kirchhoff's laws after the German scientist Gustav Kirchhoff. These laws which are two in number are described below.

Kirchhoff's Current Law This law states that the sum of currents entering a point in a circuit is equal to the sum of currents leaving that point. This law follows from the fact that charge cannot accumulate at any point in a conducting path. The number of electrons entering is equal to the number of electrons leaving the point. If the direction of current entering point P in Fig. 4.17(a) is taken as positive, the sign of currents leaving the point will be negative. Thus

$$I_T = I_1 + I_2 \quad \text{or} \quad I_T - I_1 - I_2 = 0$$

In the case of Fig. 4.17(b)

$$I_1 + I_2 = I_T \quad \text{or} \quad I_1 + I_2 - I_T = 0$$

If the algebraic signs of the currents are kept in view, Kirchhoff's Current Law can also be stated as follows:

The algebraic sum of the currents entering and leaving any point in a circuit is equal to zero.

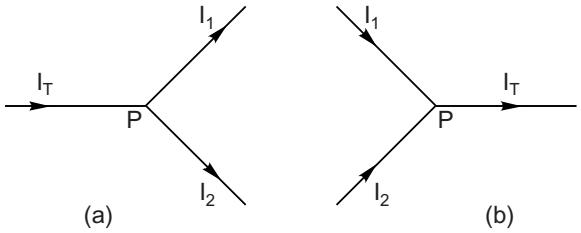


Fig. 4.17 Kirchhoff's Current Law

Kirchhoff's Voltage Law The total sum of *IR* voltage drops in a closed loop is equal to the sum of the voltage sources in the loop.

A loop is a closed path around which current can flow from any starting point back to the same point.

In finding the sum of the voltage drops round a loop, due regard must be paid to the signs of the current and voltage drops in the loop. If the current or the voltage acting in the clockwise direction is taken as positive, the current or voltage drop acting in the opposite direction must be taken as negative. In determining the polarity of the voltage *IR* drop in a resistor, it may be remembered that the point where the electron current enters gets the negative polarity.

With the signs of the various voltages taken into account, Kirchhoff's Voltage Law can also be stated as follows:

The algebraic sum of the *IR* voltage drops and voltage sources must be equal to zero around a closed loop.

The closed loop shown in Fig. 4.18 has two voltage sources V_1 and V_2 and three *IR* drops IR_1, IR_2 and IR_3 with their polarities marked in the figure.

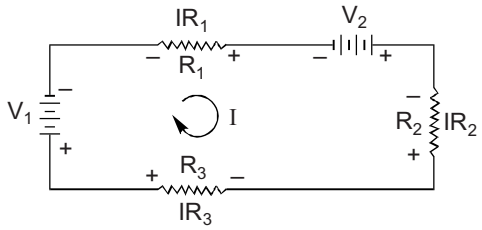


Fig. 4.18 Kirchhoff's Voltage Law

According to Kirchhoff's Voltage Law

$$V_1 - IR_1 - V_2 - IR_2 - IR_3 = 0$$

This equation can also be written as

$$V_1 - V_2 = IR_1 + IR_2 + IR_3 \tag{4.7}$$

From this equation we have to find the value of I if we know the values of V_1, V_2, R_1, R_2 and R_3.

Thus if V_1 = 12 V, V_2 = 6 V and $R_1 = 2\Omega$, $R_2 = 4\Omega$, $R_3 = 6\Omega$ then substituting these values in (4.7) we get

$$12\text{V} - 6\text{V} = I \times 2 + I \times 4\ I \times 6$$

or $$6\text{V} = 12I$$

or $$I = \frac{6}{12} = 0.5\ \text{A}$$

Kirchhoff's Voltage Law can also be expressed by the simple equation.

$$\Sigma\ V = \Sigma\ IR \tag{4.8}$$

where the symbol Σ means 'the sum of'. Thus

$\Sigma\ V = V_1 + V_2 + V_3 + ...$ (algebraic sum of all voltage sources)

and $\Sigma\ IR = I_1R_1 + I_2R_2 + I_3R_3...$ (algebraic sum of all IR drops)

Application of Kirchhoff's Laws In solving complex circuits, it often becomes necessary to apply both Kirchhoff's laws simultaneously as will be clear from the following example:

EXAMPLE: Find the current I_1, I_2 and I_3 flowing through the resistors R_1, R_2 and R_3 respectively in Fig. 4.19.

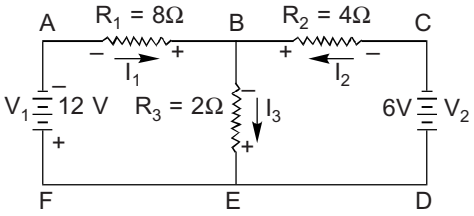


Fig. 4.19 Example

Applying Kirchhoff's Current Law to current entering and leaving point B,

$$I_3 = I_1 + I_2 \tag{4.9}$$

Applying Kirchhoff's Voltage Law to the loop ABEFA,

$$V_1 = I_1R_1 + I_2R_3 \tag{4.10}$$

Similarly, applying Kirchhoff's voltage law to the loop BCDEB

$$V_2 = I_2R_2 + I_3R_3 \tag{4.11}$$

Substituting Eq. (4.9) in (4.10) and (4.11), we have $V_1 = I_1R_1 + (I_1 + I_2)R_3$

$$V_2 = I_2R_2 + (I_1 + I_2)R_3$$

Simplifying

$$V_1 = (R_1 + R_3)I_1 + I_2R_3 \tag{4.12}$$

$$V_2 = (R_2 + R_3)I_2 + I_1R_3 \tag{4.13}$$

Substituting the values of V_1, V_2, R_1, R_2 and R_3 in equations (4.12) and (4.13)

$$12 = 10I_1 + 2I_2 \tag{4.14}$$

$$6 = 6I_2 + 2I_1 \tag{4.15}$$

Further simplifying these equations

$$6 = 5I_1 + I_2 \tag{4.16}$$

$$3 = 3I_2 + I_1 \tag{4.17}$$

Multiplying equation (4.16) by 3

$$18 = 15I_1 + 3I_2 \tag{4.18}$$

Substracting Eq. (4.17) from (4.18)

$$15 = 14I_1$$

$$I_1 = \frac{15}{14} = 1.07\text{A}$$

Substituting the value of I_1, in Eq. (4.17)

$$3 = 3I_2 + 1.07$$

$$I_2 = \frac{3 - 1.07}{3} = \frac{1.93}{3} = 0.64\text{A}$$

$$I_2 = I_1 + I_2 = 1.07 + 0.64 = 1.71\text{A}$$

II. INDUCTORS

4.12 INDUCTANCE

Inductance is defined as the property of a circuit or a component like a coil which opposes any changes in electron flow or flow of current.

The inductance provides an opposition to the sudden building up or sudden decay of current through a coil or inductor by setting up an opposing emf. This opposing emf developed by an inductor due to changes in its own current is called *self induced emf*. Its value will depend on the inductance of the coil. Inductance is a basic property of the coil itself and indicates how much voltage can be induced in it by a change of flux linkages. A coil possesses inductance in the same way as a resistor possesses resistance. Coils or inductors are designed and constructed for a known value of inductance. The main factors that determine the inductance of a coil are the number of turns and the core material, but the diameter and the spacing of turns also play a role in this regard.

4.13 UNITS OF INDUCTANCE

The unit of inductance is the henry (H). This is named after the famous scientist Joseph Henry who performed a lot of experiments with coils.

A henry is defined as the inductance that will develop a voltage of one volt across it when the current changes at the rate of one ampere per second. In other words, if a coil has an inductance of 1 H, a current change of 1 A/s will induce a voltage of 1 V in the coil.

Iron-core coils in electronic equipment have high inductance values ranging from 10 to 30 H but in some cases of iron-core chokes and in high power transmitting equipment, the inductance value may range as high as several hundred henrys.

Air-core coils used in radio and TV equipment have low values of inductance. Two smaller units of inductance are used for expressing low values of inductance. These are millihenry (mH) and microhenry (μH). A millihenry is one thousandth ($1/10^3$) and a microhenry is one millionth ($1/10^6$) of a henry. In

other words, 1000 mH is equal to 1 H and one million or 10^6 μH is equal to 1 H. Thus 5 H will be 5000 mH or 5×10^6 μH. The inductance of peaking coils in TV receivers may vary from 50 to 500 μH.

4.14 COILS—TYPES AND CLASSIFICATION

There are many ways of classifying coils or inductors.

1. Classification According to Core Material There are air-coils and iron-core coils depending upon the type of material used for the core of the coil. Air-core coils have generally air for the core material but even those coils which are wound on non-magnetic materials like rods of ceramic or plastic materials are called air-core coils. Similarly, coils having magnetic materials other than iron for the core are called iron-core coils.

2. Classification According to Frequency There are audio frequency and radio frequency coils. Audio frequency coils are generally iron-core coils known as chokes. These chokes are either used as filter chokes for providing smooth DC in radio and TV receivers or as modulation chokes in transmitters for modulating the carrier.

Filter chokes used in radio and TV receivers are coils wound on iron core with several hundred turns and may sometimes weigh 2 to 3 *lb*. The filter chokes and modulation chokes in high power transmitters, however, are much larger in size and the iron core itself may sometimes weigh several hundred pounds.

Radio frequency coils may be small coils consisting of only one or two turns as in the case of TV tuners or may consists of several hundred turns of wire wound on a form of about an inch diameter as in the case of a medium wave band of a radio receiver. Both perform the same function. When used with a capacitor they select the desired radio frequency signal and reject all others.

Radio Frequency Chokes (RFC) are small chokes used to offer opposition to flow of RF currents. These may consist of one or two turns and may be self supporting or they may consist of several hundred turns wound on a form of some non-magnetic material. The higher the frequency the less the number of turns required for the RFC.

3. Classification Based on the Method of Winding Coils sometimes have different types of windings. These could be single layer, multilayer, honeycomb, pancake or pie-section type, meant for special types of circuits.

Coils can be of fixed or variable types. In the case of the variable inductors, the inductance can be varied either in steps from tappings on the winding or made variable by a moving slug of a special magnetic material which can be moved in and out of the winding to vary inductance. This is also called permeability tuning.

Some of the common types of coils used in radio and TV receivers with their circuit symbols are shown in Fig. 4.20.

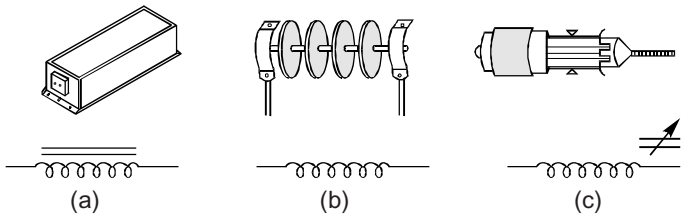


Fig. 4.20 Types of Coils (a) Iron-core (b) Air-core (c) Adjustable Iron-slug

4.15 INDUCTIVE REACTANCE

If the value of the DC voltage applied to a coil or an inductor is suddenly changed, the self induced counter emf opposes this change of current. It does not allow the current in the coil to reach its new value simultaneously with the applied voltage. The current is said to lag the applied voltage.

If an AC voltage, which is continuously changing, is applied to the above coil, the self induced voltage will have a polarity opposite to the applied voltage and will tend to limit the flow of current in the coil. But for the induced counter emf, the current in the coil will be limited only by its resistance. The opposing emf appears as an AC voltage drop across the coil produced by the opposition offered by the coil to the flow of AC current. This opposition to the flow of AC current is called *inductive reactance*. It is measured in ohms and represented by the symbol X_L. Inductive reactance is different from resistance in the sense that it opposes only AC and not DC.

The value of the counter emf and hence the value of the inductive reactance will depend upon two factors—the inductance of the coil and the rate of change of current which is determined by the frequency of the applied voltage. The inductive reactance X_L is given by the formula:

$$X_L = 2\pi fL$$

where

X_L is the inductive reactance in ohms
f is the frequency in hertz
L is the inductance in henrys
π is a constant equal to 3.14

EXAMPLE: Find the inductive reactance of a 10 H choke coil at a frequency of 50 Hz.

$$X_L = 2\pi fL$$

here $f = 50\text{Hz}$, $L = 10$ H

and $2\pi = 2 \times 3.14 = 6.28$

Therefore $X_L = 6.28 \times 50 \times 10 = 3140\ \Omega$

4.16 MUTUAL INDUCTANCE

If two coils are so situated that some of the flux produced by one coil cuts the turns of the other coil, the coils are said to be mutually coupled or they are said

to possess *mutual inductance*. For the coils to possess mutual inductance they are either wound on the same core or they are merely placed over each other.

Any change of flux in one coil will induce a voltage in the other coil. Mutual inductance M is measured in henrys like the inductance of a coil. The greater the value of M, the greater is the voltage produced in one coil due to the current change in the other coil.

Mutual inductance between two coils will be 1 H when a rate of change of current of 1 A/S in the primary coil produces a voltage of 1 V in the secondary coil.

The degree of magnetic coupling is referred to as the *coefficient of coupling* and is represented by the letter k. If all the lines from one coil cut the turns of the other coil, k equals one and these coils are said to be completely coupled. In practice, however, k is always less than one.

$$M = k\sqrt{L_1 L_2}$$

where M = Mutual inductance in henrys

k = coefficient of coupling

L_1 and L_2 are the inductances of the individual coils in henrys.

Air-cored coils are said to possess "close" coupling even when the coefficient of coupling is 0.5, while a coefficient of only a few hundredths represents "loose" coupling.

4.17 COILS IN SERIES AND PARALLEL

Coils can also be connected in series and parallel combinations like resistors. In the series arrangement shown in Fig. 4.21, the total inductance of the combination L will depend upon whether the coils are having any mutual inductance or not. Where there is no mutual inductance between the coils, the total inductance is the sum of individual inductances.

$$L = L_1 + L_2 + L_3 + ...$$

Two coils, one with an inductance of 10 H and the other with an inductance of 5 H, when connected in series without mutual inductance will have a total inductance of 10 + 5 = 15 H.

When two coils connected in series have a mutual inductance M and the coils are so connected that the flux linkages due to current in one coil add to

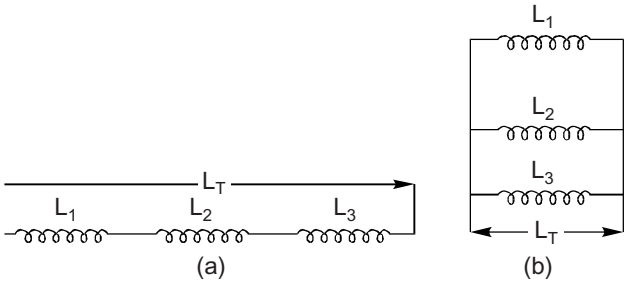


Fig. 4.21 (a) Coils in Series (b) Coils in Parallel

the self linkages of the other coil, the total inductance L of the combination is given by the formula:

$$L_T = L_1 + L_2 + 2M$$

If, however, the flux linkages of one coil oppose the self linkages of the other coil, which will be the case if the connections of one of the coils are reversed, the total inductance of the combination is given by the formula

$$L_T = L_1 + L_2 - 2M$$

In the case of the parallel combination indicated in Fig. 4.21 (b)

$$\frac{1}{L_T} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \ldots$$

For two inductors L_1 and L_2 connected in parallel, the simplified formula will be

$$L_T = \frac{L_1 \times L_2}{L_1 + L_2}$$

which is the same as the formula for two resistors connected in parallel. Thus, two inductances of 10 H each, when connected in parallel will have a total inductance of 5 H, which is less than the value of either inductance. In practice however, it is preferable to use a single coil of the required value rather than connecting coils in parallel to reduce the value of inductance.

4.18 PHASE RELATIONSHIP

In a pure inductance (coil without resistance) the application of AC voltage results in the production of self induced voltage or counter emf which does not allow the current in the inductor to reach its maximum value simultaneously with the maximum value of applied voltage. There is a delay or time lag between the applied voltage reaching this maximum value and the current reaching its maximum value. This delay which results from counter emf depends on the rate of change of current. This rate of change is different at different parts of a sine wave as shown in Fig. 4.22.

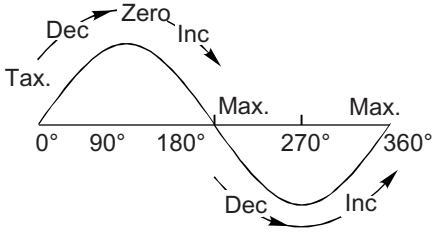


Fig. 4.22 Rate of Change at different Parts of a Sine Wave

It will be seen that the rate of change which is maximum when the current starts at 0° begins to decrease as the sine wave progresses towards 90°. At this stage the current has reached its maximum value. From here the rate of change starts increasing again and reaches its maximum value again when the sine wave passes through 180° and crosses over to be negative side.

At the 0° point where the rate of change of current is the maximum (actual value 0) the applied voltage will be passing through its maximum value. As the current approaches the 90° point, its rate of change becomes zero (actual value

maximum) and therefore, the applied voltage also falls to zero value. From this it would appear that the applied voltage reaches its maximum value 90° before the current reaches its maximum point. The current is said to lag the applied voltage by 90° or the voltage is said to lead the current by 90°. A combined picture of a current sine wave and a voltage sine wave in an inductance indicating the current lag of 90° is shown in Fig. 4.23.

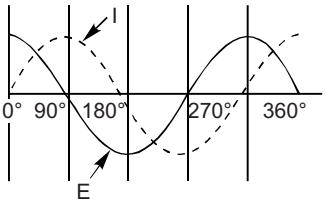


Fig. 4.23 Phase Relationship of Voltage and Current in an Inductance

The phase relationship between the current and applied voltage in an inductor can also be indicated by vector diagrams as in Figs 4.24 (a) and (b).

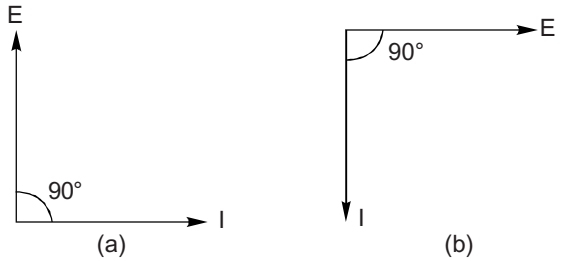


Fig. 4.24 Vector Representation of Phase Relationship Between Current and Voltage in an Inductance

4.19 RL CIRCUITS

So far we have treated a coil as a pure inductance. This means that the coil has inductance only and no resistance. This is, however, not true in actual practice. All coils are wound from wires which must have some resistance. In addition, there may be some other resistance connected separately in series with the coil. Generally, for calculation purposes, the resistance of the coil and the external resistance are lumped together and shown as a single resistance R, as in the Fig. 4.25. The coil will then have only inductive reactance X_L. If an AC voltage E is applied across the circuit, the same current I will flow through R and X_L. The voltage drop E_R across the resistance R will be in phase with the current but the voltage drop E_{XL} across the reactance X_L will be leading the current as shown in Fig. 4.26.

The total voltage drop across the combination of R and X cannot be found by merely adding E_R and E_{XL} ($E_R + E_{XL}$) because of the phase difference of 90° between the two voltages. The resultant voltage E, which is the applied voltage, can be found only by adding E_R and E_{XL} vectorially as in Fig. 4.26. This vectorial addition can be done either by completing the rectangle ABCD and measuring the length of the diagonal AC and converting this length into volts from the same scale as used for E_R and E_{XL}, or the resultant can be found by using the formula.

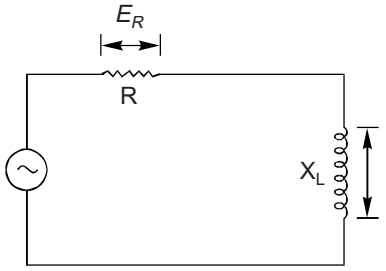


Fig. 4.25 RL Circuit

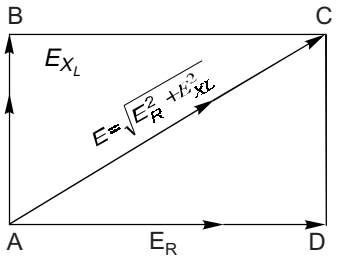


Fig. 4.26 Vectorial Addition of E_R and E_{XL}

$$E = \sqrt{E_R^2 + E_{XL}^2}$$

Let us make this clear by taking a numerical example:
If E_R =30 V and E_{XL} = 40 V, E will not be 30 + 40 = 70 V but

$$E = \sqrt{(30)^2 + (40)^2} = \sqrt{900 + 1600} = \sqrt{2500} = 50 \text{ V}$$

Thus, the resultant voltage E is not 70 V but 50 V which is the vector sum of 30 V and 40 V.

4.20 Q OF A COIL

The ratio of the inductive reactance (X_L) to the internal resistance (R) of a coil is known as Q of the coil. Expressed as an equation

$$Q = \frac{X_L}{R}$$

Q of a coil is a pure number and is not expressed in any units.

Thus, a coil with X_L equal to 200 Ω and R equal to 5 Ω has a Q value of 200/5 = 40. Q indicates the merit or the ability of a coil to develop induced voltages under the influence of magnetic fields. A high Q coil will develop a higher voltage across it than a low Q coil under the same conditions. The Q of the coil plays an important part in the design of RL circuits for electronic equipment.

4.21 IMPEDANCE

The total opposition offered by the combination of X_L and R to the flow of current is called *impedance*. This is the vector sum of the opposition offered by the resistance R and the inductive reactance X_L because R and X_L are out of phase with each other. Having found the impedance Z by the vector sum method, it is easy to calculate the current I by application of Ohm's law:

$$I = \frac{E}{Z}$$

EXAMPLE: Find the current flowing through a series combination of resistance R = 60 Ω and inductive reactance X_L = 80 Ω when an AC voltage of 50V is applied across the combination

$$R = 60\ \Omega,\ X_L = 80\ \Omega$$

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{(60)^2 + (80)^2} = \sqrt{3600 + 6400}$$

$$= \sqrt{10000} = 100\ \Omega$$

$$\text{Current } I = \frac{E}{Z} = \frac{50}{100} = 0.5\ \text{A}$$

4.22 TRANSFORMER ACTION

A changing magnetic flux produces a counter emf in a coil due to self induction. If there is another coil wound on the same core or placed in close proximity to the first one, the changing magnetic flux will cut the turns of the second coil and so induce a voltage in the second coil. There will thus be a transfer of energy from one coil to another through the magnetic flux linkages. This is the principle of transformer action.

Basically, a transformer consists of two or more coils either wound on the same core or placed close enough so that magnetic lines of force from one coil will cut the turns of the other coil. These coils are called *windings.* The coil to which the input voltage is applied is called the *primary* winding. The other coil in which the voltage is induced due to the changing magnetic flux is called the *secondary* winding (see Fig. 4.27).

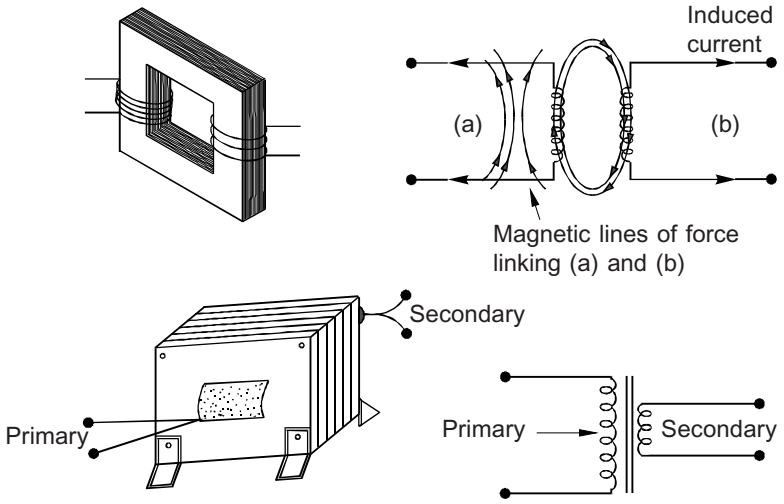


Fig. 4.27 Primary and Secondary Windings of a Transformer

4.23 SOME TECHNICAL TERMS

Turns Ratio The changing magnetic flux from the primary winding cuts each turn of the secondary winding and induces a small emf in it. The emfs induced in individual turns add up to form the total induced voltage. The magnitude of the induced voltage will be proportional to the total number of turns in the secondary winding compared to the total number of turns in the

primary winding. The ratio of the number of turns in the primary to the number of turns in the secondary is called the *turns ratio*.

$$\text{Turns ratio} = \frac{N_p}{N_s}$$

where N_p is the number of turns in the primary winding.
N_s is the number of turns in the secondary winding.

Turns ratio is a very important characteristic of a transformer. A transformer with 2000 turns in the primary winding and 1000 turns in the secondary winding has a turns ratio of two to one which is also written as 2:1.

Voltage Ratio Since the voltage induced in a winding depends on the number of turns being cut by the magnetic flux, the ratio of the voltage in the primary winding to the voltage in the secondary winding will be the turns ratio.

Thus
$$\frac{E_p}{E_s} = \frac{N_p}{N_s}$$

where E_p is the voltage in the primary winding.
E_s is the voltage in the secondary winding.
N_p and N_s are the number of turns in the primary winding and the number of turns in the secondary winding respectively.

If the number of turns in the secondary winding is *greater* than the number of turns in the primary winding, the transformer is called a *step up* transformer. A transformer with a turns ratio of 1:2 is a step up transformer. In this transformer a voltage of 110 V applied to the primary winding will appear as 220 V in the secondary winding as in Fig. 4.28. A *step down* transformer is one in which the number of turns in the secondary is less than the number of turns in the primary winding. A transformer with a turns ratio of 5 : 1 is a step down transformer.

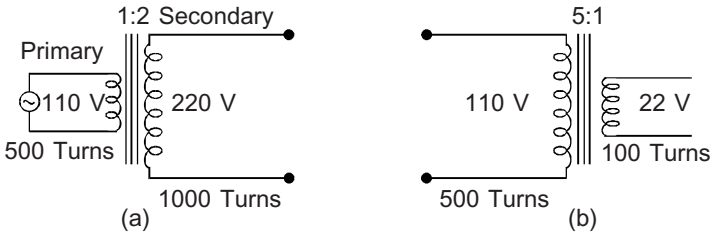


Fig. 4.28 (a) Step Up Transformer (b) Step Down Transformer

Current Ratio If there are no losses in a transformer, the input power $E_p \times I_p$ will be equal to the output power $E_s \times I_s$ where I_p and I_s are the currents flowing in the primary and the secondary windings respectively. Therefore,

$$E_p \times I_p = E_s \times I_s$$

or
$$\frac{E_p}{E_s} = \frac{I_s}{I_p}$$

But $$\frac{E_p}{E_s} = \frac{N_p}{N_s} \text{ as already proved}$$

So $$\frac{I_s}{I_p} = \frac{N_p}{N_s}$$

Combining the two relations for voltage ratio and current ratio, we have a combined formula for the two ratios

$$\frac{N_p}{N_s} = \frac{E_p}{E_s} = \frac{I_s}{I_p}$$

When expressed in words the above formula will mean that whereas the voltage ratio is directly proportional to the turns ratio, the current ratio is inversely proportional to the turns ratio. It means that in a step down transformer the voltage in the secondary winding is less than the voltage in the primary, but the current in the secondary winding is more than the current in the primary. The reverse is the case with a step up transformer.

EXAMPLE: A mains transformer has a 240 V primary winding and a 12 V secondary winding for a low voltage supply unit. How much current can be drawn from the secondary for a 250 mA current in the primary?

We use the equation $$\frac{I_p}{I_s} = \frac{E_s}{E_p}$$

$I_p = 250$ mA $= 0.250$ A

$E_p = 240\ V$ $E_s = 12$ V

So, $$\frac{0.250}{I_s} = \frac{12}{240}$$

or $$I_s = 0.250 \times \frac{240}{12} = 5.0 \text{ A}$$

4.24 TRANSFORMER LOSSES

Transformer Efficiency So far we have considered an ideal transformer in which there are no losses and the input power is equal to the output power. In practice, this is not the case. All transformers have some losses due to various causes. These losses appear in the form of heat and the output is always less than the input. Methods are adopted in the design and construction of transformers so that the losses are reduced to a minimum thereby increasing the efficiency of a transformer.

The efficiency of a transformer is the ratio of the output power to the input power

$$\text{Efficiency} = \frac{\text{Output power}}{\text{Input power}} = \frac{E_s \times I_s}{E_p \times I_p}$$

Efficiency is generally expressed as a percentage.

$$\text{Percentage efficiency} = \frac{\text{Output power}}{\text{Input power}} \times 100$$

Thus a transformer with an input power of 100 W and output power of 90 W has an efficiency of 90%. Most iron-core transformers have an efficiency of over 90% and the first grade transformers are better than 90% efficient.

Transformer Losses There are four types of losses in a transformer. These are (i) copper loss, (ii) flux loss, (iii) eddy current loss, and (iv) hysteresis loss. The causes of these and the methods adopted to reduce these losses in transformers will now be described.

(i) Copper Loss This loss is due to the resistance of the wire used for the windings and is expressed as I^2R, where I is the current in the winding and R is the resistance of the winding. This loss is also sometimes referred to as "I squared R" (I^2R) loss. This loss would be reduced by using a thicker gauge of wire for the windings.

(ii) Flux Loss or Flux Leakage This loss is caused when all the flux lines associated with the primary winding of a transformer do not pass through the secondary winding. The flux lines that do not pass through the secondary result in a flux loss as in Fig. 4.29. This loss is also referred to as flux leakage. Flux loss can be reduced by placing a core of a suitable magnetic material in the transformer to improve flux linkages. This, however, creates the problem of *eddy currents* and *hysteresis* loss. Apart from the power loss, flux leakage can induce undesirable voltages in neighbouring coils and transformers. This is prevented by suitable placement of transformers in electronic circuits or by screening the neighbouring coils and components.

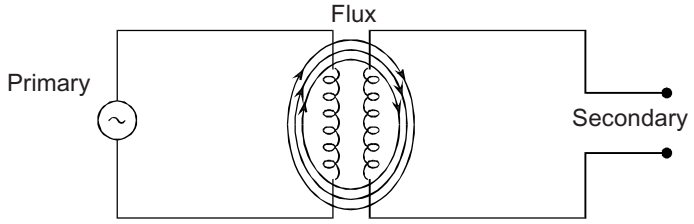


Fig. 4.29 Flux Leakage

(iii) Eddy Current Losses The iron core of transformers is a conducting material. Due to the changing flux that cuts the core, emf is induced in the core material resulting in the flow of small local currents called *eddy currents.* These currents also produce a loss which is similar to the I^2R loss.

Eddy current losses are reduced by using for core material thin strips of silicon steel called *laminations.* These laminations are coated with insulating varnish so as not to provide any paths for the eddy currents. These laminations are available in E and I forms (Fig. 4.30) to facilitate assembling a laminated core for the coils wound on a former.

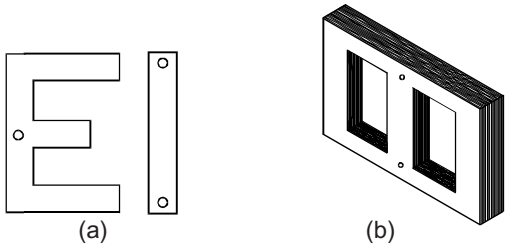


Fig. 4.30 Core Laminations

Slugs made out of powdered magnetic materials mixed with an insulating binding material are also used to reduce eddy current losses. These slugs can be moved in and out of the coils to vary their inductance value.

(iv) Hysteresis Losses Depending on the frequency of the applied AC, the core material gets magnetised and demagnetised and changes polarity a number of times in one second. This change of polarity is opposed by a magnetic friction called hysteresis and so represents a loss. Hysteresis loss cannot be completely eliminated. It can be reduced by using suitable core materials like soft iron and silicon steel.

The last three losses are due to properties of the core material and are called "core losses". In other words, the main causes of transformer loss are copper loss and core loss.

Transformers used in radio and TV equipment are broadly classified as (i) power transformers (ii) audio frequency (AF) transformers (iii) radio frequency (RF) transformers.

4.25 POWER TRANSFORMERS

Power transformers are also called mains transformers and are meant to operate on the mains AC supply of 50 or 60 Hz. These transformers are meant to supply various operating voltages required in radio receivers and other electronic equipment. This transformer has one primary winding and a number of secondary windings as shown in Fig. 4.31. The standard colour code employed

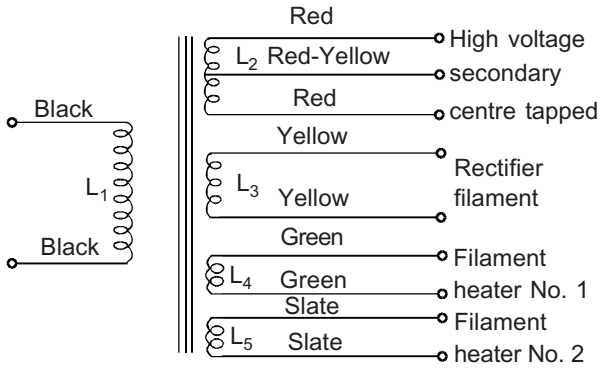


Fig. 4.31 Power Transformer Colour Coding

by EIA (Electronic Industries Association) for power transformers is also shown in Fig. 4.31. Some manufactures use their own colour codes which must be ascertained before the transformer is connected in the circuit.

4.25.1 Audio Frequency (AF) Transformers

Audio frequency transformers or AF transformers are meant to operate on audio frequencies of 20 Hz to 20 kHz. These transformers are similar to power transformers but the core material is a nickel-iron alloy of high permeability specially made to operate over the wide range of frequencies without causing much distortion. AF transformers are used in audio amplifiers for interstage coupling, for coupling microphones to amplifiers and for coupling amplifiers to speakers. Audio amplifiers use both step up and step down ratios. In the final power output stage of a radio receiver, the audio transformer is a step down transformer used to match the high output impedance of the amplifier to the low (3 to 8Ω) impedance of the speaker. Since a transformer can serve as a step up or step down device for voltage as well as currents, it can also serve as a step up or step down device for impedance which is only the ratio of voltage to current. The formula to be used for impedance matching is

$$\frac{N_p}{N_s} = \sqrt{\frac{Z_p}{Z_s}}$$

where Z_p and Z_s are the impedances in the primary and secondary, and N_p and N_s are the number of turns in the primary and secondary respectively (see Fig. 4.32).

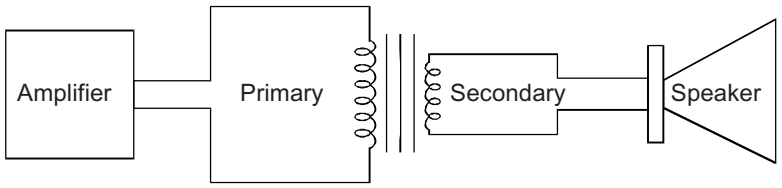


Fig. 4.32 Impedance Matching with a Transformer

EXAMPLE: What is the turns ratio of a step down transformer required for matching 8000 Ω output impedance of an audio amplifier to the 5 Ω voice coil impedance of the speaker.

$$\frac{N_p}{N_s} = \sqrt{\frac{Z_p}{Z_s}} = \sqrt{\frac{8000}{5}} = \sqrt{\frac{1600}{1}} = \frac{40}{1}$$

Therefore, the turns ratio $\frac{N_p}{N_s}$ = 40:1

4.25.2 Radio Frequency (RF) Transformers

RF transformers are meant for operation at very high frequencies compared to AF transformers. At high frequencies or radio frequencies the eddy current losses become excessive with nickel-iron laminated cores.

To reduce eddy current losses, the RF transformer is either made an air-core transformer or a powdered iron-core transformer. The powdered iron core or dust-iron core is in the form of a slug which is threaded so that it can be moved into or out of the coil for tuning purposes. This form of tuning is called permeability tuning. Radio frequency transformers are used for interstage coupling and for coupling a signal to and from an antenna. In most cases, the secondary has a variable capacitor across it to form a series resonant circuit at the desired frequency. This results in a series resonant voltage step up even when the turns ratio is 1 : 1. This phenomenon will be explained under series resonance.

4.25.3 Intermediate Frequency (IF) Transformers

IF transformers or intermediate frequency transformers are special types of RF transformers designed to operate at a fixed frequency called the intermediate frequency or IF. In modern superheterodyne radio receivers the IF commonly used is 455 kHz but in TV receivers it is much higher. The two windings are either both tuned (Fig. 4.33) or at least one of these is tuned to the IF frequency by associated capacitors and adjustable dust-iron-cores.

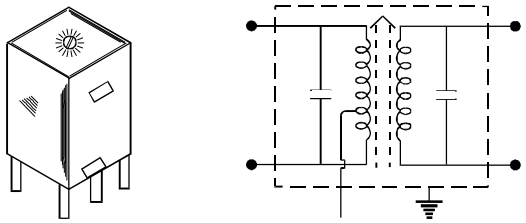


Fig. 4.33 IF Transformer (a) Screened IF Transformer (b) Circuit Symbol for IF Transformer

The magnetic coupling between the windings is crucial in selecting the pass band and is set to allow certain frequencies above and below IF also to pass through along with IF. This is necessary to include side bands corresponding to higher audio frequencies. IF transformers are enclosed in metallic cases for screening as in Fig. 4.33. Provision is made for adjustment of the capacitors or the dust-iron-core for alignment purposes.

4.25.4 Auto Transformers

In an auto transformer, the same coil is used to provide turns for the primary and the secondary windings. When the whole coil is used as the primary and a part of the coil as the secondary, it becomes a step down transformer as in

Fig. 4.34(a). In a step up transformer the entire winding is used as the secondary as in Fig. 4.34(b). Auto transformers are sometimes provided with fixed taps or a variable tap for variation of the secondary voltage.

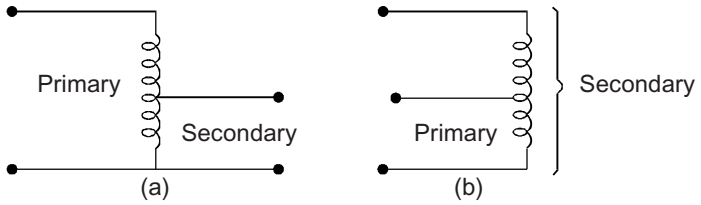


Fig. 4.34 Auto Transformer (a) Step Down (b) Step Up

An auto transformer is different from a conventional transformer in the sense that there is no isolation between the primary and the secondary windings.

III. CAPACITORS

4.26 CAPACITOR-DEFINITION

A capacitor or a condenser is an electrical device for storing electrical energy by allowing electrons to accumulate on a metallic surface. This electrical charge can then be released in the form of a current into the circuit of which the capacitor forms a part. Like resistors and inductors, capacitors are also important components used in radio, TV and other electronic circuits.

In its simplest form a capacitor consists of two metallic surfaces separated by a finite distance. The space between the two metal plates is filled with an insulating material which may be air (or a gas) as in Fig. 4.36(a), or liquid or a solid as in Fig. 4.35(b). This material separating the two plates of a capacitor is called the *dielectric*. Air is also a dielectric and when there is nothing between the plates except air, the capacitor is said to have air as a dielectric. In fact, various types of capacitors used in electronic equipments are named after the dielectric used in their construction like paper capacitors, mica capacitors, ceramic capacitors, electrolytic capacitors, and so on.

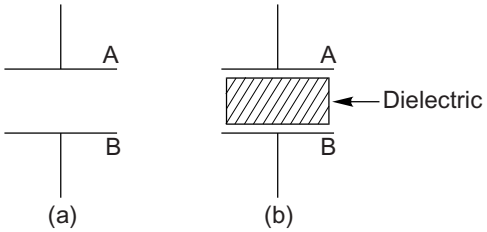


Fig. 4.35 Formation of a Capacitor (a) Air Dielectric (b) Solid or Liquid Dielectric

4.27 CAPACITOR ACTION

Figure 4.36(a) shows a capacitor with the two metallic plates *A* and *B* having air as a dielectric. The capacitor can be connected to a battery through a switch *S*. When the switch *S* is open, no electrons can flow from the negative terminal of the battery to plate *A* of the capacitor. As soon as the switch *S* is closed as in Fig. 4.36(b), electrons from the negative terminal of the battery rush to plate *A*. There will be a surplus of electrons built upon plate *A*.

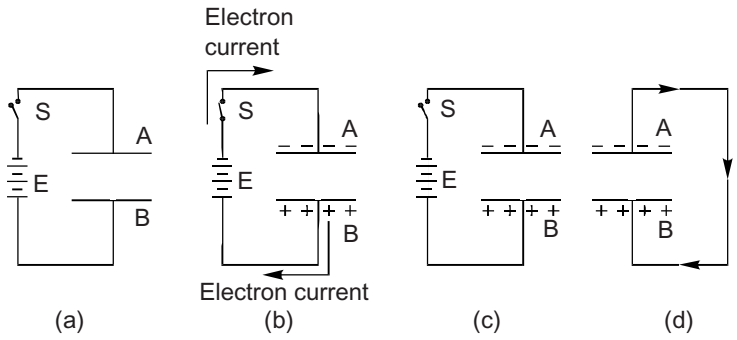


Fig. 4.36 Charging and Discharging a Capacitor

Since like charges repel, the electrons on plate *A* will repel the electrons on plate *B* and drive them back into the positive terminal of the battery *E*. At the same time, the positive charge on the positive terminal of the battery will also pull out electrons from plate *B*, leaving a deficiency of electrons on this plate. The positive charge on *B* will attract further electrons from the negative terminal of the battery on to plate *A*. There will thus be a surplus of electrons on plate *A* which will get negatively charged. Similarly, the deficiency of electrons on plate *B* will make it positively charged. This flow of electrons which constitutes a current will continue till plate *A* becomes as negative as the negative terminal of the battery and plate *B* becomes as positively charged as the positive terminal of the battery. If the switch *S* is now opened as in Fig. 4.36(c) the surplus of electrons on plate *A* and the deficit of electrons on plate *B* of the capacitor will stay. The capacitor is said to be *charged* to the voltage of battery *E*.

If plate *A* of the capacitor is now connected to plate *B* of the capacitor through a copper wire as is Fig. 4.36 (d), the surplus electrons from plate *A* will rush to plate *B* and neutralise the positive charge on this plate. When the voltage of plate *A* becomes equal to the voltage of plate *B* the capacitor is said to be discharged. This discharge current can be made to pass through a coil or a resistor and made to perform one or other of those wonderful tricks that electronic circuits are capable of performing. If the capacitor is connected to an AC voltage, the charging and discharging process will be performed a large number of times, depending on the frequency of the AC voltage.

4.28 CAPACITY

The amount of water that a water storage tank can hold depends on its size. Similarly, the amount of electrical charge that a capacitor can hold depends on its electrical size. This electrical size of a capacitor is called its *capacity* or *capacitance*.

When a capacitor is charged, it builds up a voltage which opposes the original voltage that is trying to charge the capacitor. Capacitance is also the property of a circuit that opposes any change of voltage in the circuit. It may be noted that an inductance opposes any change of current in a circuit.

In terms of charge and voltage, capacitance C can be expressed by means of the formula

$$C = \frac{Q}{E}$$

where Q = charge in Coulomb
and E = voltage in volts

In other words, the capacitance of a capacitor is the charge it will hold for every unit of potential to which its plates are charged.

$$Q = 1 \text{ Coulomb}, \; E = 1V, \quad \text{then } C \text{ will be } 1$$

The unit of capacitance is called a farad (F). This is the capacitance of a capacitor that will hold a charge of 1 Coulomb when the potential difference between its plates is 1V.

Units of Capacitance The farad, as defined above, is a large unit of capacitance. Similar units of capacitance, known as microfarad (μF) and micro-microfarad ($\mu\mu$F), are used in actual practice. A micro-microfarad $\mu\mu$F is also called a picofarad (pF). The relation between a farad and the other smaller units mentioned above is as given below

$$1 \text{ F} = 10^6 \mu\text{F}$$
$$1 \text{ F} = 10^{12} \; \mu\mu\text{F(pF)}$$
$$1\mu\text{F} = 10^6 \; \mu\mu\text{F(pF)}$$

The conversion of one unit of capacitance to another sometimes involves multiplication and division by decimal fractions and should be carefully done.

EXAMPLE: Find the charge on a 1000 μF capacitor when charged to a voltage of 24 V.

$$C = \frac{Q}{V}$$

$$C = 1000\mu\text{F} = \frac{1000}{10^6} \text{ F} = 1000 \times 10^{-6} \text{ F}$$
$$= 0.001 \text{ F}$$
$$Q = CV = 0.001 \times 24$$
$$= 0.024 \text{ Coulomb}$$

4.29 FACTORS AFFECTING CAPACITANCE

The capacitance of a capacitor depends on the following three factors: (a) the area of the plates, (b) the distance between the plates, and (c) the dielectric material.

Area of the Plates The capacitance of a capacitor is directly proportional to the area of the plates forming the capacitor. If the area is doubled the capacitance is also doubled and if the area is halved the capacitance also becomes one half.

Distance Between the Plates The capacitance is inversely proportional to the distance between the plates. In other words, the greater the distance between the plates, the smaller will be the capacitance of the capacitor.

Dielectric Material The type of material used as dielectric between the plates of a capacitor also affects the capacitance. Thus, a capacitor with mica as the dielectric will have a greater capacitance than a capacitor with air as the dielectric. The property of the dielectric which affects the capacitance is called the dielectric constant.

The higher the dielectric constant of the material between the plates, the larger will be the capacity of the capacitor.

The dielectric constants of some of the common dielectric materials are given in Table 4.4.

Table 4.4 Dielectric Constants of Some Common Materials

Air	1.00
Castor oil	4.7
Mica	5-9
Glass	4.5-7.00
Bakelite	4.5-7.5
Paper	2-2.3
Porcelanin	5.5
Quartz	1.5
Water	81

4.30 VOLTAGE RATING

Capacitors are designed and manufactured to operate at a certain maximum voltage which depends on the distance between the plates of the capacitor. If the voltage is exceeded, the electrons jump across the space between the plates and this can result in permanent damage to the capacitor. The maximum safe voltage is called the working voltage. This working voltage (WV) is marked on the capacitor in the case of bigger capacitors and indicated by a colour code in the case of capacitors having low values of capacitance. In the case of valve type radio receivers and TV receivers, capacitors with a working voltage of 200 to 600 V are used. In transistor receivers, however, capacitors with 6 or 12 V working voltage are commonly used.

In electrolytic capacitors for filter circuits of a power supply, *peak voltage* is also marked in addition to the WV. The output from a rectifier tube is a pulsating DC, where peak voltage during a part of the cycle will be much higher than the DC voltage. The capacitor should be able to stand this peak voltage also. Thus an electrolytic capacitor marked "450 V DC Wkg, peak volts 500" is meant for use in a filter circuit where the DC voltage is 450 V and the peak voltage does not exceed 500 V. A capacitor with a lower voltage rating should not be used in place of a capacitor with a higher voltage rating. However, a capacitor with a higher voltage rating can be substituted for a capacitor with a lower voltage rating, if the physical size of the capacitor is not a problem.

4.31 TYPES OF CAPACITORS

Capacitors are either fixed or variable capacitors. A fixed capacitor is one in which the value of the capacitance cannot be varied. The capacitance value remains fixed. On the other hand, in a variable capacitor the design is such that the capacitance value can be varied within certain limits. Both types are commonly used in radio and TV circuits.

4.31.1 Fixed Capacitors

Fixed capacitors use dielectric materials other than air. In fact, the fixed capacitors are classified according to the type of material used as a dielectric in their construction. The various categories of fixed capacitors are

1. Paper capacitors
2. Mica capacitors
3. Ceramic capacitors
4. Electrolytic capacitor

1. Paper Capacitors Paper capacitors are made by taking two sheets of tin foil and placing a sheet of paper between them as shown in Fig. 4.37. The foil and paper are then rolled in the form of a cylinder and wire leads are attached to the foil sheets that come out at both ends. The entire cylinder is then encased in cardboard coated with wax or put in a plastic cover.

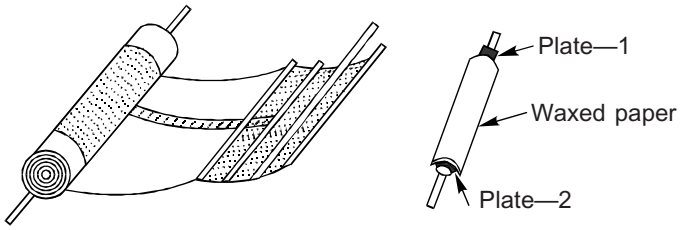


Fig. 4.37 Paper Capacitors

Modern paper capacitors are completely enclosed in a mylar type of material to make them moisture proof. These are known as moulded paper capacitors or

mylar paper capacitors. The range of paper capacitors extends from 0.0005 μF to 1 or 2 μF.

2. Mica Capacitors Mica capacitors are constructed by thin metal sheets or tin foils, arranged one over another and separated by thin mica sheets placed between the metal sheets. By combining alternate metal sheets, two sets of metal plates are formed to which separate terminals are connected as shown in Fig. 4.38. The complete unit is enclosed in a moulded ceramic or Bakelite case with terminals coming out at each end. Mica being brittle, these capacitors cannot be rolled into tubular form and are generally available in rectangular shape.

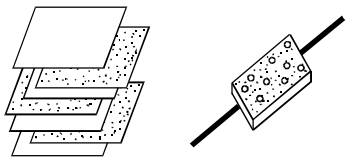


Fig. 4.38 Mica Capacitors

Mica capacitors are specially suitable for use in RF circuits. These capacitors are more expensive than paper capacitors of the same capacity. High voltage capacitors (5000 V and above) used in radio transmitters are generally mica capacitors.

Silver-mica capacitors are made by depositing a thin film of silver on both sides of a mica sheet by a chemical process. Several such silver-coated mica sheets are connected in parallel to obtain the required value of capacitance. Mica capacitors are made with capacitance values ranging from a few pF to 10,000 pF or so.

3. Ceramic Capacitors Ceramic capacitors are commonly available in three different types—disc, tabular and button type as shown in Fig. 4.39.

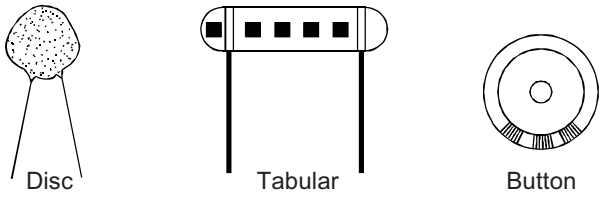


Fig. 4.39 Ceramic Capacitors

Disc capacitors are more economical to use than other types. These vary in capacitance from 1 pF to 1 μF. Voltage rating is normally 500 V but ceramic capacitors of 1000 V rating are also available.

4. Electrolytic Capacitors There are two types of electrolytic capacitors, the wet type and the dry type.

A wet electrolytic capacitor basically consists of an aluminium electrode placed in a solution of borax contained in an aluminium container as shown in Fig. 4.40. When a DC current is passed through the solution, a thin oxide film is formed on the aluminium electrode connected to the positive side of the DC supply.

The current stops flowing as soon as the positive plate is completely covered by a thin oxide film. A capacitor is formed in which the positive aluminium

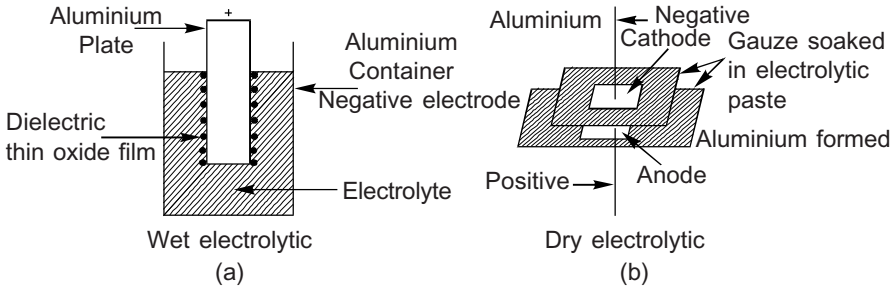


Fig. 4.40 Electrolytic Capacitors (a) Wet Electrolytic (b) Dry Electrolytic

electrode is one-plate, the thin oxide film the dielectric and electrolyte the negative plate, but actually the aluminium container which is in contact with the electrolyte serves as the negative plate of the capacitor.

In the case of the dry electrolytic capacitors, the electrolyte is made in the form of a paste or a jelly. A piece of gauze soaked in the electrolyte paste is placed between the anode and the cathode. The anode is treated chemically to form an oxide layer as dielectric, before the capacitor is assembled. The cathode or the negative plate is only a pure aluminium sheet touching the electrolyte. These capacitors are known as aluminium electrolytic capacitors. However, tantalum plates are used in place of aluminium when the size of the capacitor is a consideration. Tantalum electrolytic capacitors are expensive and are not used in radio and TV circuits.

Dry electrolytic capacitors can be rolled into tabular form like paper capacitors. Electrolytic capacitors are polarised and have a positive and negative polarity. So while connecting in a circuit, the terminal marked + (positive) should be connected to the positive side of the voltage source. Failure to do this may damage the capacitor. The polarity is either marked on the capacitor by + sign or the polarity is indicated by coloured leads. A red lead usually indicates positive polarity. Sometimes, more than one capacitor is enclosed in the same metal container and this container serves as the common negative; the positive terminals are brought out through an insulated washer or cover.

4.31.2 Variable Capacitors

Variable capacitors are so designed that the capacitance value can be varied continuously within fixed limits. The variable capacitors may use air as the dielectric as in the case of gang capacitors. However, in the adjustable type of capacitors known as trimmers and padders, the dielectric is a solid material like mica.

1. Gang Capacitors In this case the capacitance is varied by varying the overlapping area between two sets of plates. One set remains stationary and is called the *stator*. The other set of plates can be rotated by means of a shaft and is called the *rotor*. Whereas the stator is insulated from the frame, the rotor is generally connected directly to the shaft and gets automatically grounded when mounted on a metal chassis or frame. Figure 4.41(a) shows a two-gang capacitor commonly used in modern valve type receivers. Gang capacitors used in

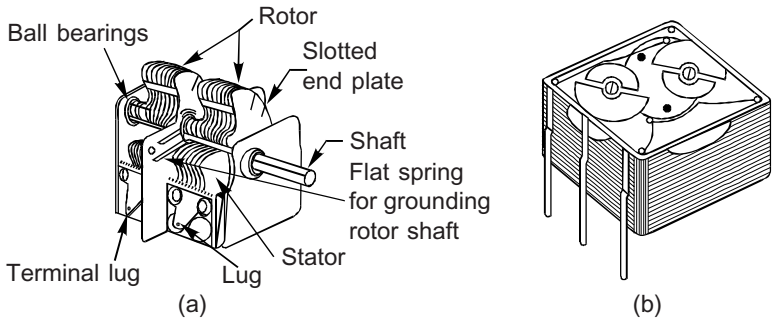


Fig. 4.41 Gang Capacitors (a) Two-gang Capacitor for Tube Type Receivers (b) Gang Capacitor for Transistor Receivers

transistor type receivers are miniature type (Fig. 4.42(b)) with the entire unit enclosed in a plastic cover. 3-gang and 4-gang capacitors are also used in certain special types of receivers where the number of stages to be tuned simultaneously is more than two.

2. Trimmers and Padders These are small adjustable types of capacitors using mica as the dielectric. Trimmers are used for finer adjustments of a resonant circuit in which the main tuning capacitor is much larger. The movable plate is of spring material which can be moved closer to or away from the fixed plate by means of a screw electrically insulated from the movable plate, as shown in Fig. 4.42.

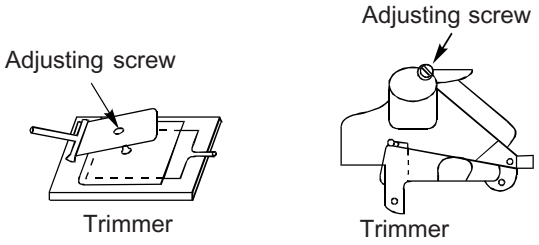


Fig. 4.42 Trimmer Capacitors

Most gang capacitors are fitted with their own individual trimmers, one for each section of the gang. Trimmers can also be mounted separately. In some cases, the trimmers and padders are miniature air dielectric capacitors like single tuning capacitors.

4.32 COLOUR CODING OF CAPACITORS

Colour coding of capacitors is very similar to the colour coding of resistors, except that in the case of capacitors the temperature coefficient is also to be indicated in addition to the capacitance value and the tolerance figures. In some capacitors the DC working voltage is also indicated by colour coding.

The capacitance value of electrolytic and paper capacitors is generally marked on the capacitor itself. However, the capacitance value of small mica and

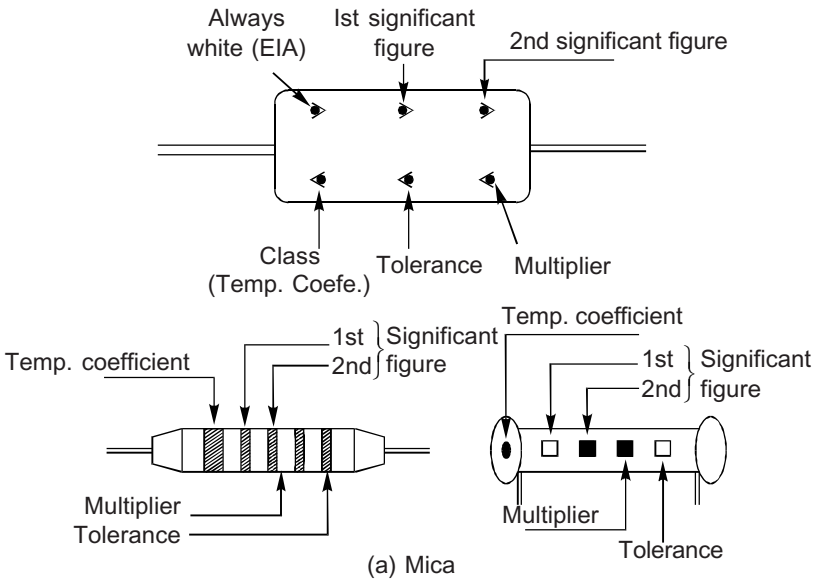
ceramic capacitors is indicated by colour coding. The colour code recommended by EIA is given in Table 4.5.

Table 4.5 Colour Coding of Capacitors

Colour	*Capacitance in Picofarads*		*Tolerance*	*Characteristics*
	Significant Figure	*Multiplier*		
Black	0	1	± 20(M)	A
Brown	1	10	± 1(F)	B
Red	2	100	± 2(G)	C
Orange	3	1000	± 3(H)	D
Yellow	4	10,000	. . .	E
Green	5	—	± 5(J)	F
Blue	6	—	—	—
Violet	7	—	—	—
Grey	8	—	—	—
White	9	—	—	—
Gold	—	0.1	± 5(J)	—
Silver	—	00.01	± 10(K)	—

Tolerance is also sometimes indicated by the letters given in brackets. Letter designations like *A*, *B*, *C*,.... indicate the temperature coefficient which is the amount of change in parts per million per degree centigrade (ppm/°C) of the nominal value at 20°C or 68°F. Detailed values of each characteristic are generally given in separate tables. Letters N and P are also used to indicate negative and positive temperature coefficients.

Methods of applying the above colour coding in mica and ceramic capacitors is as indicated in Fig. 4.43.



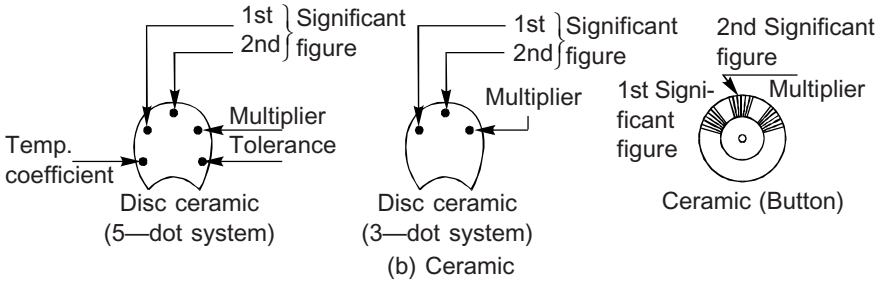


Fig. 4.43 Colour Coding of Capacitors (a) Mica (b) Ceramic

Mica Capacitors: Six dots are used. The first dot is always white indicating EIA colour code. These are read from left to right in a clockwise direction.
Ceramic Capacitors Both colour dots and colour bands are used in the case of ceramic capacitors.

4.33 CAPACITORS IN COMBINATION

Capacitors arranged in series and parallel combinations do not behave in the same way as resistors or inductors do when arranged in similar combinations. Take the case of three capacitors C_1, C_2, and C_3 arranged in a parallel combination as shown in Fig. 4.44.

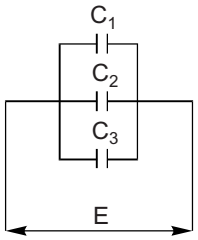


Fig. 4.44 Capacitors in Parallel

If the voltage E is applied to the combination, each capacitor will be subjected to the same voltage E but the charge flowing into each capacitor will be proportional to its capacity. The total charge Q taken by the combination will be the sum of the charges Q_1, Q_2 and Q_3 taken by the individual capacitors C_1, C_2, C_3 respectively.

$$Q = Q_1 + Q_2 + Q_3$$

Now $C = \dfrac{Q}{E}$ by definition of C.

If C_T is the total capacitance of the combination.

$$C_T \times E = C_1 \times E + C_2 \times E + C_3 \times E$$
$$= E \times (C_1 + C_2 + C_3)$$

Cancelling E from both sides of the equation we get

$$C_T = C_1 + C_2 + C_3$$

So the total capacitance of the combination is the sum of the individual capacitances.

Thus, if a 4 μF capacitor is combined in parallel with an 8 μF capacitor, the total capacitance of the combination will be

$$4\mu\text{F} + 8\mu\text{F} = 12\mu\text{F}$$

Compare this with the case of resistors where this formula applies to resistors in *series.*

If the three capacitors C_1, C_2, and C_3 are arranged in series as in Fig. 4.45 the total voltage E is the sum of the voltages E_1, E_2, and E_3 across individual capacitors C_1, C_2, and C_3 respectively.

The capacitors being in series, the same current will flow through each, and each will get the same charge Q.

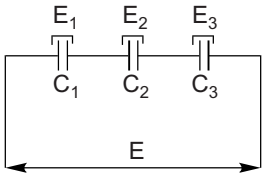


Fig. 4.45 Capacitors in Series

$$E = E_1 + E_2 + E_3$$

$$\frac{Q}{C_T} = \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3} = Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}\right)$$

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

If there are only two capacitors in series, it can proved that

$$C_T = \frac{C_1 \times C_2}{C_1 + C_2}$$

Thus if the 4 μF and 8 μF capacitors mentioned in the earlier example are connected in series, the capacitance of the combination will be

$$C_T = \frac{C_1 \times C_2}{C_1 + C_2} = \frac{4 \times 8}{4 + 8} = \frac{32}{12} = 2.66\mu\text{F}$$

Notice that this formula, in the case of resistors, is applicable to resistors in parallel.

4.33.1 Voltage Ratings of Capacitors in Combination

The working voltage of a parallel combination of capacitors is the lowest working voltage of any of the capacitors in the parallel combination. If two capacitors connected in parallel have 600 and 400 V as the working voltages, the working voltage of the combination will be 400 V only.

In the case of a series combination, however, the working voltage of the combination is the sum of the working voltages of the individual capacitors. Thus, in the case of the above two capacitors with working voltages of 600 and 400 V, the working voltage of the combination in series will be 600 V + 400 V = 1000 V.

EXAMPLE A 0.05 μF 250 V capacitor is connected in series with a 0.1 μF 450 V capacitor. Find the capacity and the working voltage of the combination. Applying the formula

$$C_T = \frac{C_1 \times C_2}{C_1 + C_2}$$

$$C_T = \frac{0.05 \times 0.1}{0.05 + 0.1} = \frac{0.005}{0.15} = 0.033\ \mu\text{F}$$

Working voltage of the combination = 250 + 450
= 700 V

The combination is equivalent to a 0.033 μF, 700 V capacitor.

4.34 CAPACITIVE REACTANCE

Let us see what happens when an AC voltage is applied across a capacitor as in Fig. 4.46. During the negative half of the cycle, the electrons from the negative pole of the generator will flow to the plate *A* which gets an excess of electrons and gets negatively charged. These electrons repel the electrons from plate *B* and drive them into the positive pole of the generator which also attracts these electrons. Plate *B* gets positively charged due to a deficiency of electrons. An electron current flows through the circuit in the direction shown by the arrow. During the positive half of the cycle the process is reversed and plates *A* and *B* get charged to the opposite polarity and a current flows through the circuit in the opposite direction. Thus, an AC current flows through the circuit even though no electrons jump across the capacitor, which behaves like a conductor in this case. If, however, a DC source is connected across the capacitor, a current flows only momentarily and stops as soon as the capacitor is charged to the voltage of the DC source. Thus, a capacitor allows an AC current to flow through it while it blocks the DC current. This is the most important property of a capacitor and is made use of in different ways in radio and TV circuits.

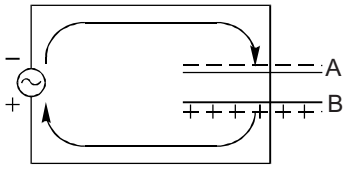


Fig. 4.46 Flow of AC Current Through a Capacitor

A capacitor does not allow the to-and-fro motion of the electrons without offering any opposition. The opposition offered by a capacitor to the flow of AC current through it is called *capacitive reactance*. This opposition is measured in ohms and is represented by the symbol X_c. The value of the capacitive reactance X_c is given by the formula

$$X_c = \frac{1}{2\pi fC}$$

where f = frequency of the AC supply and
C = the capacitance of the capacitor in farads.

It can be seen from this formula that the higher the frequency the lower is the capacitive reactance. Also, the bigger the capacitor, the smaller the value of the capacitive reactance. It is interesting to note that the behaviour of capacitive reactance (X_c) is exactly opposite to that of inductive reactance (X_L) which increases both with frequency and the value of inductance L.

EXAMPLE: Find the capacitive reactance of a 1000 μF capacitor at 50 Hz. What will be the capacitive reactance of this capacitor at 100 Hz?

$$X_c = \frac{1}{2\pi fC}$$

$f = 50\text{Hz},\ C = 1000\mu\text{F} = 100 \times 10^{-6}\text{F}$

$$X_c = \frac{1}{6.28 \times 50 \times 1000 \times 10^{-6}} = \frac{10^6}{314 \times 10^3}$$

$= 3.2\ \Omega$

At 100 Hz, the capacitive reactance will be

$$X_c = \frac{1}{6.28 \times 100 \times 1000 \times 10^{-6}}\ \Omega = \frac{10^6}{628 \times 10^3}$$

$= 1.6\Omega$

Thus when the frequency is doubled, the capacitive reactance is reduced to one half.

4.35 PHASE RELATIONSHIP

We have seen earlier that in an inductive circuit the current flowing through the inductor lags the voltage by 90°. The reverse is the case in a capacitive circuit. In a capacitive circuit, the current leads the voltage by 90°. The relation between the current and the voltage in a capacitive AC circuit is shown in Fig. 4.47. The explanation is simple. In the beginning of the AC cycle, there is no voltage developed across the capacitor and the current flowing is maximum. When the AC voltage across the capacitor builds up, the current decreases and drops to zero value by the time the capacitor is fully charged. As the voltage starts falling, the current starts rising in the opposite direction and reaches its negative peak when the voltage reaches zero. During the other half of the AC cycle, the same process is repeated in the negative direction, the current always reaching its maximum 90° ahead of the voltage.

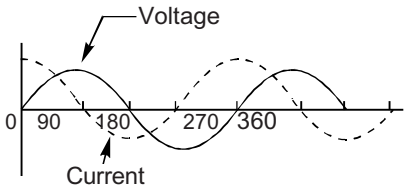


Fig. 4.47 Phase Between Current and Voltage in a Capacitive Circuit

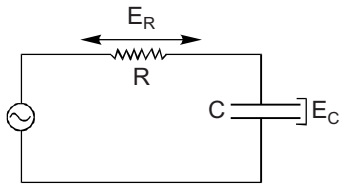


Fig. 4.48 RC Circuit

When an AC voltage is applied across a series combination of a resistor R and a capacitor C as shown in Fig. 4.48, the same current flows through R and C. There is a voltage drop E_R across the resistor R and a voltage drop E_c across the capacitor C. The current I flowing through the resistor R and the voltage drop E_R across the resistor are in phase. However, the same current I flowing through the capacitor C and the voltage drop E_c across it are not in phase. The current in this case leads the voltage E_c by 90°. These relations are shown in Fig. 4.49.

In order to find the total voltage drop across the combination of R and C, we cannot simply add E_R and E_c because these are not in phase. The resultant total

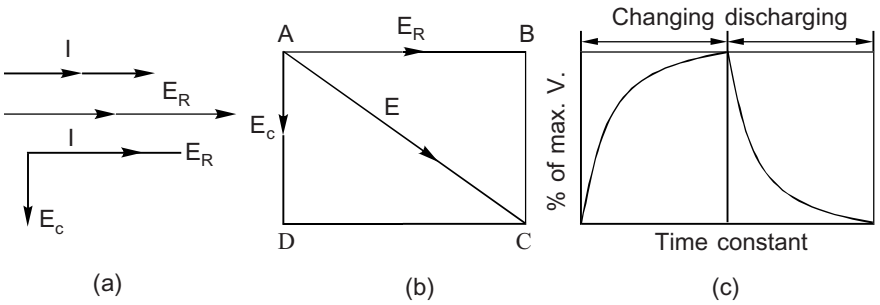


Fig. 4.49 Vector Addition of E_R and E_c

voltage drop E can be found by adding E_R and E_c vectorially as shown in Fig. 4.49(b). The resultant can be found either by completing the rectangle $ABCD$ and measuring the length of the diagonal AC or by the use of the formula:

$$E = \sqrt{E_R^2 + E_C^2}$$

Thus if E_R = 30 V and E_c = 40 V the total voltage drop will be

$$E = \sqrt{(30)^2 + (40)^2} = \sqrt{900 + 1600}$$

$$= \sqrt{2500} = 50 \text{ V}$$

and not merely 30 + 40 V = 70 V

An RC time constant reflects the relationship between time, resistance and capacitance. The time it takes a capacitor to charge and discharge is directly proportional to the amount of the resistance and capacitance. The time constant reflects the time required for a capacitor to charge to 63.2% of the applied voltage or to discharge by 63.2%.

Impedance When a current flows through the series combination of resistance R and the capacitance C, the total opposition offered by the combination to the flow of current is the impedance Z. Since the voltage developed across the resistor is proportional to the resistance R and the voltage developed across the capacitor is proportional to the impedance Z. The resistance R and the capacitive reactance X_c have the same phase relationship as the voltages E_R and E_c across them. To find the impedance Z, the resistance R and the capacitive reactance X_c must be added vectorially and the total impedance in the circuit Z found by the formula

$$Z = \sqrt{R^2 + X_C^2}$$

Since $X_c = 1/2\pi fC$ where f is the frequency of the applied voltage, the above formula becomes

$$Z = \sqrt{R^2 + \left(\frac{1}{2\pi fC}\right)^2}$$

The same result can also be obtained by the rectangle method as shown in Fig. 4.50.

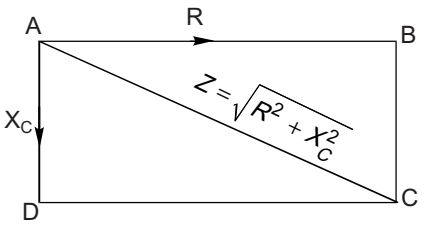


Fig. 4.50 Impedance of an RC Circuit

EXAMPLE: Find the current flowing through the circuit when a 10Ω resistance is connected in series with a 100 μF capacitor and an AC voltage of 120 V at 50 Hz is applied across the combination.

Here
$$X_c = \frac{1}{2\pi fC} = \frac{1}{6.28 \times 50 \times 0.0001} = 32\ \Omega \text{ approx.}$$
$$R = 10\ \Omega$$
$$Z = \sqrt{(10)^2 + (32)^2} = \sqrt{100 + 1024} = \sqrt{1124} = 33.5\ \Omega \text{ approx}$$
$$I = \frac{E}{Z} = \frac{120}{33.5} = 3.6A$$

4.36 LR, CR AND LC CIRCUITS

So far you have studied the properties of resistors (R), coils (L) and capacitors (C) and the behaviour of these components in circuits when used singly or in combination with each other. When resistor (R) and inductor (L) are used together, the combination is called an RL circuit. A circuit in which a resistor (R) and a capacitor (C) are combined together is called an RC circuit.

One of the most important combinations is, however, the circuit in which the inductance (L) and the capacitor (C) work together. This circuit is called the LC circuit and forms the basis of resonant circuits used in radio and TV receivers.

SUMMARY

Resistance of a material is the opposition to the flow of current. Factors that determine the resistance of a material are length, area of cross-section, nature of the material and its temperature. At a fixed temperature, the resistance R is given by the formula

$$R = \rho \frac{l}{A}$$

where ρ is the specific resistance, l the length and A is the area of cross-section. Specific resistance is defined as the resistance of a cubic centimeter of the material. Resistance of certain materials increases with temperature and in certain cases it decreases with temperature.

A resistor is an electronic component having a known value of resistance. There are two types of resistors, carbon resistors and wire-wound resistors A resistor can either be a fixed resistor or a variable resistor. Potentiometers and rheostats are variable resistors. The three ratings of a resistor are its ohmic

value, its tolerance and its wattage. In the case of small resistors, the ohmic value and tolerance are given by a colour code and the physical size indicates the wattage. The rating of wire-wound resistors are marked on the body of the resistor.

When resistors are connected in series the total value of the combination is the sum of the individual resistances. In a parallel combination the reciprocal of the total value is equal to the sum of the reciprocals of individual resistors. The wattage of a combination of resistances is always the sum of the individual wattages whether the resistances are connected in series, in parallel or in any other combination.

Kirchhoff's two laws—the Current Law and the Voltage Law—are helpful in solving complex circuits which cannot be easily solved by applying *Ohm's law*.

Inductance is the property of a coil that determines the amount of opposing emf produced in the coil due to a given change in the magnetic flux linkages. Inductance is a basic property of a coil in the same way as resistance is a basic property of a resistor.

The unit of inductance is henry but smaller units called millihenry and microhenry are also used. 1 H = 10^3 mH = 10^6 μ H.

Coils are classified either according to the core material used or according to the frequency of the input signal used. Thus there are air-core coils and iron-core coils. Similarly, there are audio frequency (AF) and radio frequency (RF) coils.

The counter emf induced in a coil does not allow the current to reach its maximum value at the same time as the applied voltage. The current lags the voltage by 90°.

The opposition offered by a coil to the flow of AC current is called inductive reactance. This is represented by X_L and its value is given by the formula $2\pi fL$.

The inductance of a circuit can be varied by connecting inductors in series and in parallel in the same way as resistors are concerned in series and in parallel.

The total voltage drop across an *RL* circuit is the vector sum of the individual voltage drops across the resistor and the inductor.

Transformers If two coils are placed near each other, the changing flux in one coil will cut the turns of the other coil and induce a voltage in it. This is the principle of transformer action.

The coil to which the input voltage is applied is called the primary winding and the coil in which voltage is induced is called the secondary winding. The ratio of the number of turns in the primary to that in the secondary is called the turns ratio. It is this turns ratio that determines the voltage induced in the secondary due to a given voltage applied to the primary. A transformer is a step up transformer or a step down transformer depending on whether the turns ratio is greater than or less than one.

The efficiency of a transformer is the ratio of the output power to input power. The efficiency is less than one due to losses in the transformer. Transformer losses are broadly classified as copper losses and the core losses. Copper losses are the I^2R losses and the core losses include losses due to flux leakage, eddy current losses and hysteresis losses.

Transformers are classified as power transformers, audio frequency (AF) transformers and radio frequency (RF) transformers.

Auto transformers are transformers which use the same coil to provide turns for the primary and the secondary windings.

A capacitor or a condenser is a device for storing electrical energy. The amount of electrical charge that a capacitor can hold depends on its electrical size. This electrical size of a capacitor is called its capacity or capacitance.

Farad is the unit of capacitance but smaller units of capacitance known as microfarad and micro microfarad (picofarad) are also in use. One farad is equal to 10^6 microfarads of 10^{12} micro microfarads.

Capacitance is directly proportional to the area of the plates and the dielectric constant and inversely proportional to the distance between the plates.

There are two main types of capacitors—the fixed capacitors and the variable capacitors. Fixed capacitors are available as mica capacitors, paper capacitors, ceramic capacitors and electrolytic capacitors. Examples of variable capacitors are gang capacitors and trimmer capacitors. Whereas the value of electrolytic and paper capacitor is either marked or stamped on the capacitor itself, the value of small mica and ceramic capacitors is indicated by colour coding.

Total capacitance increase when capacitors are connected in parallel and it decreases when capacitors are connected in series. The voltage rating of capacitors connected in series increases.

Opposition offered by a capacitor to the flow of AC current is called capacitive reactance and is measured in ohms. The capacitive reactance X_c is given by the formula $X_c = 1/2\pi fC$.

In a capacitive circuit the current leads the voltage by 90°. The series impedance of a capacitive circuit is found by the vector addition of the resistance and the capacitive reactance.

Review Questions

1. What is resistance? On what factors does the resistance of a wire depend?
2. What is specific resistance? A heater element of resistance 10 Ω is to be wound with nichrome wire having an area of cross-section of 1 mm^2. If the specific resistance of nichrome is 108 μΩ/cm^3, what length of wire will be required for the purpose? *Ans:* 926 cm
3. What is the effect of temperature on the resistance of a material? Name two materials that have negative coefficient of resistance.
4. What are the two types of resistors? Give the relative merits and demerits of each type.

5. How will the resistance of a wire vary if,
   (a) the length of the wire is halved,
   (b) the area of cross-section is doubled,
   (c) the specific resistance is made three times,
   (d) the temperature is raised.
6. Give the circuit symbols for a
   (a) fixed resistor,
   (b) variable resistor,
   (c) tapped resistor,
   (d) potentiometer.
7. Give the colour code for the following:
   (i) 100 k Ω ± 5%
   (ii) 470 Ω ± 10%
   (iii) 10 M Ω ± 20%
   (iv) 5 Ω ± 10%
8. What value of resistor is indicated by the following colour code:
   (i) Orange, orange, red, silver.
   (ii) Brown, green, green, gold.
   (iii) Orange, white, yellow, no band.
   (iv) Red, red, gold, gold.
9. Describe the use of a potentiometer as a voltage divider.
10. Select the correct answers.
    (a) In a series combination of resistors:
        (i) The voltage across each resistor is the same.
        (ii) The current flowing through each resistor is the same.
        (iii) Wattage of each resistor is the same.
    (b) In a parallel combination of two equal resistors:
        (i) The voltage across each resistor is the same.
        (ii) The current in each branch is the same.
        (iii) The total effective resistance R_T is double the value of each resistor.
        (iv) The total effective resistance is half the value of each resistance.
        (v) The total wattage of the combination is half the wattage of each resistor.
11. Define inductance of a coil. Give the unit of its measurement.
    Convert into millihenrys the inductance of a 0.01 H coil. (*Ans.* 10 mH)
12. What is counter emf? How does it depend on the value of the inductance?
13. What is an iron-core coil? Is it necessary for the core to be iron?
14. What is inductive reactance? How is it calculated?
    Find the inductive reactance of a 10 H coil at 500 Hz. What will the inductive reactance be at (a) 1000 Hz (b) 200 Hz?
    (*Ans*. 31,400 Ω (a) 62,800 Ω (b) 12,560 Ω)

15. What is the phase relationship between the current and voltage in an inductive circuit? Does the current lag behind or lead the voltage?
16. What is the series impedance of an *RL* circuit in which $R = 10\ \Omega$ and $X_L = 15\ \Omega$? (*Ans.* 18Ω)
17. What is transformer action and what is turns ratio? What is the output voltage of a transformer with a turns ratio of 5:1 when a voltage of 120 V is applied to the primary winding? (*Ans*. 24 V)
18. What is a step down transformer? State whether the transformer with the following turns ratio is a step down transformer or a step up transformer: (a) 1 : 5, (b) 40 : 1, (c) 1 : 1.
19. What are transformer losses? What are core losses?
20. What is the efficiency of a transformer? Why is the efficiency of a transformer always less than 100%? A mains transformer has one primary and two secondary windings. A current of 0.5 is drawn by the primary when a voltage of 200 V is applied. One of the secondary windings draws a current of 300 mA at 250 V and the other secondary draws a current of 2.5 A at 6.V. Calculate the efficiency of the transformer. (*Ans*. 90%)
21. What is an auto transformer? What is its weakness?
22. What is capacitance? What are the factors affecting the capacitance of a capacitor? If the distance between the plates of capacitor is increased, will the capacitance increase or decrease?
23. Name the unit of a capacitance and define it. Find the capacitance in farads of a 1000 μ F capacitor.
24. What is a variable capacitor? Give an example of a variable capacitor with a solid dielectric.
25. Give the formula for the total capacitance of two capacitors connected in series. A 50 μF 6 V capacitor is connected in series with a 100 μF 12 V capacitor. What is the total capacitance and voltage rating of the combination? (*Ans.* 33.3 μF, 18 V)
26. What is capacitive reactance? How is it calculated? Find the capacitive reactance of a 0.1 μF capacitor at 500 Hz and at 200 Hz? (*Ans*: 3184.7 Ω, 7961.7 Ω))

chapter 5

Resonance and Filter Circuits

5.1 RESONANCE

It has already been seen that in a circuit containing only inductance the voltage leads the current by 90° but in a circuit containing only capacitance the voltage lags behind the current by 90°. Thus we find that in one case the voltage leads the current and in the other case the voltage lags behind the current by the same phase angle of 90°. In other words, the phase angle between the inductive and capacitive voltages is 180° and the two voltages oppose each other.

In circuits where coils and capacitors are used together, the two more or less oppose each other. In some of these circuits the inductive effect is greater than the capacitive effect and the circuit will act like a coil and the voltage will lead the current. If the capacitive effect is greater than the inductive effect the circuit will behave like a capacitor and the voltage will lag behind the current. If, however, the inductive effect of the coil is exactly equal to the capacitive effect of the capacitor, the two effects will cancel each other out and the voltage and current will be in phase. Such a circuit is called a resonant circuit and the phenomenon is called *resonance*. Resonant circuits are quite important in Radio and TV receivers because these resonant circuits enable one to separate various stations and make the selection of the desired station possible.

5.2 TYPES OF RESONANT CIRCUITS

There are two types of resonant circuits—series resonant circuits and parallel resonant circuits. In series resonant circuits, the capacitor and the coil are connected in series, and in parallel resonant circuits the two components are connected in parallel. These resonant circuits will now be described in detail.

5.2.1 Series Resonant Circuits

Consider first the simple case of a coil L and a capacitor C connected in series.

An AC voltage E at a frequency F is applied across the combination as shown in Fig. 5.1(a).

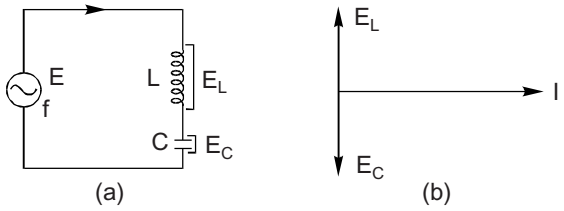


Fig. 5.1 Series Resonant Circuit

Since L and C are in series, the same current will flow through L and C. The voltage E_L developed across L will be leading the current I by 90°. The two voltages will be opposed to each other as shown in Fig. 5.1(b). If the values of L and C are so chosen that the inductive reactance X_L is equal to the capacitive reactance X_C at the given frequency, the circuit will neither be inductive nor capacitive but will behave like a resistance. The value of the impedance $(X_L - X_C)$ or the reactance will be theoretically zero and a very high current will flow through the circuit. The circuit is then said to be *resonant* at this frequency where $X_L = X_C$. However, in actual practice a coil will always have some resistance R and the current flowing at resonance will be limited by the value of resistance R.

5.3 FREQUENCY OF RESONANCE

X_L $(2\pi fL)$ increases with frequency and X_C $(1/2\pi fC)$ decreases with frequency. At a certain frequency the two reactances X_L and X_C will be equal. This is the frequency of resonance or the *resonant frequency*.

At resonance $$X_L = X_C$$

or $$2\pi f_0 L = \frac{1}{2\pi f_0 C}$$

where f_0 is the frequency of resonance

$$f_0^2 = \frac{1}{4\pi^2 LC}$$

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

In this formula f_0 will be in hertz when L is in henrys and C is in farads. This formula helps us to make some important calculations regarding resonant circuits. In the circuit given, if

$$L = 100 \text{ mH} = 0.1 \text{ H}$$

$$C = 1\mu\text{F} = 10^{-6} \text{ F}$$

we can calculate the resonant frequency f_0 by substituting the values of L and C in the formula

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

$$f_0 = \frac{1}{2\pi\sqrt{0.1\times10^{-6}}}$$

$$f_0^2 = \frac{1}{4\pi^2 \times 0.1\times10^{-6}} = \frac{1}{39.44\times0.1\times10^{-6}}$$

$$= 253550$$

$$f_0 = \sqrt{253550} = 503 \text{ Hz}$$

Thus the frequency of resonance is about 500 Hz.

This formula also helps us to select suitable values of components required for resonance at a particular frequency as in the following example.

EXAMPLE: What is the value of the capacitor required to be connected in series with a 10 H choke coil so that the combination resonates at 50 Hz.

In the formulae $f_0 = \frac{1}{2\pi\sqrt{LC}}$

$f_0 = 50$ Hz $\quad L = 10$ H, $C = ?$

$$50 = \frac{1}{2\pi\sqrt{10\times C}}$$

$$(50)^2 = \frac{1}{4\pi^2 \times 10\times C}$$

$$2500 = \frac{1}{4\times9.86\times10\times C}$$

$$C = \frac{1}{2500\times40\times10} = \frac{1}{1000000}$$

$$= 1\times10^{-6} \text{ F} = 1\mu \text{ F}$$

So a 10 H choke coil and a 1 μ F capacitor will resonate at a frequency of 50 Hz.

5.4 LCR RESONANT CIRCUITS

As mentioned earlier, a pure inductance or coil (without resistance) is not obtainable in practice. Any coil will have some resistance depending upon the size of the wire used. Even the capacitance and the connecting leads in an LC circuit will have resistance. So any resonant LC circuit is actually an LCR resonant circuit in which resistance R plays an important part. Let us study the properties of an LCR resonant circuit of the type shown in Fig. 5.2. Since R, L and C are in series, the impedance of the circuit will be given by the formula

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

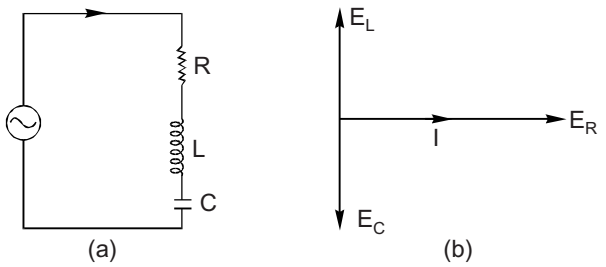


Fig. 5.2 LCR Resonant Circuit (a) Actual Circuit (b) Vectorial Representation

If the circuit is resonant to the frequency of the applied voltage E, then $X_L = X_C$ and the impedance of the circuit $Z = \sqrt{R^2 + O^2} = R$

The circuit is resistive and the current I is given by Ohm's law

$$I = \frac{E}{Z} = \frac{E}{R}$$

The voltages across L and C are equal and opposite in phase and they cancel out leaving only the resistance R in the circuit. A vectorial representation of this effect is given in Fig. 5.2(b).

In the circuit shown, if $R = 120\ \Omega$ is in series with a 100 mH coil and a 1 μF capacitor, the resonant conditions can be examined for a 240 V, 500 Hz generator connected across the RLC combination.

$$X_L = 2\pi fL = 6.28 \times 500 \times 0.1 = 314\Omega$$

$$X_C = \frac{1}{2\pi fC} = \frac{1}{6.28 \times 500 \times 0.000001} = 318\Omega$$

If the values of X_L and X_C calculated above are rounded off to a common value of 316 Ω, we find that $X_L = X_C$. The circuit is, therefore, resonant at this frequency. X_L and X_C being of opposite phases, if we call X_L positive, X_C will be given the negative sign (–) and the impedance of the circuit will be

$$Z = \sqrt{(120)^2 + (316 - 316)^2} = 120\ \Omega$$

Actually, the resistance 120 Ω may be the sum of the resistance of the coil (10 Ω) and the resistance (110 Ω) placed externally in the circuit. These resistances are generally lumped together and shown as one resistance R for calculation purpose.

$$\text{Current } I = \frac{E}{Z} = \frac{240V}{120\Omega} = 2\,A$$

Voltage across the coil (E_L) = $I \times X_L = 2 \times 316 = 632$ V

Voltage across capacitor (E_C) = $L \times X_C = 2 \times 316 = 632$ V.

These two voltages are equal but opposite as shown in Fig. 5.2 (b).

It is important to notice that the voltage developed across the coil (E_L) and the voltage developed across the capacitor (E_C) are both several times greater than the applied voltage E. This phenomenon is called resonant voltage step-up.

The lower the value of resistance in the circuit, the higher is the voltage step up. In a practical resonant circuit, we have only a coil and a capacitor in series and the only resistance in the circuit will be the resistance of the coil.

As $Q = X_L/R$, Q will be higher if R is low. There will be more resonant voltage step up in a high Q coil than in a low Q coil. To sum up, in a series resonant circuit,

(i) there is low impedance and high current.
(ii) the voltage across the coil and the capacitor is several times the value of the source voltage.

5.5 RESONANCE CURVES

A certain combination of L and C resonates at a particular frequency when the capacitive reactance becomes equal to the inductive reactance. At this frequency the impedance of the circuit is minimum and the current flowing is the maximum.

When the frequency is lower than the frequency of resonance the capacitive reactance is greater than the inductive reactance. The circuit will then behave like a capacitive circuit with current leading the voltage. If, however, the frequency of the applied voltage is higher than the resonant frequency the inductive reactance is greater than the capacitive reactance and the circuit behaves like a coil and the voltage will lead the current. In either case the impedance of the circuit will be minimum only at resonance. It will be higher both above and below the resonant frequency. Accordingly, the current is maximum at resonant frequency and is less at frequencies higher and lower than the resonant frequency. The variation of current with frequency is shown in Fig. 5.3. Such a curve is called a resonance curve.

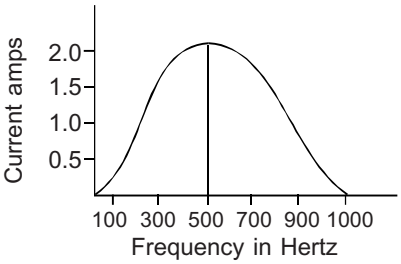


Fig. 5.3 Resonance Curve

Changing the resistance has an important effect on the shape of the resonance curve. If the Resistance is high, Q of the circuit is low and the curve is broad. Series resonant circuits are used to select one frequency and reject all other frequencies. A high Q circuit is more selective than a low Q circuit.

L and C can be varied in a number of ways to get resonance at a particular frequency. The ratio of inductance L to the capacitance C is called the L to C ratio. In a series resonant circuit a high L to C ratio will give a sharper resonance curve than a low L to C ratio.

A series resonant circuit is also known as an *acceptor circuit*.

5.6 PARALLEL RESONANT CIRCUITS

In a parallel resonant circuit, the inductance L and the capacitance C are connected in parallel and the same voltage appears across both L and C as shown in Fig. 5.4.

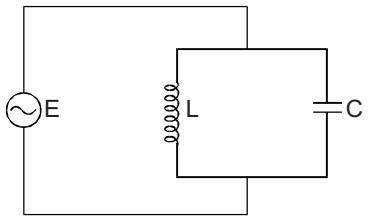


Fig. 5.4 Parallel Resonant Circuit

Resonance will occur when the inductive reactance X_L is equal to the capacitive reactance X_C

or $$X_L = X_C$$

For fixed values of L and C, the resonance will occur at a frequency given by the formula

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

This formula is the same as that discussed in the case of series resonance.

The behaviour of a parallel resonant circuit is exactly opposite to that of a series resonant circuit. Whereas the impedance of a series resonant circuit is minimum at resonance and the line current high, the impedance of a parallel resonant circuit is maximum at resonance and the line current minimum. Since the coil and the capacitor in a parallel resonant circuit are connected in parallel, a current I_L will flow through the inductive branch and a current I_C will flow through the capacitive branch. The values of these currents depend upon the inductive and capacitive reactances of the two branches. At resonance these two currents are theoretically equal. In the capacitive branch the current leads the voltage and in the inductive branch the current lags behind the voltage by 90°. The currents will be 180° out of phase with each other. In other words, the currents in the inductive and capacitive branches flow in opposite directions. This means that during half the cycle the capacitor discharges through the coil and the electric energy is stored in the coil in the form of a magnetic field during the other half of the cycle, the coil releases this electrical energy in the form of current and charges the capacitor with this current. Thus the coil and the capacitor pass current back and forth between them inside the resonant circuit. This is called the circulating current.

A coil has some resistance and the connecting leads also have some resistance. This resistance is kept very low. There can be very high currents flowing back and forth between the coil and the capacitor. There is always some loss of energy in the form of heat because of this resistance. This loss will be made up by a low current that will be supplied by the generator. This current is called the line current.

5.7 BANDWIDTH OF A RESONANT CIRCUIT

In a resonant LC circuit the maximum effect is produced at the resonant frequency *fr*. However, other frequencies close to the resonant frequency are also effective. There is a band of frequencies above and below the resonant

frequency which also produce some resonance effects. The width of the resonant band of frequencies, centered around *fr*, is called the bandwidth of the resonant circuit, Fig. 5.5(a) represents a series resonant circuit whose resonance curve is shown in Fig 5.5(b).

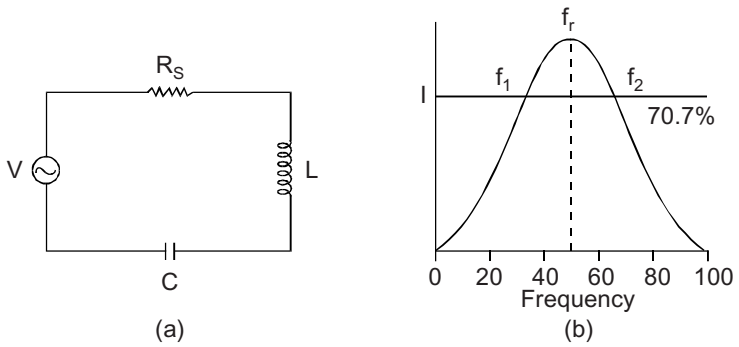


Fig. 5.5 (a) Resonant Circuit (b) Resonance Curve

At frequencies f_1 and f_2 the response of the circuit is 70.7% of the maximum response at the resonant frequency *fr*. If the bandwidth is represented by Δf, then

$$\text{Bandwidth } \Delta f = f_2 - f_1$$

Circuit with high Q have a sharp resonance curve and the bandwidth is low but circuits with low Q are less sharply lined and have a broader bandwidth. There is, thus, a relationship between the bandwidth Δf and the Q value of a tuned circuit. This relationship is given by the formula

$$\Delta f = \frac{fr}{Q}$$

or $$\Delta f \cdot Q = fr$$

When stated in word the relationship will mean that the product of bandwidth and the Q of a resonant circuit is equal to the resonant frequency of the circuit. The bandwidth is expressed in the same units as the resonant frequency. It is generally written as BW.

EXAMPLE: Find the Q of a circuit which is resonant at 1000 kHz and has a BW of 10 kHz.

In this formula

$$\Delta f\, Q = fr$$

Or $$Q = \frac{fr}{\Delta f}$$

If fr = 1000 kHz, BW = Δf = 10 kHz

$$Q = \frac{1000 \text{ kHz}}{10 \text{ kHz}} = 100$$

Thus the Q of the circuit is 100.

5.8 TUNING

Tuning is the process by which an *LC* circuit is made to resonate to different frequencies over a given range of frequencies by varying either *L* or *C*. When the resonance is obtained by the variation of capacitance *C* it is called capacitive tuning and when the inductance *L* is varied to obtain resonance it is called *inductive tuning*.

A common example of capacitive tuning is the tuning of a radio receiver to different stations whose carrier frequencies are marked on the dial of the radio receiver. The tuning is done by the variation of the capacitance *C* of a variable air capacitor for a fixed value of the inductance of the coil *L*. Figure 5.6 shows

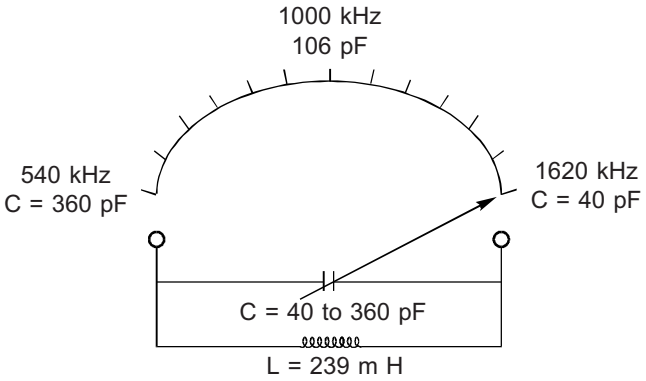


Fig. 5.6 Medium Wave Tunning

how an *LC* circuit is turned over the medium wave band of 540 to 1620 kHz for a fixed value of *L*(239 μH), the value of *C* at any frequency of resonance is calculated from the formula

$$f = \frac{1}{2\pi\sqrt{LC}}$$

It may be noted that the lowest frequency of 540 kHz corresponds to the highest value of *C* = 360 pf and the highest frequency of 1620 kHz is turned when the capacitor is at its lowest value of 40 pf.

The tuning ratio which is the ratio of the highest frequency to the lowest frequency is inversely proportional to the square root of the ratio of lowest value of *C* to the highest value of *C*.

In this case

$$\text{Tuning ratio} = \frac{\text{max frequency}}{\text{min frequency}} = \frac{1620\,\text{kHz}}{540\,\text{kHz}}$$

$$= \frac{1}{\sqrt{\frac{C_{min}}{C_{max}}}} = \frac{1}{\sqrt{\frac{40\text{pf}}{360\text{pf}}}} = \sqrt{9}$$

$$= \frac{1}{\sqrt{\frac{1}{9}}} = \sqrt{\frac{9}{1}} = \frac{3}{1}$$

EXAMPLE: The tuning ratio of frequency in an *LC* turned circuit is 2:1. Find the ratio of the variable *L*.

As stated above, the tuning ratio in frequency is inversely proportional to the square root of the ratio of the lowest value of *L* to the highest value of *L*.

$$\text{Turning ratio} = \frac{2}{1} = \frac{1}{\sqrt{\frac{L_{\min}}{L_{\max}}}} = \sqrt{\frac{L_{\max}}{L_{\min}}}$$

$$\therefore \qquad \frac{L_{\max}}{L_{\min}} = \left(\frac{2}{1}\right)^2 = \frac{4}{1} = 4:1$$

Thus the ratio of the value *L* in this case is 4:1.
This ratio will also be the ratio of the variable *C*.

5.8.1 Slug Tuning

Slug tuning or permeability tuning is the tuning in which the value of *L* or inductance of the *LC* tuned circuit is varied by moving a ferrite core slug up and down the winding of the coil. Slug tuning is used in the tuning of IF transformers (IFTs) in a radio transistor circuit.

5.8.2 Electronic Tuning

In this type of tuning use is made of the properties of a varactor diode (varicap) whose junction capacitance varies inversely as the reverse bias applied across the junction of the diode. The varactor diode is connected in parallel with a capacitor of the *LC* circuit and the DC reverse bias is applied through a variable resistor (pot) to provide a control for fine tuning, Fig. 5.7 shows the circuit for electronic tuning. DC reverse bias is applied through the resistor R_3 from a DC battery. R_2 is an isolating resistance and R_1 is the resistance that varies the

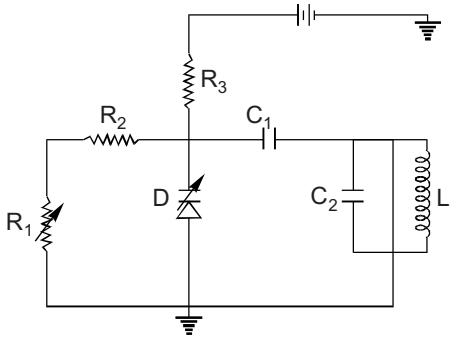


Fig 5.7 Electronic Tuning with a Varactor Diode

reverse bias on the diode D and hence serves as the fine tuning control. C_1 blocks the DC supply but it is almost a short at the high frequency of operation and the diode capacitance is virtually in parallel with main tuning capacitor C_2. At very high frequencies of operation capacitor C_2 is not required and the capacitance of the varactor diode is enough to provide the necessary capacitance for tuning.

This type of tuning is often used in the RF tuned circuits of a TV receiver'.

5.9 FILTER CIRCUITS

Filters are devices meant to separate out components that are mixed together. Filters are either mechanical or electrical. Mechanical filters separate particles from liquid or small particles from large particles. Electrical filters separate out components of different frequencies. Electrical or Electronic filters will be described in this section.

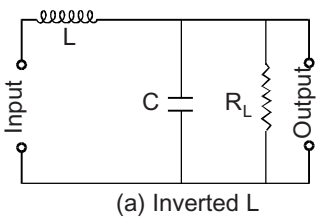
Electronic filters make use of frequency sensitive components L and C because of the opposite frequency characteristics. X_L increase but X_C decreases with higher frequencies. Depending on the filtering action required these components are arranged in different configurations, for separating audio frequencies from radio frequencies or vice versa. The most common types of configurations are L, T and π types as described below. Resistance R_1 in this configuration represents the load resistance in which the filtered frequencies develop the required voltage.

5.9.1 Classification of Filters

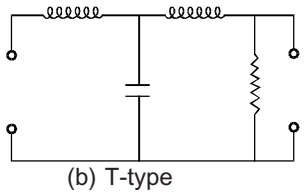
Depending on their functions, the filter circuits can be broadly divided into three types viz low-pass filters, high-pass filters and band-pass filters as described below:-

1. *Low Pass Filters:* A low pass filter allows the lower frequency components to pass through and develops the output voltage across the load resistance R_L. Three types of configuration for low pass filter are shown below.

With the applied input voltage having different frequency components, the low pass filter action results in max low frequency voltage across R_L while most of the high frequency voltage is developed across the series choke L or the resistance R. In Fig. 5.8(a) the use of both the series choke and by pass capacitor improves



(a) Inverted L


(b) T-type

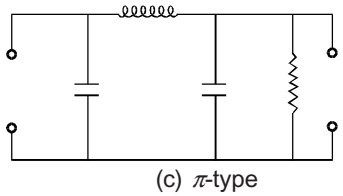

(c) π-type

Fig. 5.8 Low Pass Filters Circuits

filtering action by providing sharper cut off between the low frequencies that can develop voltage across R_L and the higher frequencies are stopped from the load by producing max voltage across R_L. Similar action takes place in t-type filter in Fig. 5.8(b) and the II-type filter in Fig. 5.8(c).

Attenuation This is the ability of the filter to reduce the amplitude of the undesired frequencies.

Cut-off Frequency: The frequency at which the attenuation reduces the output of the filter to 70.7% of the response is the cut-off frequency.

Pass Band and Stop Band In a low pass filter, all frequencies that lie below the cut-off frequency are in the pass band of the filter and frequencies above the cut-off frequency constitute the stop band for the filter. Figure 5.9 below shows the response of the low pass filter with a cut off frequency of 15 kHz.

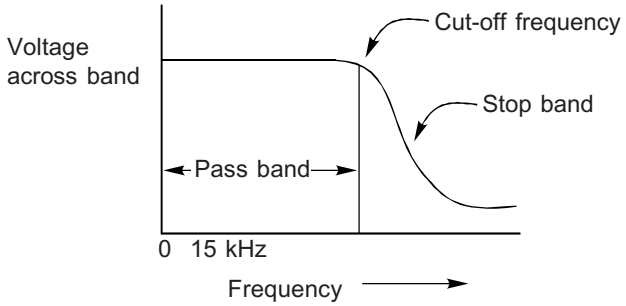


Fig. 5.9 Frequency Response of a Low Pass Filter Reducing the Cut Off Frequency (15 kHz) and the Pass Band and Stop Band of the Filter

High Pass Filter A high pass filter allows all frequencies above the cut off frequencies and cannot develop appreciate voltage across the load. A high pass filter circuit of T-type is shown in Fig. 5.10(a). In this the high frequency components of the input voltage can develop very little voltage across the R_L series capacitors C_1 and C_2, allowing most of the voltage to develop across the load R_L. The inductance across the line has higher reactance with increasing frequencies, allowing the shunt impedance to be not lower than the load resistance R_L. For lower frequencies R_L is effectively short circuited by the lower inductive reactance across the line.

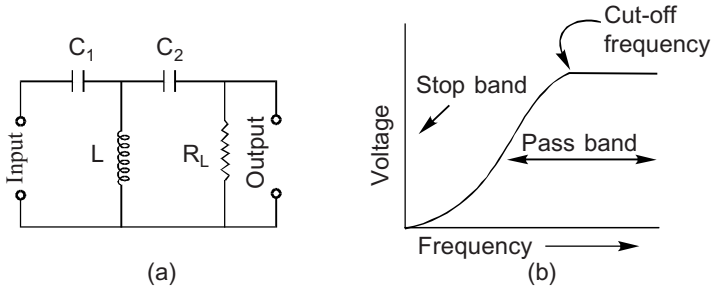


Fig. 5.10 High Pass Filter

Figure 5.10 (b) gives the response of the high pass filter with the cut-off frequency and the pass band and stop band indicated there on.

Band Pass Filters A high pass filter combined with a low pass filter to form a band pass filter is shown in Fig 5.11. In this case the frequencies between the cut off frequencies of high pass filter and the cut off frequencies of the low pass filter form the pass band of the band pass filter. Band pass filters are useful at radio frequencies for filtering a certain band of frequencies. A tuned circuit provides a convenient method of band pass filtering. The pass band depends on the Q-value of the tuned circuit.

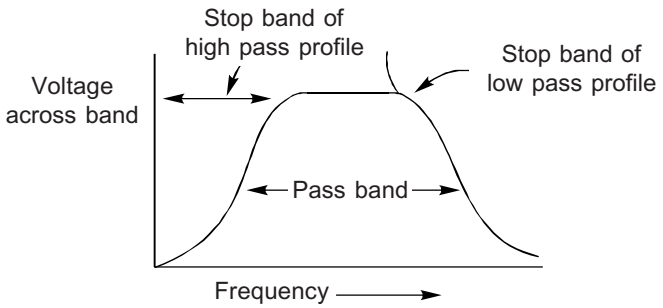


Fig. 5.11 High Pass Filter

Constant K Filter In a filter circuit (say of L-Type) the values of inductance and capacitance can be designed to make the product of X_L and X_C constant at all frequencies. This enables the filter circuit to present constant impedance at the input and output terminals. A constant *K* filter can be high pass or low pass.

M-Derived Filter This is a modified form of constant *K* filter. The ratio of the filter cut off frequency to the frequency of infinite attenuation is called the m factor of the filter circuit. The m-derived filter can be a high pass or a low pass, filter. Such a filter has the advantage of providing a very sharp cut-off.

Summary

Resonance When the inductive reactance of a circuit equals its capacitive reactance, the two cancel each other out and the circuit behaves like a resistance. The circuit is said to be a resonant circuit and the phenomenon is called resonance.

There are two types of resonant circuits—series resonant circuits and parallel resonant circuits. In a series resonant circuit the impedance is minimum at resonance and the current is maximum. In a parallel resonant circuit, the impedance is maximum and the current is minimum. The frequency of resonance in both the cases is given by the formula

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

Resonant circuits are extensively used in radio and TV receivers.

A relationship exists between the bandwidth Δf and the Q value of the tuned circuit and is given by the formula

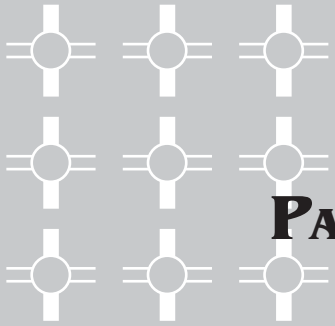
$$\Delta f = \frac{fr}{Q}$$ where fr is the resonant frequency of the circuit.

Tuning Tuning is the process by which an LC circuit is made to resonate over a given range of frequencies. Methods commonly used for tuning are capacitor tuning, slug tuning and electronic tuning.

Filter Circuits Filters are devices meant to separate out components that are mixed together. Electrical filters separate out components of different frequencies. Filters are classified as low pass filters, high pass filters and band pass filters.

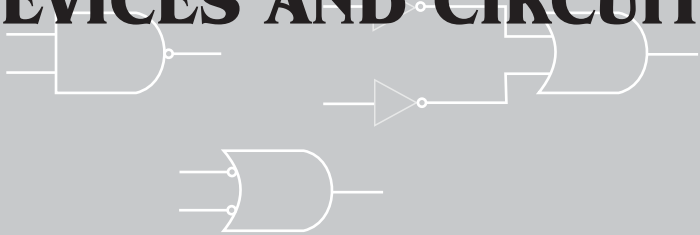
Review Questions

1. What is resonance and what is resonant frequency?
2. What is the formula for calculating the frequency of resonance of an LC circuit. A 20 H choke coil and a capacitor are to the resonated at 50 Hz. What is the capacitance of the capacitor required (ans 0.5 μf)?
4. How is series resonance different form parallel resonance.
5. Describe some commonly used methods for tuning.
6. What is a filter circuit? Give the classification of filter circuits.
7. Describe a band pass filter.



PART II

SEMICONDUCTOR DEVICES AND CIRCUITS



chapter 6

Semiconductor Diodes and Transistors

6.1 INTRODUCTION

Semiconductor diodes and transistors are solid-state devices which have gained tremendous importance in every field of the electronic industry. Transistors have brought about a complete transformation in electronic technology. New equipments and gadgets have been made possible and old ones made more compact, reliable and economical. A transistor can perform almost the same functions as an electronic tube or a valve. However, because of the many advantages possessed by a transistor over its counterpart, the electronic tube; the former has almost replaced completely the latter in all branches of electronics except those applications requiring very high output powers (above 1 kW) and high frequency operation. Some of the advantages possessed by transistors compared to vacuum tubes are: (a) small size and rugged mechanical construction; (b) No filament or heating power required and less heat formation; (c) easy operation in any position and a very long life. Basic to the study of transistors is the study of semiconductor materials.

6.2 SEMICONDUCTORS

A semiconductor is a material whose electrical properties lie between those of a conductor and an insulator. Germanium and silicon are the most commonly used semiconductors. In their pure crystalline form, both germanium and silicon are insulators or bad conductors of electricity. However, when small quantities of certain impurities are added to these crystals, their electrical properties are modified and they become partial conductors or semiconductors. This process of adding controlled quantities of certain impurities to the pure crystals of germanium and silicon is known as *doping*.

6.3 N-AND P-TYPE SEMICONDUCTORS

An atom of silicon (or germanium) has four electrons in its outermost shell called valence electrons as shown in Fig. 6.1.

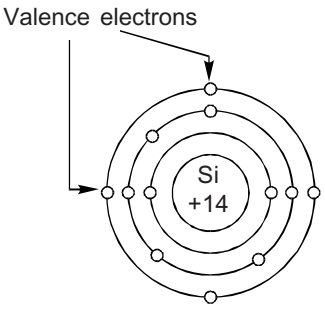


Fig. 6.1 Silicon Atom

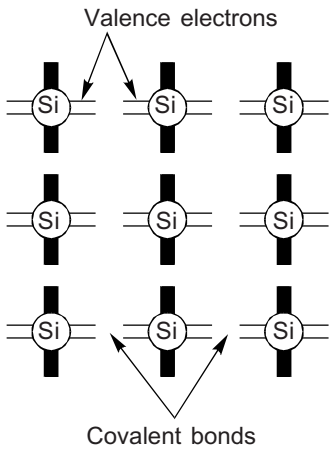


Fig. 6.2 Silicon Crystal Lattice Showing Electron Pairs or Covalent Bonds

In their pure crystalline form or intrinsic form, as it is called, the atoms of the semi-conductor materials like silicon arrange themselves in a crystal lattice structure. In a crystal lattice, each atom shares its outermost or valence electrons with those of an adjacent atom forming what are known as an electron-pair or covalent bonds between the atoms as shown in Fig. 6.2. With these covalent bonds, the electron pairs are so strongly bound to each other and to the nucleus, that no free electrons are easily available to allow any conduction of current through the material which in its intrinsic or pure form behaves as a nonconductor of electricity.

When silicon in its pure form is doped with a pentavalent (five electrons in the outermost shell) atoms like arsenic or antimony, four of its five valence electrons form covalent bonds with the valence electrons of four adjacent silicon atoms, but the fifth valence electron of arsenic remains unattached and becomes a free electron (Fig. 6.3). Thus, a silicon (or germanium) crystal doped with arsenic (or antimony) develops an excess of free electrons and is called an N-type semiconductor. Arsenic which donates its electrons to silicon is called a "donor" atom or 'donor' impurity.

If silicon is doped with a trivalent (three electrons in the outermost shell) atom like indium, the three valence electrons of indium form covalent bonds with the valence electrons of three adjacent silicon atoms but one valence electron from the adjacent silicon is unable to form an electron pair bond. As a result one electron is missing or an electron valency exists in the crystal lattice of silicon. This electron vacancy or absence of an electron is called a hole. The same conductor so formed has a deficiency of electrons or excess of holes and

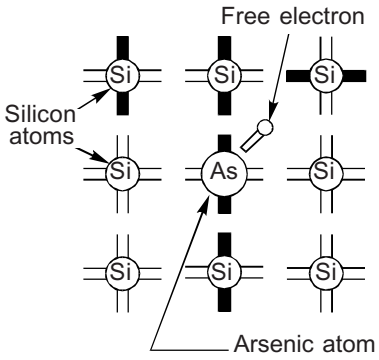


Fig. 6.3 Formation of N-type Semi-conductor (Silicon 'Doped' with Arsenic)

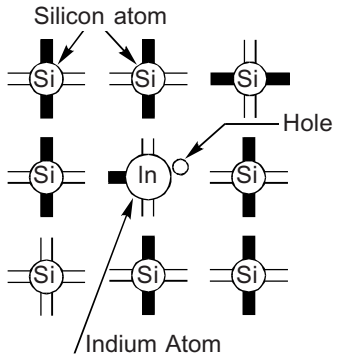


Fig. 6.4 Formation of 'Holes' in a P-type Semiconductor (Silicon 'Doped' with Indium)

is called a P-type semiconductor (Fig. 6.4). Indium is called an acceptor atom or acceptor impurity because it can accept an electron from an adjacent covalent bond.

Current Flow through Semiconductors An N-type semiconductor has an excess of free electrons. These free electrons act as current carriers when an electric field or a voltage is applied across an arsenic doped silicon crystal.

In the case of P-type semiconductors, the 'holes' indicating absence of electrons behave like positively charged particles when an electric field is applied across the crystal. Under the influence of the field an electron from an adjacent electron pair bond breaks loose and falls into a hole near the positive pole of the battery. This creates a new hole that can accept another electron which has broken loose from its electron pair bond. This process continues and constitutes a movement of electrons towards the positive pole of the battery and the movement of holes towards the negative terminal of the battery. As the holes reach near the negative terminal, electrons from this terminal enter the crystal and neutralise the holes. At the same time, the loosely held electrons that filled the holes are pulled away by the positive terminals thereby creating new holes. The movement of holes in one direction and the movement of electrons in the opposite direction, however, constitutes a current flow in the same direction. Thus, in a semiconductor, the current conduction is the result of the movement of holes inside the crystal and the movement of electrons through the external circuit and the battery.

6.4 P-N JUNCTION DIODE

A P-N junction diode is formed by combining a P-type semiconductor with an N-type semiconductor. The P-N junction so formed exhibits the interesting and useful property of offering a low resistance to current flow in one direction. The P-N junction has the same rectifying characteristics as a vacuum diode and is widely used as a detector in electronic circuits including radio and TV receivers.

A P-N junction cannot be formed by simply putting together a P-type and an N-type semiconductor. The construction of a P-N junction diode involves special manufacturing techniques of various types. In one method, a single crystal is grown from a melt which contains impurities of one kind at the start but in the middle of the process impurities of the opposite kind are added to the melt in sufficient quantities, so that the type of the material grown thereafter is changed. This is the grown junction diode. In the second method, a small 'dot' of indium is fused on a germanium immediately below the surface resulting in the formation of a P-N type junction between the P-region and the body of the N-type germanium.

Figure 6.5 shows the constructional details of a point-contact diode used as a small signal device in radio, TV and tape recorders and other electronic circuits.

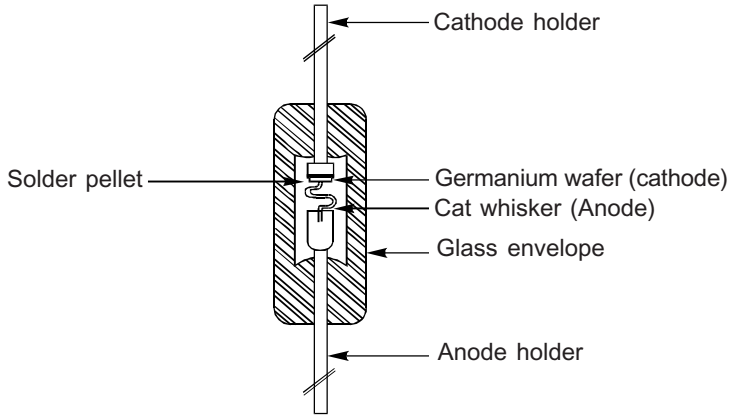


Fig. 6.5 Construction of a Germanium Point Contact Diode (Courtesy: BEL)

6.5 OPERATION OF A SEMICONDUCTOR DIODE

Figure 6.6(a) is a representation of a P-N junction diode and Fig. 6.6(b) is its circuit symbol.

In the circuit symbol, the terminal marked 'anode' is connected to the P-type material while that marked 'cathode' is connected to the N-type material. The arrow head also indicates the direction of conventional current flow in the diode. The cathode end of a diode is usually marked by a circular band or by a plus (+) sign as in Fig. 6.6(c).

The operation of a P-N junction diode can be explained by the fact that there is a movement of holes and electrons even when no external voltage is applied across the junction. Electrons and holes in the junction area combine leaving certain ions uncovered or unneutralized. These uncovered ions set up a field called the barrier field or barrier potential. This barrier potential drives the remaining electrons and holes away from the junction, which in turn uncovers more ions and so on. This barrier potential is equivalent to a small battery

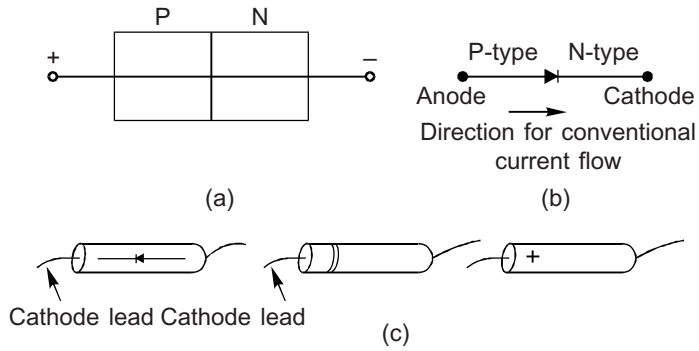


Fig. 6.6 Identifying Cathode Lead (a) Arrow Pointing Towards Cathode (b) Cathode Marked by Circular Band (c) Cathode Marked by + Sign

across the junction as in Fig. 6.7. This barrier potential depends on the nature of the semiconductor material and is 0.7 V in the case of silicon and 0.3 V in the case of germanium as in Fig 6.7.

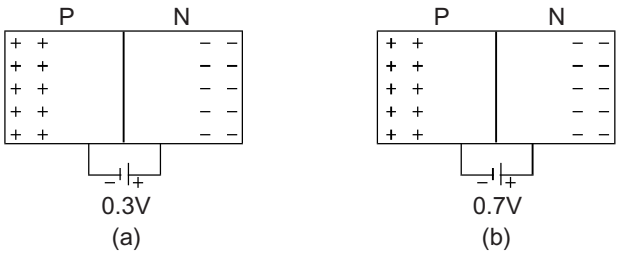


Fig. 6.7 Barrier Potential (a) Germanium Barrier Potential (b) Silicon Barrier Potential

For any current to pass through the P-N junction, the barrier potential must be overcome by application of some external voltage or battery to oppose the barrier potential. This is called biasing the P-N junction diode.

6.6 FORWARD BIAS AND REVERSE BIAS

A P-N junction is said to be forward biased if an external battery is connected across the junction so that the polarity of the external battery is opposite to the barrier potential as in Fig. 6.8. This lowers the barrier potential and allows easy flow of current through the diode as explained below.

Free electrons from the negative terminal of the battery repel the free electrons in the N-type material and these electrons move towards the P-N junction. The holes in the P-type material are also repelled by the positive terminal of the battery and move towards the junction. At the junction the free electrons and holes combine and are lost in the process. However, the current carriers lost in these combinations are replenished by new current carriers resulting from the separation of electron hole pairs. The free electrons produced in the

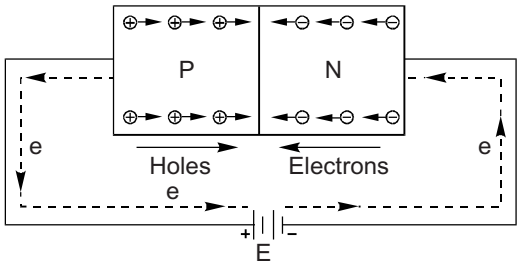


Fig. 6.8 Forward Bias

P-type material are attracted by the positive terminal of the battery and flow in the external circuit as shown in Fig. 6.8. This is a continuous process and constitutes a current flow by electrons in the external circuit but inside the diode both holes and electrons carry currents.

The increase of current with the increase of applied voltage in the case of a silicon diode is indicated graphically by its volt-ampere (*V-I*) characteristics as in Fig. 6.9. It will be noted from the graph that current increase is very little with increase of forward bias below 0.7 V. For voltage equal to or above 0.7 V the diode permits flow of current and even a slight increase in forward bias results in a large increase of current. The silicon diode is then said to have a low forward resistance.

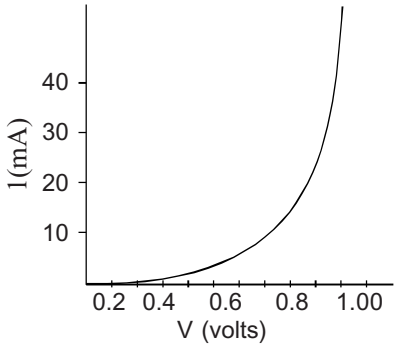


Fig. 6.9 Volt-ampere Characteristics of a Silicon Diode

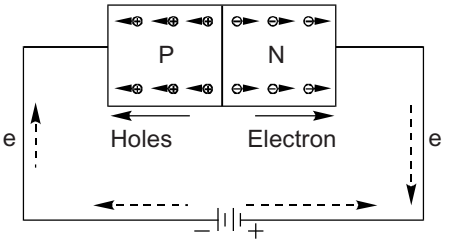


Fig. 6.10 Reverse Bias

A P-N junction diode is said to be *reverse biased* when the external voltage or battery connected across the P-N junction aids the barrier potential as in Fig. 6.10. Since the barrier potential is actually raised by the reverse bias, there is practically no flow of current through the diodes.

The free electrons in the N-type material are attracted away from the P-N junction and the holes in the P-type material are similarly attracted away from the P-N junction by the negative terminal of the battery and there are practically no holes or electron carriers left in the neighbourhood of the P-N junction and the current flow stops completely. However, a minute current of the order of a few microamperes still flows through the diode as a result of the minority

carriers liberated due to the thermal energy in the crystal. The reverse current remains low with increase of reverse bias till the breakdown voltage is reached when the current suddenly shoots up. This phenomenon is called *Zener action* of the diode and the current is known as a *Zener current*. The breakdown voltage when the current suddenly shoots up is the Zener voltage.

Zener Diodes Zener diodes are silicon diodes which are designed for a specific reverse breakdown voltage. Figure 6.11 shows the typical current-voltage characteristic of a zener diode.

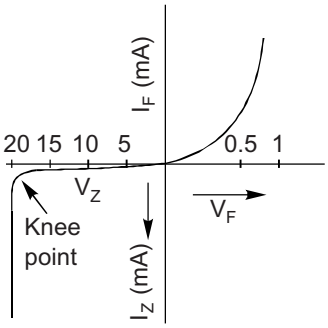


Fig. 6.11 Characteristics of a Zener Diode

With forward bias the Zener diode behaves like a closed switch and the forward current increases with increase of forward bias. When, however, the diode is reverse biased, a small reverse current flows and remains practically constant with an increase of reverse bias till the Zener voltage V_Z is reached. When this happens the reverse current abruptly increases to a very high value as a result of the covalent bonds near the junction breaking down and releasing a large number of electron hole pairs in the form of an avalanche.

Zener diodes are widely used in electronic circuits as voltage regulators and voltage reference standards. When the *avalanche-current* flows, the voltage across the diode remains constant. Figure 6.12 (a) shows the circuit symbol for a Zener diode and the use of a Zener diode as a voltage regulator is shown in Fig. 6.12 (b).

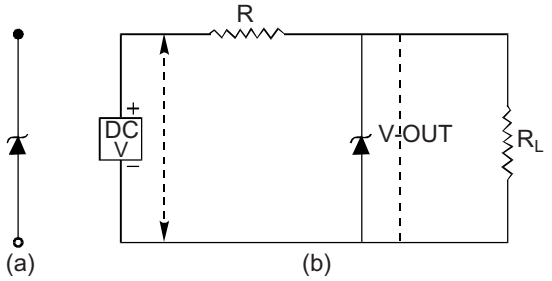


Fig. 6.12 Zener Diode in Circuit (a) Circuit Symbol of a Zener Diode (b) Zener Diode as a Shunt Voltage Regular

The Zener diode is connected in parallel with the load resistance to keep the output voltage (V_{out}) across the load R_L constant within specific limits. The variations of voltage may either be due to the change in DC supply voltage (V_{in}) or the changes in load resistance or load current. The value of R depends on design considerations, such as constant output voltage (V_{out}) required and the load current variations.

The ratings of the Zener diode and its specifications provided by the manufacturers include Zener voltage, Zener current limits, maximum power dissipation, maximum operating temperatures, maximum Zener impedance, reverse leakage current and the nature of the material like silicon from which the Zener diode is manufactured.

Besides the Zener diode, there are a number of other special types of diodes which find useful applications in electronic circuits. Some of these diodes are briefly described below.

6.7 OTHER TYPES OF SEMICONDUCTOR DEVICES

Tunnel Diodes A tunnel diode is a P-N junction device which differs from the junction diode and the Zener diode in the sense that it has a high percentage of impurities in its semiconductor elements. As a result of heavy doping, the charge carriers are able to pass through the depletion layer near the junction by a process called "tunneling". The device permits both forward and reverse current with the very important feature that it presents a negative resistance region for a specific range of forward voltages. In this negative resistance region, the current decreases with increase of applied voltage. Because of this characteristic of negative resistance, tunnel diodes are used as amplifiers and oscillators particularly at low power microwave frequencies to avoid radiation effects. The circuit symbol for a tunnel diode is given in Fig. 6.13(a).

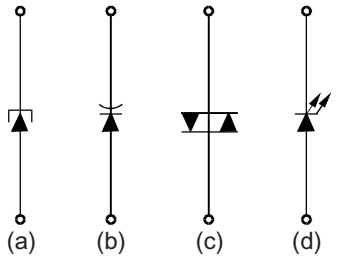


Fig. 6.13 Circuit Symbols for Various Types of Diodes (a) Tunnel Diode (b) Varactor Diode (c) Varistor Diode (d) LED

Varactor Diodes A reverse-biased diode junction behaves like a capacitor because the reverse bias keeps the charges separated away from the junction and the depletion zone. Moreover, the capacitance at the junction can be controlled by the reverse bias applied to the diode which makes the depletion zones wider or narrower as the reverse bias is varied. Such a voltage sensitive capacitor is called a varactor and finds extensive use in television circuits as a fine frequency control. The circuit symbol for a varactor is given in Fig. 6.13(b).

Varistor Diodes A varistor consists of two junction diodes of opposite polarities connected as shown in its circuit symbol in Fig. 6.13(c). A varistor is generally used as shunt across the collector in transistor circuits as a protection against sudden jumps of voltage either of a positive or of a negative nature.

Light Emitting Diode (LED) P-N junctions made of certain special materials like gallium compounds emit light as a result of energy released by the recombination of charges. Radiations of red, green or yellow light are emitted when a forward bias of the order of 1.2 V is applied from special types of cells.

LED units are arranged in segments which are switched on or off by digital control circuit to produce a display of various combinations of digits from 0 to 9. LEDs find extensive use in electronic watches and many electronic metering devices. Figure 6.13 (d) shows the circuit symbol of LED.

Liquid crystal displays (LCD) which depend on the ambient light for their light emitting properties, need a much less load current than LED displays.

6.8 JUNCTION DIODE AS A RECTIFIER

A diode presents a high resistance in one direction and a low resistance in the other direction. It can, therefore, be used as a rectifier for converting AC into DC. Power supply circuits using diodes as rectifiers are commonly used in transistor radios, tape recorders and record players as eliminators for the mains operation of these equipments.

There are two methods of rectification: (i) half-wave rectification, and (ii) full-wave rectification.

6.8.1 Half-wave Rectification

This is the simplest method of rectification as shown by the arrangement in Fig. 6.14. When the transformer secondary voltage is on its positive half cycle the diode *CR*, is forward-biased and current flows through the load R_L and a voltage is developed across the load. When the secondary voltage is on its negative half cycle, the diode is reverse-biased and no current flows through R_L and no voltage is developed across it. The input and output waveforms are as shown in Fig. 6.14, only the positive peaks are obtained and the negative peaks are suppressed. Thus, we get a pulsating DC as the output voltage, the average value being 0.318 ($1/\pi$) of the peak value. The pulsating DC can be converted to steady DC by the use of filter circuits described earlier.

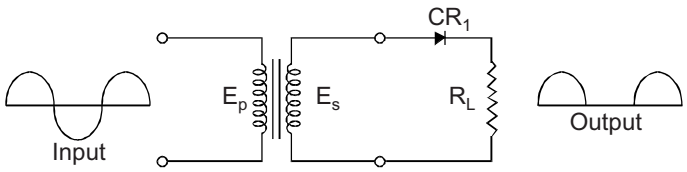


Fig. 6.14 Half-wave Rectifier Circuit and Waveform

6.8.2 Full-wave Rectification

A generally more useful and efficient method of converting AC into DC is to recover both the positive and negative portions of the AC input voltage. The circuit for this method, known as full-wave rectification, is given in Fig. 6.15. With the polarities shown, the CR_1 diode is forward-biased and conducting and the CR_2 diode is reverse-biased and non-conducting. The situation reverses itself when the polarity of the transformer secondary voltage E_s reverses itself on the next AC half cycle. Since each diode conducts only half the time on

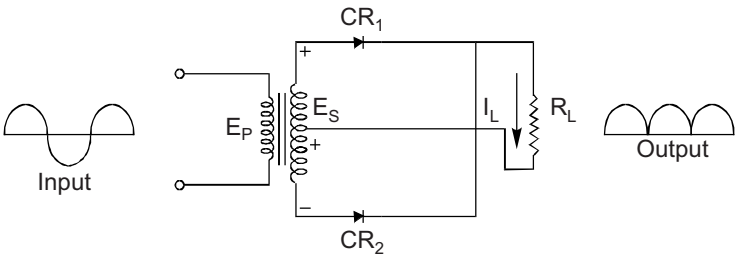


Fig. 6.15 Full-wave Rectification

alternate half cycles, a load current of double the rating of each diode can be obtained. The frequency of the pulsating DC output is double the frequency of the input waveform. Full-wave rectifiers have a smoother output voltage than half-wave rectifiers. With filtering, the load current can be made quite smooth.

6.8.3 Bridge Rectifier Circuit

The centre tap circuit was the most popular full-wave rectifier till the advent of low cost highly reliable and small sized silicon diodes. This has made the bridge circuit shown in Fig. 6.16 the most popular circuit today. The reason for this is a reduction in transformer size for the same available output power as for a centre tap circuit. In the centre tap circuit, the current is drawn through opposite halves of a secondary on alternate half cycles whereas the bridge circuit uses the entire transformer secondary during both half cycles allowing higher output power before core saturation than the centre tap circuit for an equivalent sized transformer.

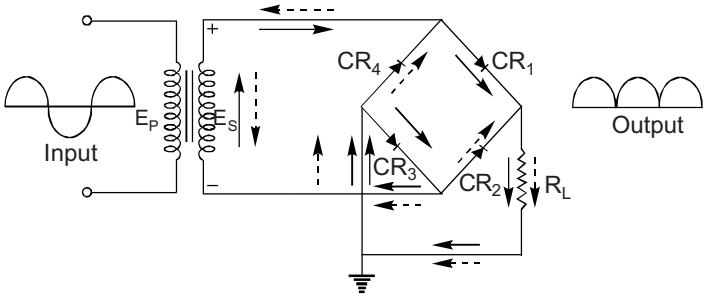


Fig. 6.16 Bridge Rectifier

The obvious disadvantage is the need for twice as many diodes as that for the *CT* version. However, because of the availability of low cost semiconductor diodes, the bridge circuit is still the most economical. Diodes CR_1 and CR_3 being forward-biased conduct during the positive half cycle of the input AC and diodes CR_2 and CR_4 do not conduct because of their being reverse-biased. The direction of conventional current flow is shown by full-line arrows in Fig. 6.16. During the negative half cycle, diodes CR_2 and CR_4 are forward-biased and CR_1 and CR_3 are reverse-biased and the direction of current flow is

indicated by the dotted arrows. It may be noted that the direction of conventional current is the same through the load R_L which develops a pulsating DC output voltage of the waveform shown in the figure.

6.9 TESTING A DIODE

The property of a diode presenting a low resistance when forward-biased and a high resistance when reverse-biased can be used to test a diode with a multimeter used as an ohmmeter. It may be remembered that a multimeter working in the ohmmeter position has a battery connected inside the meter which can either forward bias the diode or reverse bias it, depending upon whether the negative or the positive terminal is connected across the diode as shown in Fig. 6.17 (a) and the resistance of the diode noted. The connections of the meter of the diode leads are then reversed as in Fig. 6.17 (b), and the resistance noted again. If the resistance of the diode is significantly greater in one direction than in the other, the diode is in good condition. If the test shows direct continuity or the same amount of deflection on both sides, the diode is defective and needs replacement.

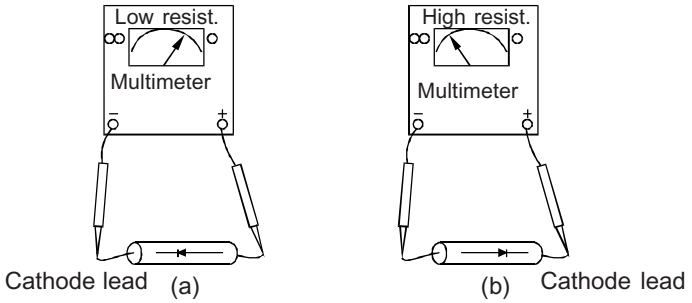


Fig. 6.17 Testing a Diode (a) Methods of Connecting Multimeter Probes to Diode Leads (b) Diode Leads Reversed

The above test can also be used to identify the leads of an unmarked diode or when the markings on the diode are not clearly visible. For this the polarity of the multimeter or ohmmeter lead must be determined by checking with a DC voltmeter or another multimeter. Then the lead of the diode which shows low resistance when connected to the negative lead of the meter is the cathode lead.

It is useful to remember that in certain multimeters the terminal marked negative (–) on the meter is actually connected to the positive terminal of the battery inside.

6.10 TRANSISTORS

A transistor is a three-element semiconductor device formed by placing two P-N semi-conductor junctions back-to-back as shown in Fig 6.18. The two inner segments of N- or P-type material are actually combined into a single segment which is lightly doped and made thin compared to the two outer

Fig. 6.18 Transistor as a Combination of Two Junction Diodes

segments as in Fig. 6.19. Three distinct regions of the material are actually combined into a single segment which is lightly doped and made thin compared to the two outer segments as in Fig. 6.19. Three distinct regions of the material are formed in this way and are called Emitter, Base and Collector, represented by letters E, B and C respectively. Two different types of transistors are formed depending upon whether the N-type material is sandwiched between two sections of the P-type material or the P-type material is sandwiched between two section of the N-type material. The two types of transistors are known as PNP and NPN type.

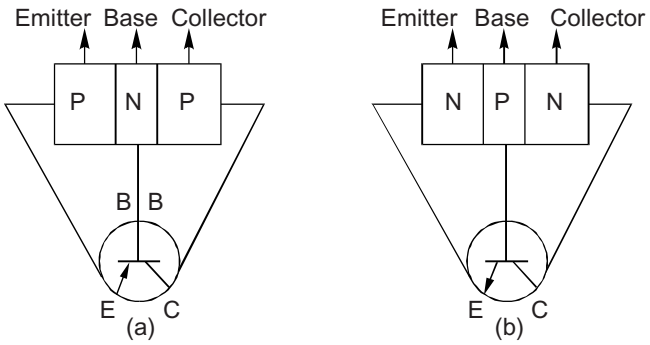


Fig. 6.19 Emitter, Base and Collector in a Transistor (a) PNP Transistor (b) NPN Transistor

PNP Transistor In this type of transistor, a thin layer of N-type material is surrounded by two comparatively thicker layers of P-type semiconductor material. The middle layer is called the *base* (B), the left layer the *emitter* (E), and right side layer is called the *collector* (C) as shown in Fig. 6.19 (a) which also indicates the circuit symbol used for the PNP type of transistor. Notice that the arrow for the emitter points inwards towards the base.

NPN Transistor In this type of transistor a thin layer of P-type material is sandwiched between two comparatively thicker layers of N-type semiconductors. The construction and the circuit symbol are shown in Fig. 6.19 (b). The arrow for the emitter points outwards away from the base.

6.11 CONSTRUCTION OF A TRANSISTOR

A transistor is a combination of two P-N junction diodes. Accordingly, it is constructed either as a *grown junction* or as a fused junction crystal. Transistors are called silicon transistors or germanium transistors depending upon

whether the basic material used for its construction is silicon or germanium. Figure 6.20 shows the constructional details of a low frequency germanium output transistor. After completion of the assembly, the wires are bounded to the electrodes and connected to their respective pins. The entire unit is then enclosed in either a metal can or a plastic case which is filled with some inert material and sealed to prevent any moisture getting in. In most cases the transistors are manufactured with flexible leads and are meant to be soldered directly into the printed circuit boards.

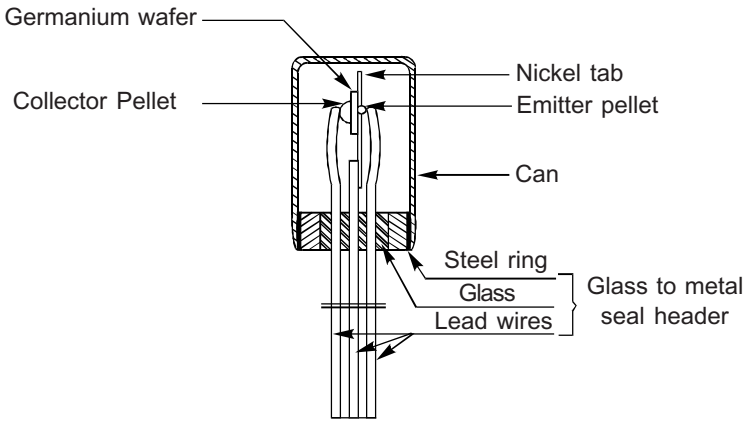


Fig. 6.20 Construction of a Low Frequency Germanium Output Transistor *(Courtesy:* BEL)

6.12 SYMBOLIC IDENTIFICATION OF TRANSISTORS

In schematic diagrams the transistors are identified as PNP or NPN type by their circuit symbols as given in Fig. 6.21. In PNP type transistors, the emitter arrow points inwards towards the base and in the NPN type, the emitter arrow points outwards away from the base. It is interesting to note that the direction of the arrow indicates the direction of flow of conventional current and is opposite to the direction of electron current flow inside the transistor.

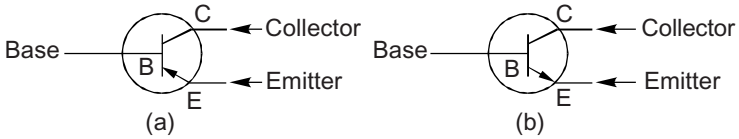


Fig. 6.21 Circuit Symbol for Transistors (a) PNP Type (b) NPN Type

6.12.1 Type Number Codes

To distinguish one transistor from another, transistors are generally given a type number. In American transistors, letter N indicates a semiconductor with a numerical prefix for the number of junctions. Thus 1N will indicate a diode,

2N a transistor, and so on. The digits that follow will indicate the specific type as in 2N137. Certain other types of transistor are labelled as 2*SA* for PNP and 2*SC* for NPN types. Type designation codes for most Indian and European semiconductor devices like diodes and transistors generally consist of two letters followed by a serial number. The first letter distinguishes between junction and non-junction devices and also gives an indication of the semiconductor material like germanium, silicon and gallium, arsenide, etc. The second letter indicates the class and main application. The serial number consisting of three figures or one letter and two figures is meant to indicate the use for which the device is designed. For example, AC128 is a germanium (A) transistor for AF applications (C) and BF194 is a silicon (B) transistor for HF applications (F).

6.12.2 Identification of Leads

For the identification of the leads of a transistor, certain standard methods are used by the manufacturer for arranging these leads in the base of the transistor. When the three leads are arranged in a line as shown in Fig. 6.22(a), these can be identified as emitter, base and collector in the order shown. However, the collector spacing is kept greater than the spacing of the other leads. When the three leads are equidistant as in Fig. 6.22 (b), a coloured dot (red or green) is painted on the case close to the collector lead.

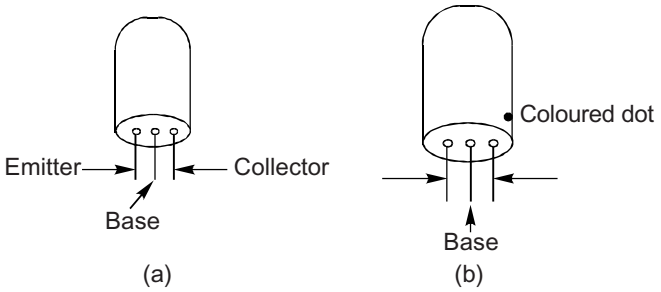


Fig. 6.22 Identification of Transistor Leads

Neither the identification codes nor the arrangement of leads in the base indicate whether a particular transistor is of the PNP or NPN type. This information together with other technical data is available in transistor manuals or application notes supplied by the manufacturers. These manuals also supply information regarding the use of transistors in various types of circuits. Manufacturers generally provide a schematic diagram indicating the position and identification of each lead.

Figure 6.23 indicates the bottom view of some of the common transistors used in radio receivers and television circuits. Transistors in a metal case are often provided with a fourth lead called *shield*(S) which is internally connected to the case as in Fig. 6.23 (d) (BF 167/173).

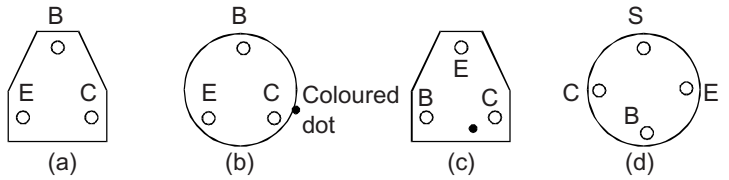


Fig. 6.23 Bottom View and Positioning of Leads in some Common Transistors (a) BC148B (b) AC 128 (c) BF 194B, BF 195C, BF 195D (d) BF 167/173

6.13 TRANSISTOR OPERATION

A transistor is a combination of two P-N junctions so arranged as to have a P or N type semiconductor sandwiched between opposite types. The middle section called the *base* forms two separate junctions with the outer sections known as *emitter* and *collector*. When the *emitter-base* junction is forward biased and the *collector-base* junction is reverse-biased, the emitter supplies charges, either holes or electrons, which pass through the base and are collected by the collector. The base exercises necessary control on the flow of collector current.

PNP Transistor In Fig. 6.24 (a), the emitter-base junction of a PNP transistor is forward-biased by a small battery B-1 and the collector-base junction is reverse-biased by a bigger battery B-2. The holes in the p-region on the left are repelled by the positive terminal of the B-1 battery and under the influence of the electric field, these holes cross over to the N-region or the base region which is comparatively thin and is lightly doped. A majority of the holes are able to drift across the base region and enter the collector region. A small number of holes, about 5 per cent, are lost in this area due to recombination with electrons. The holes that cross the base-collector junction into the collector region are attracted towards the collector by the negative polarity of the B-2 battery. As each hole reaches the collector electrode, it is neutralised by an electron from the negative terminal of the B-2 battery. For each hole that is lost by combination with an electron in the base region and the collector region, an electron is released near the emitter electrode by the breaking down of a covalent bond and the electron so liberated by the positive terminal of the B-1

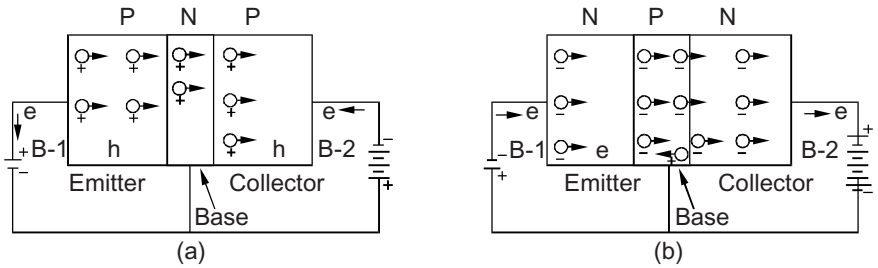


Fig. 6.24 How a Transistor Operates (a) Operation of a PNP Transistor (b) Operation of a NPN Transistor

battery. The new hole formed by the breaking down of the covalent bond then moves towards the emitter-base junction and this process goes on and on.

It is clear from the above that conduction within the PNP transistor is the result of hole conduction and the conduction in the external circuit is the result of electron movement. It is also clear that the collector current is less than the emitter current by an amount (5 per cent) proportional to the number of electron hole combinations occurring in the base area.

NPN Transistor In this case again, the emitter-base junction is forward-biased and the collector base junction is reverse-biased as in Fig. 6.24(b). The polarities of the B-1 and B-2 batteries are reversed compared to the polarities in the case of a PNP transistor. The free electrons in the N-type emitter region are repelled by the negative polarity of the B-1 battery and these electrons are able to cross the barrier potential at the emitter-base junction and enter the P-type base region which is thin and lightly doped. Excepting a few electrons (5 per cent), which combine with the holes in the base region, the rest of the electrons are swept over to the collector electrode by the positive potential of the B-2 battery and enter the positive terminal of the B-2 battery. For every electron that leaves the collector electrode, an electron from the negative terminal of B-1 battery enters the emitter region and is pushed toward the emitter-base junction. Thus the conduction in this case is entirely due to electrons, both inside and outside the transistor.

It is, therefore, clear from the above explanation that the majority carriers are holes in the case of PNP transistors, whereas in the NPN transistors, the majority carriers are electrons. In both cases, the collector current is less than the emitter current by an amount equal to the recombination of holes and electrons.

The above explanation for the operation of a transistor indicates that there are three type of currents that flow in a transistor circuit. These currents are, the emitter current I_E, the base-circuit I_B, and the collector current I_C. Figure 6.25 indicates the direction of flow of these currents in the NPN type of transistor.

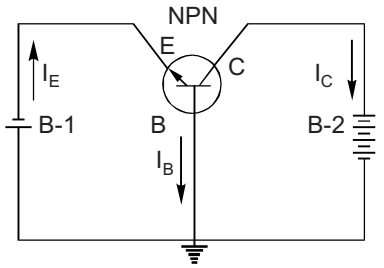


Fig. 6.25 Current Flow in an NPN Transistor

Furthermore, it is clear that most of the charge carriers supplied by the emitter are able to diffuse through the comparatively thinner and lightly doped base and reach the collector electrode, thus constituting the collector current I_C. Only a very small percentage (less than 5 per cent) of the charges recompense

in the base region to provide the very small base current I_B. These facts can be quantitatively represented as follows:

$$I_E = I_C + I_B \tag{6.1}$$

EXAMPLE: In a transistor circuit, $I_E = 10$ mA and $I_B = 200\ \mu$ A. Find the value of I_C. From Eq. (6.1)

$$I_C = I_E - I_B$$
$$= 10 \text{ mA} - 0.2 \text{ mA} = 9.8 \text{ mA}$$

(In this case $I_B = 200\ \mu\text{A} = 0.2$ mA).

6.14 TRANSISTOR AS AN AMPLIFIER

The operation of a transistor as an amplifier is based on the fact that base current (I_B) in a transistor can control the collector current (I_C). The base current can be varied by variations of forward bias and this will produce corresponding variations in the collector current. Here the action of a transistor can be compared to the action of a triode valve, where the grid bias controls the anode current. The only significant difference between the operation of the two electronic devices is that whereas the triode valve is a voltage operated device, a transistor is a current operated device.

The emitter-base junction in a transistor is forward-biased and, as such, the input impedance (resistance) is low. On the other hand, the base-collector junction is reverse biased and the output impedance is very high. Any input AC signal will either oppose or help the forward bias and as a result the base current will either decrease or increase. This will produce corresponding variations of collector current in the external output circuit. A high value of load resistor (R_L) in the collector circuit will produce a varying voltage drop across the load resistor. This voltage drop will be generally greater than the input signal voltage variations and the input signal will appear as an amplified voltage in the output circuit.

A transistor has only three elements or electrodes. The input voltage can be applied between any two electrodes and the output voltage is also available between the other two electrodes. Since there are only three electrodes available, one of the electrodes will be common between the *input* and the *output*. There are thus, three possible methods of connecting a transistor in an amplifier circuit. These are (i) common-base (CB) circuit, (ii) common-emitter (CE) circuit, and (iii) common-collector (CC) circuit. All the three configurations are shown in Fig. 6.26 for a PNP transistor.

1. Common Base (CB) Circuit Figure 6.26(a) shows a common base circuit of a PNP transistor in which the base (B) is the common or grounded electrode. The input signal is applied between the collector and the ground (base). V_{EE} supplies the forward bias and V_{CC} the reverse bias for the transistor. The input has a low resistance because of high I_E and the output resistance of the circuit is high. The circuit has no current gain because I_C is less than I_E. As

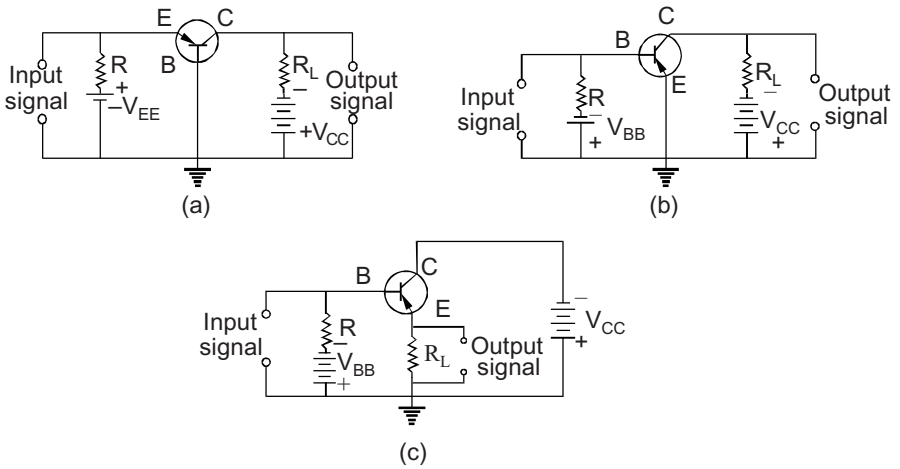


Fig. 6.26 PNP Transistor as an Amplifier-CB, CE and CC Configurations (a) Common Base (CB) (b) Common Emitter (CE) (c) Common Collector (CC)

such, the circuit is seldom used in actual practice. The main characteristics of the circuit are:

1. It has no current gain but voltage gain can be high.
2. It has a low input resistance but high output resistance.
3. The output voltage is in phase with the input voltage.

2. Common-emitter (CE) Circuit In the CE circuit shown in Fig. 6.26 (b), the input single is applied between the base and the emitter which is the grounded or common terminal. The output is taken between the collector and the emitter (ground). In this case, the current flowing in the input circuit is I_B which is much less than I_E. As a result, the input resistance of a CE circuit is much higher than that of a CB circuit. Since I_C is much greater than I_B the CR circuit has a high current gain but the same voltage gain as in a CB circuit. Because of the combined advantages of voltage gain and current gain, the CE circuit is the one most commonly used in amplifier circuits. When used as a multistage amplifier, the CE circuit does not overload the previous stage because of its high input resistance. The disadvantage with this circuit is that it needs bias stabilisation to avoid amplification of reverse leakage current in the CE circuit. The CE circuit possesses the following characteristics.

1. It can provide both current gain and voltage gain.
2. It has a high input resistance compared to a CB circuit.
3. The output voltage is 180° out of phase with the input voltage.
4. It needs bias stabilisation to avoid damage due to amplification of reverse leakage current.

3. Common Collector (CC) Circuit In this circuit, the collector is connected to the ground through the reverse bias battery V_{CC} which is supposed to have practically no internal resistance. The input voltage is applied between the base and the ground (C) and the output is available between the emitter and

ground (C). Thus, C is the grounded or common electrode. There is no voltage gain because the bypassed resistor R_L in the emitter circuit provides 100 per cent feedback.

6.15 STANDARD LETTER SYMBOLS FOR TRANSISTORS

In order to distinguish between various types of voltages and currents (dc and ac) in a transistor (or vacuum tube) circuit, standard letter symbols are used to represent these quantities. In general the capital letters V and I are used for average dc values. Double subscripts such as V_{CC} are used for voltages that do not change. The value that change with time are represented by letters like v and i. Table 6.1 gives the standard letter symbols for various voltages in a transistor circuit.

Table 6.1 Standard Symbols for Transistor Voltages

Definition of voltage	*Symbol*		
	Collector	*Emitter*	*Base*
Supply voltage	V_{CC}	V_{EE}	V_{BB}
Average dc voltage	V_C	V_E	V_B
AC component	v_c	v_e	v_b
Instantaneous value	v_c	v_E	v_B
RMS value of a componet	V_c	V_e	V_b

A similar table can be prepared for standard symbols for the current values for different electrodes in a transistor.

Other symbols that are commonly used are ICBO which means the current between the collector and base when the third electrode i.e emitter is open. BVCBO indicates breakdown voltage, collector to base when the emitter is open.

6.16 CURRENT GAIN IN TRANSISTORS—ALPHA (α) AND BETA (β) CHARACTERISTICS

A transistor is a current operated device. Any change of current in one circuit affects the current in the other circuit.

α (Alpha) In common base (CB) circuit, a change in emitter current ΔI_E will result in a corresponding change ΔI_C in the collector current. The relation between the two currents is expressed by a control characteristic α (alpha), which is defined as the ratio of a change in collector current ΔI_C to the corresponding change in emitter current ΔI_E with the collector to base voltage V_{CB} remaining constant. Expressed mathematically,

$$\alpha = \frac{\Delta I_C}{\Delta I_E} \, (V_{CB} \text{ constant}) \tag{6.2}$$

Since the collector current in a CB circuit is slightly less than the emitter current, α is less than one in this case. It lies between 95 to 99 per cent. Thus, in a CB circuit, there is no current gain. Actually, there is a current loss. However, the circuit exhibits voltage and power gain.

β (Beta) In the common emitter (CE) circuit, the current gain is represented by the Greek letter β (beta) and is defined as the ratio of the change in collector current I_C resulting from the change in base current I_B with the collector to emitter voltage (V_{CE}) remaining constant. Thus,

$$\beta = \frac{\Delta I_C}{\Delta I_E} \quad (V_{CE} \text{ constant}) \tag{6.3}$$

I_C being greater than I_B, β is always greater than one. A transistor with $\beta = 50$ means that 50 times more current is flowing through the collector than through the base.

It has already been established in Eq. (6.1) that

$$I_E = I_C + I_B$$

Representing the changes in emitter current, collector current and base current by ΔI_E , ΔI_C and ΔI_B respectively Eq. (6.1) can be written as

$$\Delta I_E = \Delta I_C + \Delta I_B \tag{6.4}$$

or
$$\Delta I_B = \Delta I_E - \Delta I_C \tag{6.5}$$

The ratio
$$\frac{\Delta I_C}{\Delta I_B} = \frac{\Delta I_C}{\Delta I_E - \Delta I_C} \tag{6.6}$$

Dividing the numerator and denominator of the right-hand side of Eq. (6.6) by ΔI_E, we have

$$\frac{\Delta I_C}{\Delta I_B} = \frac{\Delta I_C}{\Delta I_E} \Big/ \left(1 - \frac{\Delta I_C}{\Delta I_E}\right) \tag{6.7}$$

We know that the ratio $\Delta I_C/\Delta I_E$ (V_{CB} constant) is α and the ratio $\Delta I_C/\Delta I_B$ (V_{CE} constant) is β. Substituting in Eq. (6.7) we get

$$\beta = \frac{\alpha}{1 - \alpha} \tag{6.8}$$

Knowing the value of α in a transistor, the value of β can be derived from Eq. (6.8). Thus if $\alpha = 95\%$ (or 0.95), the value of $\beta = 0.95/1 - 0.95 = 0.95/0.05 = 19$.

Similarly, a transistor with an α of 0.99 will have β equal to 99.

6.17 CHARACTERISTIC CURVES FOR TRANSISTORS

A transistor is a non-linear device and as such the relation between the various electrode voltages and currents is best expressed by curves known as characteristic curves. These curves are generally supplied by the manufacturers and help in determining the performance of a particular transistor.

Figure 6.27 (a) is an experimental arrangement for determining the I_C versus V_C characteristics of a transistor in CE configuration with a fixed value of I_B in each case. A family of curves obtained in a typical case is shown in Fig. 6.27(b). I_C is measured for different values of collector to emitter voltage V_{CE} while I_B is kept fixed at a particular value. For another curve the value of I_B is changed to a different value and the same process of noting I_C for different values of V_{CE} is repeated.

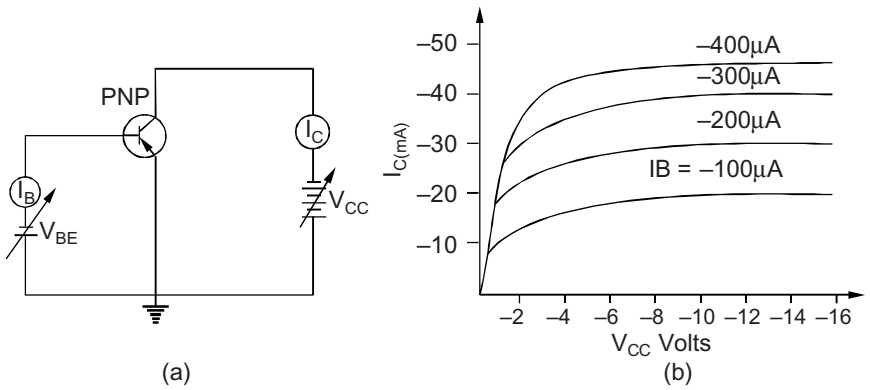


Fig. 6.27 Characteristic Curves for Transistors (a) Experimental Set Up for I_C Versus V_C Curves (b) Family of Characteristic Curves

It is possible to determine the value of β from these curves. Thus, for fixed $V_{CE} = -4$V, $I_C = 10$ mA when $I_B = 100$ μA and for the same fixed value of V_{CE} (4V) $I_C = 18$ mA when $I_B = 200$ μ A. From these values we can work out ΔI_C and ΔI_B as follows:

$$\Delta I_C = 18 \text{ mA} - 10 \text{ mA} = 8 \text{ mA}$$

and

$$\Delta I_B = 200\ \mu\text{A} - 100\ \mu\text{A} = 100\ \mu\text{A}$$

$$\beta = \frac{\Delta I_C}{\Delta I_B} = \frac{8 \text{ mA}}{0.1 \text{ mA}} = 80$$

α can then be determined from the relation $\beta = \alpha/(1 - \alpha)$
In this particular case when $\beta = 80$,

$$80 = \frac{\alpha}{1 - \alpha}$$

$$80 - 80\ \alpha = \alpha$$

$$81\ \alpha = 80$$

$$\therefore \quad \alpha = \frac{80}{81} = 0.9 \text{ approx.}$$

6.18 TRANSISTOR ANALYSIS—LOAD LINE

The characteristic curves shown in Fig. 6.28(b) have been drawn with the experimental setup as in Fig. 6.28 (a) when there is no load in the output circuit.

In order that this circuit behaves as an amplifier, an external load (resistance or impedance) in the output circuit is necessary as shown in Fig. 6.28(a).

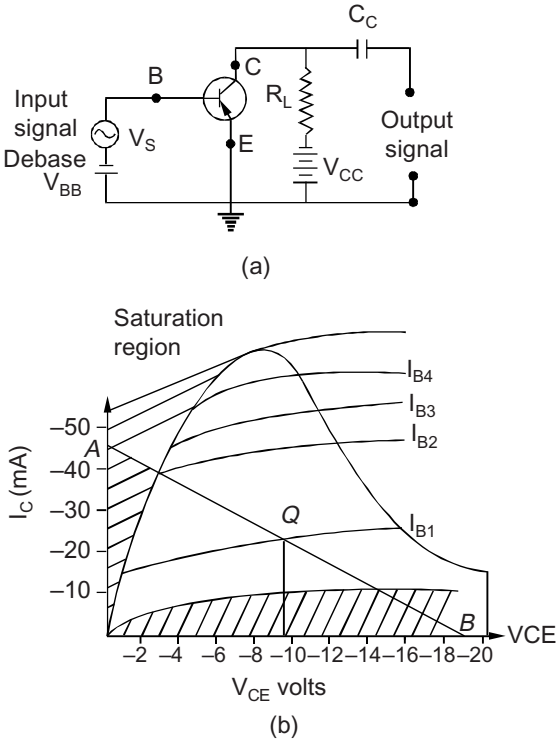


Fig. 6.28

The resistance R_L has linear characteristics compared to the non-linear characteristics of the transistor and can be represented by a straight line AB superimposed on the characteristics of the transistor. The straight line AB which can be used to analyse the behaviour of the circuit with the load R_L is called the *load line*.

How to Draw the Load Line The end points A and B of the load line can be determined if the value of the load R_L and the value of the supply voltage V_{CC} are known. For point B the value of $I_C = 0$ and the value of $V_{CE} = V_{CC} - I_C \times R_L$. The point B has the co-ordinates (V_{CC}, 0). For point A, $V_{CE} = 0$, $V_{CE} = 0 = V_{CC} - I_C R_L$.

$$\therefore \qquad I_C = \frac{V_{CC}}{R_L}$$

So the point A is $\left(\frac{V_{CC}}{R_L}, 0\right)$

The straight line joining the two points $A\left(\frac{V_{CC}}{R_L}, 0\right)$ on the current axes and the point B ($V_{CC,\,O}$) on the voltage axes is called the *load line*. For V_{CC} = 20 V and R_L = 5Ω, the two points on the dc load line will be A (4, 0) and B(0, 20). For any value of I_C, the corresponding V_{CE} must be on the load line which takes into account R_L voltage drop.

Q-point The point where the load line intersects the curve for the base-bias current (I_B) is called the Q point or Quiescent point.

This point specifies the static dc values without any ac signal input. This point which is also called the operating point for an amplifier is so chosen that it lies in between the saturation collector current and cut off collector current. This is necessary to avoid distortion in the output of the amplifier.

After selection of the operating point an ac voltage is applied to the input. Due to this voltage the base current (I_B) varies from time to time. As a result the collector current and collector voltage also vary with time. These variations of collector current and voltage can actually be plotted on the characteristics of the transistor.

If the operating point Q, is very near to the saturation region both the output voltage and current are clipped at the positive peaks as shown in Fig. 6.29 (a).

It the operating point Q_2 is very near to cut off region, the output signal is clipped near the negative peaks as in Fig. 6.29 (b).

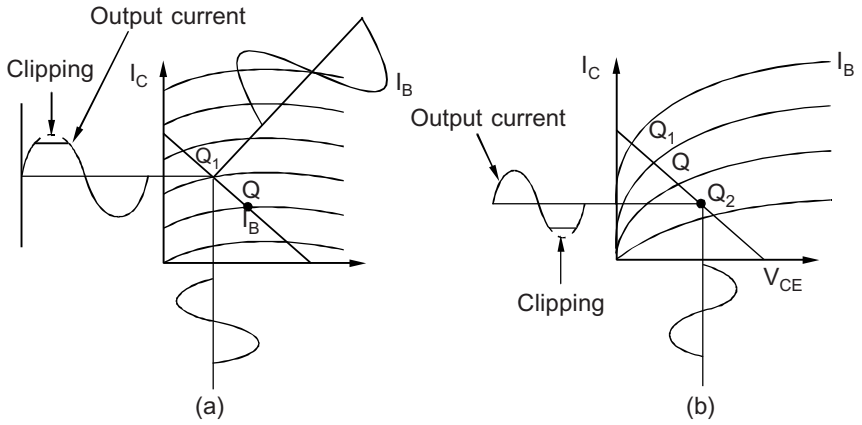


Fig. 6.29 (a) Clipping at Positive Peak (b) Clipping at the Negative Peaks

It is only if the operating point is at Q that the output is without distortion as shown in Fig. 6.30.

It is however necessary that the input signal should not be too large to avoid clipping of the output voltage both at the positive and negative peaks.

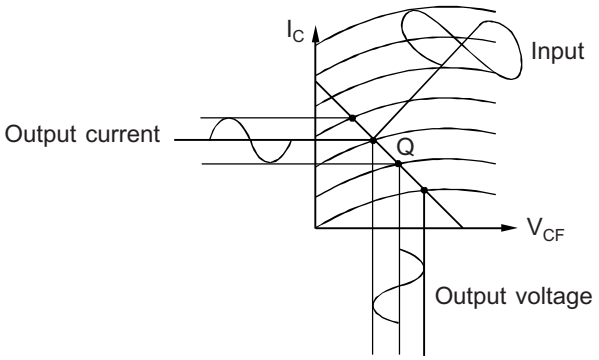


Fig. 6.30 Output Distortion

6.19 h-PARAMETERS FOR A TRANSISTOR

A transistor is a three-terminal device. The input and output to a transistor requires two terminals each. Since there are only three terminals available it becomes necessary to make one terminal common both to the input and the output as shown in Fig. 6.31. In this Fig. (1, 3) are input terminals and (2, 3) are output terminals.

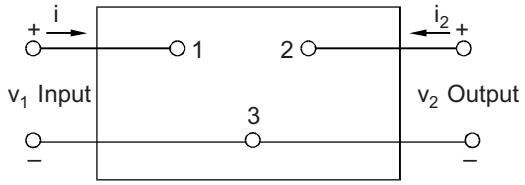


Fig 6.31 Common Terminal in a Transistor

Terminal 3 is the common terminal. The four variables involved in this transistor circuit are the input voltage and input current (v_1, i_1) and the output voltage and output current (v_2, i_2). The relation between these four variables can be represented by the following two equations:

$$v_1 = h_{11}\, i_1 + h_{12}\, v_2 \qquad (6.9)$$

$$i_2 = h_{21}\, i_1 + h_{22}\, v_2 \qquad (6.10)$$

The parameters h_{11}, h_{12}, h_{21} and h_{22} which are used to relate the four variables in the above equations are called h-parameters or *hybrid* parameters. These are called hybrid parameters because these are generally defined by mixture of constants having different units.

The h-parameters are convenient for specifying the characteristics of transistors and these can also be measured easily. To define the h-parameter in terms of the variables mentioned above the two equations have to be solved under specific circuit conditions as explained below:

If the output terminals (2, 3) are shorted $v_2 = 0$ and

$$v_1 = h_{11}\, i \quad (h_{12}\, v_2 = 0)$$

and $$i_2 = h_{21}\, i \quad (h_{22}\, v_2 = 0)$$

Form these equations we have

$$h_{11} = \left.\frac{v_1}{i_1}\right|_{v_2 = 0} = \text{Input impedance, being the ratio of voltage to current}$$

$$h_{21} = \left.\frac{i_2}{i_1}\right|_{v_2 = 0} = \text{pure number}$$

Similarly, if the input terminals are kept open then $i_1 = 0$, the two equations become

$$v_1 = h_{12}\, v_2$$

and $$i_2 = h_{22}\, v_2$$

from these equations we have

$$h_{12} = \left.\frac{v_1}{v_2}\right|_{i_2 = 0} = \text{pure number, being the ratio of two voltages}$$

$$h_{22} = \left.\frac{i_2}{v_2}\right|_{i_1 = 0} = \text{ratio of current to voltage which is admittance}$$

That the h-parameters are really hybrid in nature is shown by the fact that $h_{11} = \dfrac{\text{voltage}}{\text{current}} = h$ and has the units of resistance (Ω) and $h_{22} = \dfrac{\text{current}}{\text{voltage}} = h_0$ and has the units of admittance or mho (℧). However, $h_{12} = \dfrac{\text{voltage}}{\text{voltage}}$ is a pure number = h_r, and $h_{21} = \dfrac{\text{current}}{\text{current}} = h_f$ (forward current ratio) is also a pure number. These have no units. Quite often the manufacturers specify the transistor characteristics in terms of h-parameters which can be easily measured.

An additional suffix e added to h-parameters sometimes indicates that the transistor is being used in the CE mode. In this case the terminal is the base terminal and terminal 2 is the collector terminal. With this notation the two equations (6.9) and (6.10) take the form

$$v_b = h_{ie} i_b + h_2 \cdot e v_e \tag{6.11}$$

$$i_c = h_{fe} i_b + h_{cs}\, v_c \tag{6.12}$$

Applying Kirchoff's voltage law to equation 6.11 and Kirchoff's current law to equation 6.12, the complete h-parameter equivalent of a transistor can be represented by the Fig. 6.32.

The h-parameters at a given operating point can be determined from the static characteristics of the transistor.

6.20 α AND β CUT-OFF FREQUENCIES

The frequency response of a junction transistor at higher frequencies is limited by its internal capacitances and the finite time taken by the charge carriers to

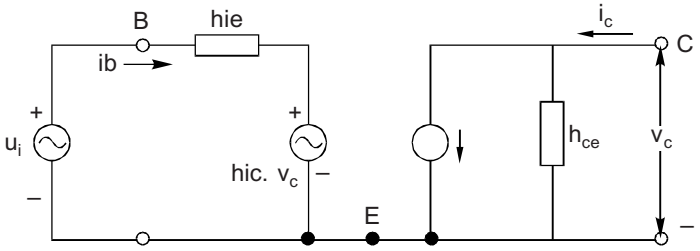


Fig. 6.32 Complete h—equivalent of a Transistor

diffuse through the base region. The upper frequency limit for a transistor is called the α (alpha) *cut-off frequency* and is defined as the frequency at which the value of α drops down to 0.707 of its value at 1 kHz. For small signal transistors used in RF circuits, α cut-off frequency is around 300 MHz.

β cut-off frequency can also be defined in the same manner.

6.21 TRANSISTOR BIASING

Transistor circuits discussed so far have used two voltage sources or batteries, one for the emitter-base bias (forward bias) and the other for the base-collector bias (reverse bias). However, there are practical methods by which we can eliminate the need for two separate batteries. A number of methods are available for doing this but the simplest is the potential divider method. In this method, the base-bias voltage is obtained from a potentiometer formed by two resistances R_1 and R_2 connected across the main battery as shown in Fig. 6.33. The values of R_1 and R_2 can be calculated from Ohm's law.

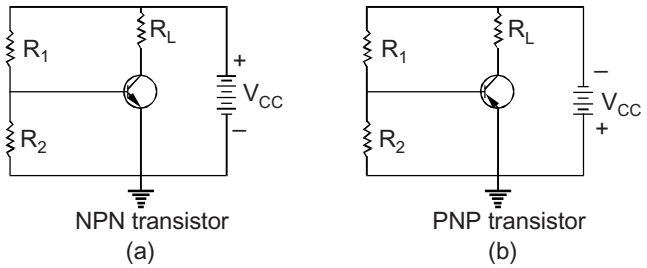


Fig 6.33 Potential Divider Method of Transistor Biasing

6.21.1 Transistor Bias Stabilisation—Thermal Runaway

Transistors in general, and germanium transistors in particular, are very sensitive to increase in temperature. As the temperature of a semiconductor rises, due to dissipation of heat at the collector junction, an increasing number of electrons in the crystal are set free producing an increase of collector current unrelated to the emitter or base current. This leakage current, though small, can cause a further rise in the collector temperature which releases still more electrons which further increase the electron current. This process, called

thermal runaway, is a cumulative effect which can cause permanent damage to the collector junction resulting in an internal short-circuit to the base. The effect is not so marked in silicon transistors, which can be used at much higher temperatures than their germanium equivalents. Most transistors manufactured these days are silicon transistors.

One method to prevent damage to the transistors is the use of *heat sinks* which are metal structures and when attached to the transistor radiate heat from the collector junction. Use of heat sinks becomes necessary only in the case of power transistors which develop enough heat during normal operation to be unbearably hot in spite of their not having any filament or heater.

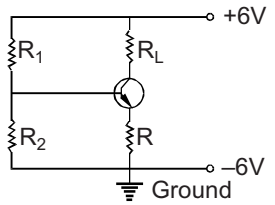


Fig. 6.34 Bias Stabilisation by Self-bias Resistor

Heat sinks can minimise stabilisation problems but cannot eliminate them. One of the simplest methods is to provide a stabilising resistor between the emitter and the common point as shown in Fig. 6.34.

Resistance R develops a self-bias due to IR drop which has a polarity opposite to forward bias and acts as a reverse bias for the base-emitter circuit and automatically compensates for change in temperature.

6.21.2 Bias Stabilisation with Thermistors and Diodes

Thermistors are special types of resistors whose resistance decreases with temperature compared to copper or tungsten resistors which have a positive temperature coefficient (resistance increasing with temperature). These temperature compensating components when suitably connected in the circuit as in Fig. 6.35 (a), can be used for bias stabilisation in a transistor circuit. Stabilisation can also be achieved with a compensating diode when connected to control the bias current in the base circuit as in Fig. 6.35 (b).

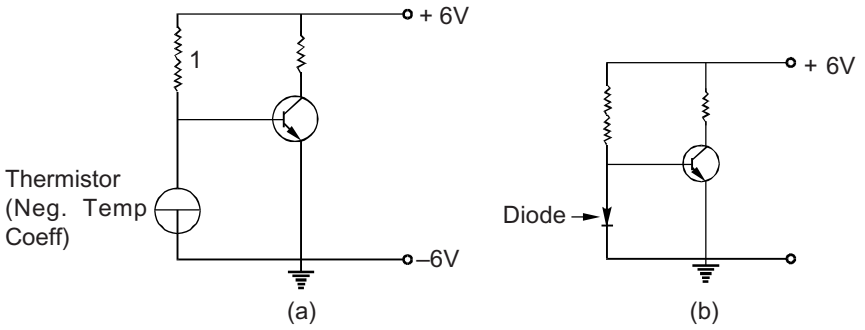


Fig. 6.35 Bias Stabilisation with Thermistors and Diodes (a) Thermistor for Bias Stabilisation (b) Bias Stabilisation with a Diode

6.22 FIELD EFFECT TRANSISTOR(FET)

The transistors described so far were two junction devices known as *bipolar* transistors because their operation depended on the action of two types of charge carriers, holes and electrons. There is another type of transistor, called *filed effect transistor* (FET), which is a *unipolar* device because its operation depends on the action of only one type of charge carrier. Field effect transistors are of two types-the Junction type FET (JFET) and the Metal Oxide Semiconductor type (MOSFET).

6.22.1 JFET

The main body of a JFET consists of an N-type (or P-type) material known as the channel and the other three elements are the Source (S), the Gate(G) and the Drain (D) as shown in Fig. 6.36(a). The P-type material which is embedded on both sides of the channel forms a junction with the N-type channel but the other two terminals, source and drain, are only ohmic contacts on the two sides of the channel. When the two gates are connected internally, the device is a single-gate FET, as in Fig. 6.36 (a), but when separate leads are provided for the two gates a dual-gate FET is formed. Circuit symbol for N-channel JFET is shown in Fig. 6.36 (b). For the P-channel JFET the arrow will be pointing outwards.

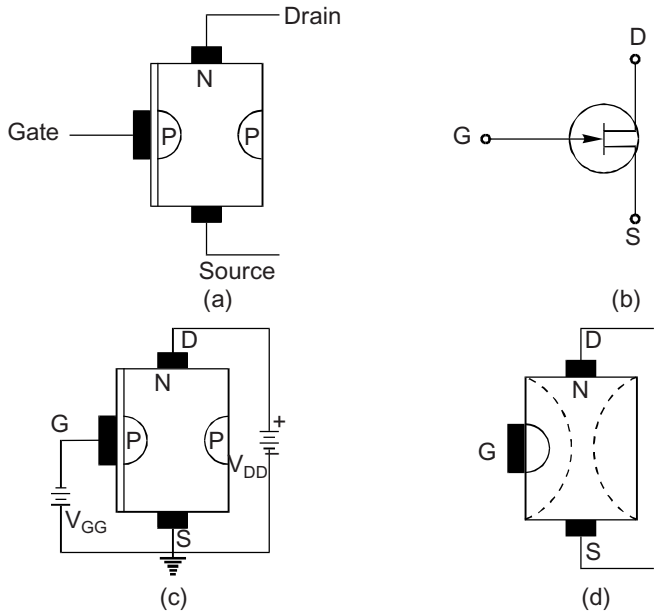


Fig. 6.36 JFET Circuit Symbol and Operation (a) Construction (b) Circuit Symbol for N-channel JFET (c) Operation of JFET (d) Reverse Bias on Gate Reduces Channel Width

In a JFET, the drain (D) corresponds to the collector, source (S) to the emitter and the gate (G) to the base of a bipolar transistor but the two types operate on different principles. When a voltage V_{DD} is applied between the drain and the source as in Fig. 6.36 (c), the free electrons in the N-channel are attracted by the positive terminal of the battery and electrons from the negative terminal enter the channel to replace those leaving the drain. This current can be controlled by the reverse bias V_{GG} applied to the gate. The channel, which has already been made narrow by the physical presence of the P-material, is further reduced in width by the electric field produced at the gate by the reverse bias voltage in Fig. 6.36 (d). This restricts the flow of current through the channel. If the reverse bias gate is made sufficiently high, the drain current is completely cut off due to blocking of the channel. It is the negative voltage on the gate that controls the drain current I_D and not the gate current.

The main advantage of JFET is that it has got a very high input resistance of the order of megaohms and is less sensitive to temperature changes. The disadvantages are less gain for a given bandwidth and smaller power rating compared to a bipolar transistor.

6.22.2 MOSFET or IGFET

The construction of the Metal-Oxide Semiconductor Field Effect Transistor (MOSFET) as shown in Fig. 6.37 is slightly different from that of the JFET already described. In MOSFET, the bulk of the material is neutral or lightly doped silicon which is known as *substrate* on which the rest of the structure is built. Deposited on the N-channel is a thin film of silicon-dioxide (SiO_2) which insulates the N-channel from the gate (G), which is a thin metallic film

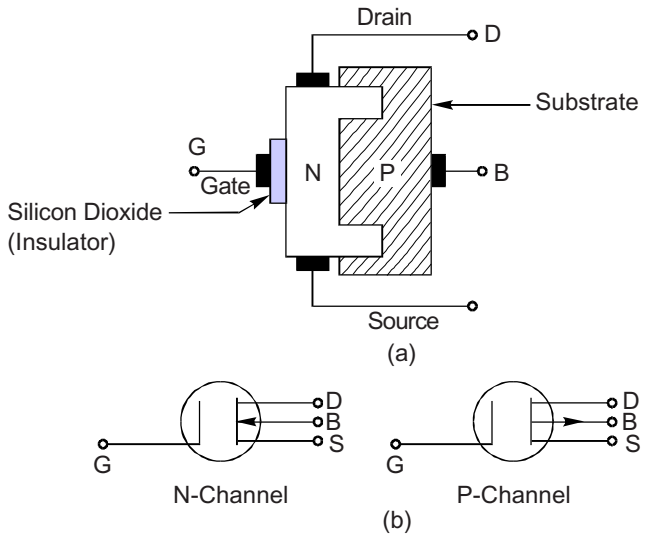


Fig. 6.37 MOSFET Construction and Circuit Symbols (a) Construction of MOSFET (b) Circuit Symbol for MOSFET (Depletion Type)

deposited on the silicon-dioxide insulator. The gate is insulated from the body of the FET and, as such, the MOSFET is also known as insulated gate FET or IGFET. Ohmic contacts are brought out for the gate, drain, source and the substrate as in Fig. 6.37(a). Figure 6.37 (b) shows the circuit symbols for a MOSFET.

The drain current I_D is controlled by the voltage on the gate. When the gate is made positive the MOSFET operates in the enhancement mode and when the gate is negative, it is the depletion mode. The insulated gate forms an electrostatic device with a very high input resistance of the order of several megaohms. To avoid a build up of static charges, the leads of an IGFET are kept shorted with a shorting ring when not used in a circuit.

Figure 6.38 shows the use of IGFET as an amplifier. This circuit corresponds to the CE circuit of a bipolar transistor and is known as the common-source configuration. Other possible configurations are common gate and common drain corresponding to CB and C configurations respectively in bipolar transistors.

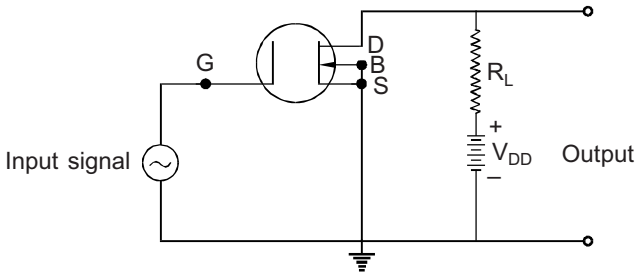


Fig. 6.38 IGFET as an Amplifier

The main advantages of FET are a very high input impedance and its not being very sensitive to temperature effects. Field Effect Transistors, therefore, find wide applications in low power audio amplifiers, electronic timers and test instruments.

6.22.3 Other Semiconductor Devices

Silicon Controlled Rectifier (SCR) A silicon controlled rectifier (SCR) is a four layer semiconductor device having three external connections. These connections are the anode, the cathode and a control connection called the gate. Figure 6.39(a) is a representation of NPNP SCR and Fig. 6.39(b) gives its circuit symbol. The SCR is a solid state rectifier which does not conduct appreciably even when forward-biased till the anode voltage reaches or exceeds a value called the forward break over voltage. Once the forward breakover voltage is reached the SCR is switched on to a highly conductive state and the voltage drop across it comes down to as low a value as one volt. In this highly conducting state, the current in the SCR is limited only by the supply voltage and the load resistance. The flow of current continues till the circuit is intercepted for a brief moment as in a thyratron. In actual practice, the SCR is

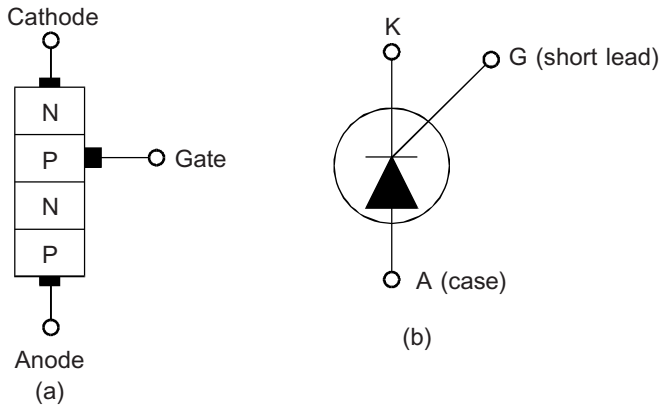


Fig. 6.39 Silicon Controlled Rectifier (SCR) (a) Representation (b) Circuit Symbol

operated at a voltage slightly less than the forward breakover voltage and then it can be switched on by a small trigger pulse applied to the gate. This is a novel method of rectifier operation where low levels of gate current (1.4 V, 30 mA) can control high levels of anode current. SCRs are manufactured in various capacities starting with low current SCRs with less than 1 A anode current to high current SCRs which can pass anode currents of hundreds of amperes. This property of control over powerful currents by means of small Gate pulses makes the SCR useful for many relay, switching and control applications. Typical applications are the lamp-dimmer circuits and the speed control circuits in motors.

The silicon controlled rectifier (SCR) belongs to a family of semiconductor devices called *thyristors* which are generally used in switching and control circuits.

6.22.4 Thyristor

This is a general name for gate controlled rectifiers. Two special types are TRIAC and DIAC.

TRIAC The TRIAC is a bidirectional SCR. There is no anode or cathode in the TRIAC, as current can flow in either direction between terminals 1 and 2.

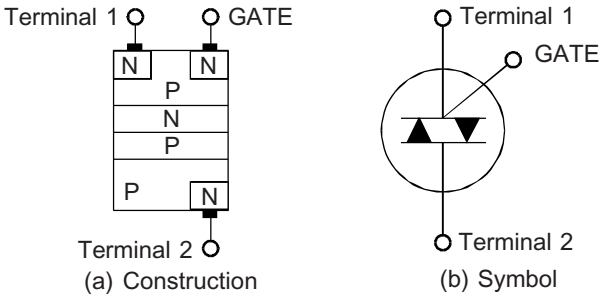


Fig. 6.40 The TRIAC—A Gate Controlled Rectifier

DIAC This is also bidirectional between main terminals 1 and 2 but it does not have a gate: there are just three layers.

6.22.5 Unijunction Transistor (UJT)

UJT has an emitter and two connections to the base, without a collector terminal. UJT is not used as a transistor amplifier but only as a switching device.

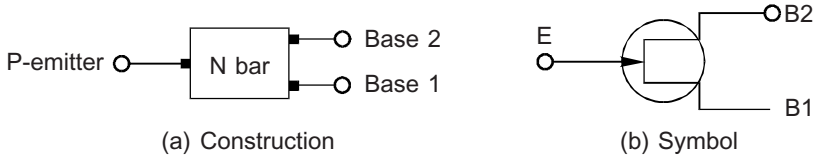


Fig. 6.41 Unijunction Transistor (UJT)

6.23 TESTING A TRANSISTOR

A transistor is a combination of two junction diodes. Each junction can be tested as a diode with a multimeter used as an ohmmeter. The transistor leads are first identified and the ohmmeter leads are connected between the base and the emitter after the ohmmeter has been adjusted for either the $R \times 10$ or the $R \times 100$ range. Resistance reading is noted in this position. The ohmmeter leads to the base and the emitter are then reversed as in Fig. 6.42 and the resistance reading noted again. If the resistance value is significantly higher in one direction than in the other direction, the base-emitter junction can be assumed to be in good condition. The ohmmeter leads are then applied to the base-collector junction and the above procedure is repeated. If any of the above tests show either a direction continuity or an infinite resistance, the transistor has either an internal short or an open junction and should be rejected. Finally, the test is also repeated between the emitter and collector leads when the order of resistances in the two directions will be considerably higher.

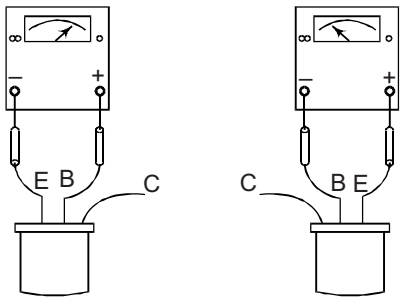


Fig. 6.42 Transistor Testing

It is important to remember that the $R \times I$ range should not be used while testing low power transistors with a multimeter as the high current supplied by the ohmmeter in this position can damage the transistor. If the transistor under

test is already connected in a circuit, it should be tested after isolating it from the circuit by disconnecting the base lead.

Transistor testers are available which can not only test a transistor but can also measure its current gain and other parameters. Most transistor testers can perform either in-circuit or out-of-circuit tests.

6.23.1 Transistor Troubles

Transistor troubles are mainly due to internal short circuit. Although transistor testers are available to test the transistors in and out of circuit, for open circuit, short circuit leakage and β tests. However, open and short circuit tests can be easily performed with a multimeter.

The types of tests that can be performed with a multimeter are

(i) *In-circuit tests:* These tests are performed on a transistor which is already soldered into a circuit. These tests consist of voltage measurements to determine if the junctions are in lact and the transistor is conducting properly.

Check the forward bias by measuring the base to emitter voltage by putting the voltmeter leads directly across their terminals. V_{BE} should be about 0.2V for germanium and 0-6 V for silicon transistors.

If the reading for BE is zero, the base emitter junction is short-circuit. If the V_{BE} is 0.8 V or higher the base emitter junction is probably open.

Substitution by a new and tested transistor is the best way to correct faults in an in-circuit suspected transistor.

(ii) *Out-of-circuit tests:* These tests are meant to test each transistor junction as a diode.

Check the resistance between base and collector and reverse the leads. Do the same thing between base and emitter. The resistance will be very high when the junction is reverse biased very low when the junction is forward biased.

Avoid the use of $R \times I$ scale in the multimeter to avoid damage to the transistor junction particularly in the case of low power transistor.

The actual reading of resistance in these tests depends on the ohmmeter range selected and the type of transistor. Reverse resistance is in front in the case of silicon transistors.

6.23.2 Identifying the Leads of a Bipolar Transistor

If the terminal markings on a transistor are not known and also if the identification or number of the transistor is defaced or erased, it is possible, with an ohmmeter to identify the base, emitter and collector terminals and to find the type of the transistor, i.e. pnp or npn.

(i) Make resistance measurements between each pair of leads both in the forward and reverse bias direction. A low resistance of about 250 ohms shows forward biasing of the junction.

The highest forward bias reading is obtained between the emitter and the collector.

The third lead which is not connected to the ohmmeter is the base lead.

(ii) To identify which of the two unknown leads is collector and which is emitter, it is necessary to know whether the transistor is pnp or npn type. If the ohmmeter indicates forward resistance when the negative lead is connected to the base, the transistor is pnp type. The transistor is npn type if the forward resistance is indicated with the positive lead of the ohmmeter connected to the base.

(iii) Now carefully observe the polarity that gives lower resistance indication.

In the case of pnp transistor the lower resistance indication is obtained with the negative lead of the ohmmeter connected to the collector. The positive lead of the ohmmeter will give lower resistance when connected to the collector in the case of npn type of transistor. Having established the identity of collector lead, all the three leads, viz. base, collector and emitter have been identified.

6.23.3 Transistor Testers

Various types of transistor testers ranging from simple 'good-bad' indicating types to those using programmed testing are available.

The more commonly used battery operated types of testers provide for the measurements of open, shorts, alpha, beta and facilitates for measurement of ac current and leakage currents at various junctions.

The indicators used are either meters or simply a neon-lamp indicator.

6.24 PRECAUTIONS IN THE USE AND HANDLING OF TRANSISTORS

1. Transistors, though rugged and sturdy in construction, can be easily damaged if not handled with care. Small power transistors have flexible leads which are very fragile and can easily break if not handled carefully.
2. Unlike vacuum tubes, transistors are not able to tolerate any over-loads and are instantly damaged. Any accidental shorting of base-collector leads in an operating condition is likely to damage the transistor. Care should be taken to avoid such shorting of leads while taking voltage measurements on a live transistor.
3. Excessive heat can damage a transistor. Soldering a transistors in circuits should be done quickly and the use of low wattage soldering irons of 30 to 35 W rating is recommended.
4. To avoid transistors being damaged due to transients, the power should be switched off before inserting or removing the transistor from the circuit.
5. Voltages applied to the collector and emitter leads should be of correct polarity and should not exceed the specified limits.
6. While testing transistors with a multimeter, $R \times I$ range should not be used to avoid damage to the transistor due to excessive current.
7. While testing with a signal generator, the output from the signal generator should be kept low in the beginning and gradually raised, if necessary. The use of a capacitive coupling in the generator leads is recommended.

SUMMARY

Transistors are replacing vaccum tubes in every field of electronics because of their small size, rugged mechanical construction, long life and non-dependence on heating power. Transistors are made of semiconductor materials whose electrical properties lie between those of conductors and insulators. Germanium and silicon are the most commonly used semiconductors. In their pure crystalline form germanium and silicon are insulators but when doped with certain impurities like antimony or indium, they become semiconductors. Silicon doped with antimony becomes a P-type semiconductor but when doped with indium becomes an N-type semiconductor. In P-type semiconductors, the conduction takes place through positive charge carriers called holes while in N-type semiconductors the charge carriers are electrons.

A P-N junction is formed by combining a P-type material with an N-type material. It is either of the grown type or of the fused type. A P-N junction has the important property of offering a low resistance to current flow in one direction and a high resistance in the opposite direction.

A P-N junction is forward-biased when a positive potential is applied to the P-region and negative potential to the N-region. In this condition, the external potential lowers the barrier potential and allows conduction. When reverse-biased by applying negative potential to the P-region and positive potential to the N-region, it offers a very high resistance. A P-N junction diode which conducts during positive half cycles of AC and does not conduct during negative half cycles is used as a rectifier for conversion of AC into DC and as a detector of modulated waves.

A zener diode is a reverse-biased P-N junction diode which, operated in the breakdown region, is useful as a voltage regulator. A tunnel diode is a P-N junction which presents a negative resistance region for a specific range of forward voltages and is useful as an oscillator at microwave frequencies. Varactor diodes behave like capacitors whose capacitance can be varied by variation of reverse bias. Varistor diodes are used to prevent sudden jumps of voltage in transistor circuits. LEDs and LCDs are diodes which are useful electronic diodes because of their light emitting properties.

A transistor is a three terminal semiconductor device formed by placing two P-N junctions in a back-to-back fashion. When the N-type material is sandwiched between two P-type regions, the transistor formed is a PNP transistor. When the P-type material is sandwiched between two N-type regions the transistor so formed is the NPN type.

Like a diode, a transistor is also constructed either as a grown junction or as a fused junction crystal.

In a transistor, the centre region is called the base and the two outer regions, are called the emitter and the collector respectively. Type numbers are given to transistors to distinguish one type from another and standard methods are used by manufacturers for the identification of the leads of a transistor.

For the operation of a transistor the emitter-base junction is forward-biased and the collector base junction is reverse-biased. Battery polarities used for the operation of PNP transistors are opposite to the polarities for the operation of NPN transistors.

The fact that the base current I_B in a transistor can control the collector current I_C makes it suitable as an amplifier like a triode vaccum tube. In a transistor amplifier circuit, the input signal can be applied between any two electrodes and the output can be obtained between the other two electrodes. There being only three electrodes, there are three possible ways of using a transistor as an amplifier. These are the common base (CB), common emitter (CE) and the common collector (CC) configurations.

The CB configuration provides no current gain, has low input resistance but high output resistance and the output voltage is in phase with the input voltage. The CE configuration is the one most commonly used. It provides both current gain and voltage gain, has a high input resistance and its output voltage is 180° out of phase with the input voltage. It needs stabilisation. The CC configuration known as the emitter-follower circuit, has a voltage gain of less than one and it provides a high input impedance and a low output impedance making it useful as a matching device.

A transistor is a current operated device in which change of current in one circuit affects the current in the other circuit. The ratio of change of collector current to the change of emitter current is called α and the ratio of the change of collector current to the change in base current is called β. α and β are connected by the relation

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

Thermal runaway is the cumulative increase in collector current caused by the leakage current which increases with rise of temperature. Heat sinks and other bias stabilisation methods are used to prevent damage to the transistor due to thermal runaway.

Load line is a straight line on the characteristic curves of a transistor that is used to analyse the behaviour of the transistor.

Q point or operating point of an amplifier is so chosen as to avoid any distortion in the output of the amplifier.

h-parameters for a transistor are the hybrid parameters which are used to relate the four variables of a transistor (v_1, i_1) and (v_2, i_2) are called h-parameters.

Field effect transistors (FETs) are unipolar semiconductor devices whose operation depends on the action of only one type of charge carriers. FETs are of two types, junction FET (JFET) in which the main body consists of N-type or P-type material known as the channel and the other three elements are the source (S), Drain (D) and the Gate (G). In JFET, the drain current can be controlled by the application of reverse bias to the Gate. JFET has a very high input resistance and is useful as a small power amplifier.

In the Metal-Oxide Semiconductor FET (MOSFET) the Gate is insulated from the channel by a layer of silicon dioxide and this is also known as insulated gate FET (IGFET). This type of FET has even higher input resistance than the JFET.

Although rugged in construction, small power transistors have very fragile leads and should be handled carefully. Transistors are also damaged by excessive heat while soldering, a sudden urge of current due to accidental shorting of leads while measuring voltages on transistors in operation.

Review Questions

1. Why are transistors replacing vaccum tubes in the electronic industry?
2. What is doping? How is an N-type semiconductor formed by doping and how does it differ from a P-type semiconductor?
3. What property of a junction diode makes it suitable for rectification? Draw the circuit of a full-wave rectifier using silicon diodes.
4. How is a PNP type transistor constructed from two PN junctions? Draw a diagram indicating the biasing for the proper operation of a PNP transistor.
5. What is meant by thermal runaway in a transistor? Give two methods of preventing thermal runaway in transistors.
6. What property of a transistor makes it suitable as an amplifier? Draw the circuit of a CE configuration with NPN transistor and describe its characteristics.
7. Define α and β and indicate by a formula the relation between the two constants. The characteristics of a transistor indicate that with $V_{CE} = 5$ V, the values of I_C corresponding to I_B values of 50 μA and 75 μA are 4.5 mA and 9.5 mA respectively. Calculate the value of β and α.
   (Ans. $\beta = 200$ $\alpha = 0.9$)
8. What is FET? Describe the working of a MOSFET. Why is it called IGFET?
9. Fill in the blanks:
   (a) When the arrow in a transistor circuit symbol points towards the base, the transistor is _____ type.
   (b) A P-N junction diode is forward-biased when the P-section is made _____ with respect to the N-section.
   (c) The Emitter = Base junction of a transistor is _____ biased and the collector base function is _____ biased.
   (d) The FET is _____ while a junction transistor is bipolar.
   (e) A CC circuit has _____ input resistance and _____ output resistance.
   (f) In a transistor the _____ current controls the _____ current and transistor is a _____ operated device.
12. What precautions are necessary in the use and handling of transistors?
13. What is a load line? How is it used to describe the behaviour of a transistor.
14. What are h-parameters? How are these parameters used to relate the various parameters of a transmitter?

chapter 7

Integrated Circuits

7.1 DEFINITION

Integrated circuits (ICs) are small packages containing dozens of circuit components such as diodes, transistors, resistors and capacitors formed into or mounted on the surface of a block of semiconductor material by special manufacturing processes. These components are arranged or "integrated" into specific circuits on extremely small wafers called "chips". These chips containing several transistor circuits are only tiny bits to begin with but the size becomes relatively bigger when packaged and provided with terminals and leads. Even as a complete package, the size of an IC is comparable with the size of a normal transistor package. Figure 7.1 shows two IC packages of different shapes.

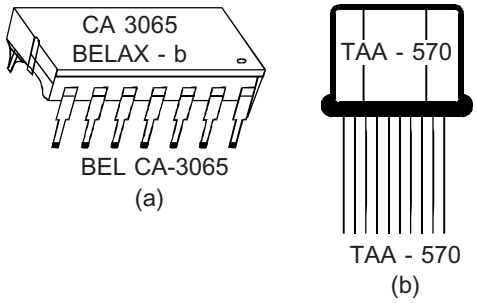


Fig. 7.1 IC Packages (a) BEL 3065 (b) TAA 570

One single IC can replace a number of complex circuits or stages in a TV receiver, radio receiver or in other electronic circuits. This has introduced a new trend of circuit fabrication in the electronic industry. In circuit construction, ICs provide the advantage of extremely reduced size and weight combined with reliability and economy in component and assembly cost. In addition to their use as audio and RF amplifiers in radio and TV circuits, the IC chips are also used in electronic calculators, watches and digital computers.

7.2 CONSTRUCTION OF ICs

ICs are constructed as monolithic ICs and thin or thick film ICs.

Monolithic ICs In this type of IC all the components like diodes and transistors are formed out of a single P or N type wafer called *substrate*. The process is a diffusion process which causes impurities to spread out into given areas of the wafer or the substrate.

Thin or thick film ICs In this type, the components are formed by the deposition of certain materials upon the substrate. In the thin film type, the substrate is ceramic or glass and all the components are formed on this insulating platform by an evaporation process. In the thick-film type, however, *R* and *C* are formed on the substrate itself but transistors are then added as discrete chips.

Hybrid ICs consists of components which are both of monolithic and film type.

7.2.1 Classification

As for their functional use, ICs can be classified as linear or Digital ICs. Linear ICs generally contain circuits operating on the linear portions of the characteristics so that input and output waveforms are similar. Linear ICs are mostly used in linear audio and *RF* circuits. Digital ICs are for pulse circuits employed in calculators and digital computers.

Complete circuit details and applications of a particular IC are generally available in application notes furnished by the manufacturers. Figure 7.2 is a functional diagram of an IC-*CA* 810 used as an audio amplifier for radio receivers. Other applications of the IC include stereo amplifiers, Hi-Fi amplifiers and audio amplifiers for TV receivers.

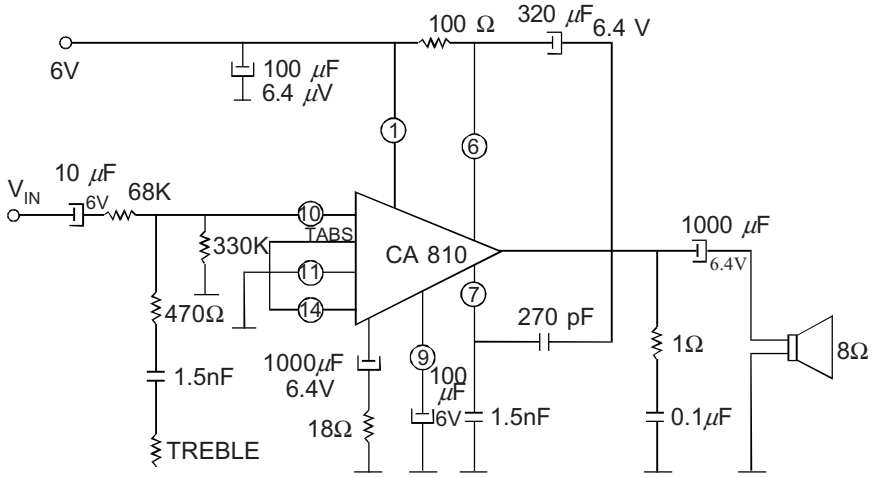


Fig. 7.2 0.5 W Audio Amplifier for a Radio Receiver (Courtesy: BEL)

Linear Integrated Circuits A linear integrated circuit is an IC in which the output is proportional to input. These ICs are mainly designed for dc amplifi-

ers, audio amplifiers, RF amplifiers, IF amplifiers, power amplifiers, differential amplifier etc.

An important class of linear ICs is the operational amplifier (op-amp).

7.2.2 Operational Amplifiers (Op-amps)

Basically an operational amplifier is a high gain direct coupled amplifier. These amplifiers were originally used in analog computers to perform certain mathematical functions such as addition, subtraction, integration and differentiation and hence the term operational amplifier was used. However, these days the operational amplifier refers generally to a high gain amplifier together with its feedback circuits.

An operational amplifier is a complete amplifier circuit constructed as an IC on a single silicon chip. Inside the package there are a number of transistors and other components integrated to form a single unit. Monolithic construction leads to a very small size, balancing of temperature effects, reduced cost and improved reliability by elimination of much of inter-connecting wiring etc.

Characteristics of an op-amp The following are the main characteristics of an ideal op-amp. 1. Infinite gain 2. Infinite input impedance 3. Zero output impedance 4. Infinite bandwidth 5. Zero rise and fall times for Quiescent point, zero voltage output level for a zero input level. 6. Infinite common mode rejection (CMMR) 7. Insensitivity to ageing 8. Insensitivity to power supply voltage variations.

7.2.3 Typical Applications of op-amps

(i) Notation Symbolically an op-amp is represented by a triangle as shown in Fig. 7.3(a).

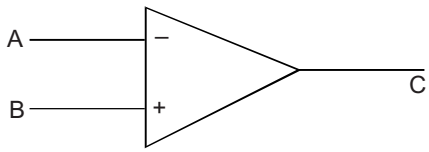


Fig. 7.3 (a) Symbol for an op-amp

The triangle indicates the direction of signal flow. A and B are the signal input connections and C the output signal connection. The (–) minus input is the inverting input and (+) plus input is the non-inverting input.

Identification: The following information is generally indicated on the package of the op-amp.

(i) Device type—μ A710
(ii) Package type—T (Mini DIP-numero)
(iii) C—Temperature

Package Type—generally represented by one letter:

D — Dual-in-line package (Herrnatic ceramic)
F — Flat pack
H — Metal can package
J — Metal power package (TO-66 outline)
K — Metal power package (TO-3 outline)
P — Dual-in-line package (Moulded)

R — Min-DIP (Hermeic Ceramic)
T — Min-DIP (Moulded)
U — Power package (Moulded-To-220 outline)

Temperature range

Three basic temperatures ranges in common use are

C — Commercial– 0 to + 70°C
M — Military – 55 to + 85°C
V — Industrial – 20 to + 85°C
– 40 to + 85°C

EXAMPLE: **μ A 725 HC**

Indicates a μ A 725 type operational amplifier in a metal can package with a commercial temp. range capability (0 to 70°C).

7.2.4 Practical Applications

(a) As a Summing Amplifier Let the voltages V_1, V_2, V_3, etc. be applied to the inverting input terminal of the op-amp as shown in Fig. (7.4).

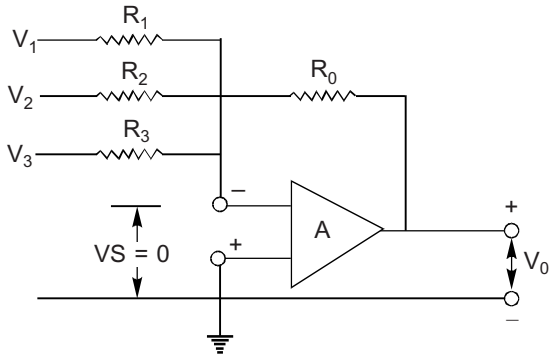


Fig. 7.4 Summing Amplifier

It can be shown that

$$V_0 = -\left(\frac{R_0}{R_1} V_1 + \frac{R_0}{R_2} V_2 + \frac{R_0}{R_3} V_3\right) \tag{7.1}$$

If $R_1 = R_2 = R_3 = R_0$, then $V_0 = -(V_1 + V_2 + V_3)$ and the virtual ground is really the summing point for all input circuits.

This circuit is very useful in analog computers and can be used to solve linear equations.

(b) As an Inverting Amplifier If A is the closed loop gain of the Amp, then

Given $$A = \frac{V_0}{V_1} = \frac{-R_2}{R_1}$$

This is an important fact that the gain of an inverting op-amp depends only on the feedback resistors R_2 and R_1. The minus sign only shows that the output signal is inverted i.e. 180° out of phase with the input signal.

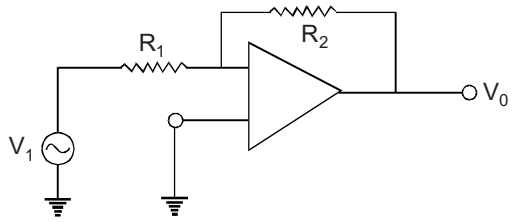


Fig. 7.5 Inverting amp

If $R_1 = 3\ k$ and $R_2 = 30\ k$

then $$A = \frac{30k}{3k} = 10$$

(c) As a Non-inverting Amplifier The closed loop gain A of the non-inverting amp. is given by.

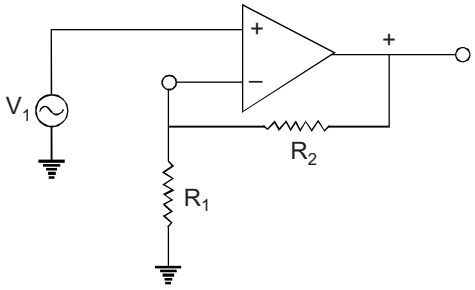


Fig. 7.6 Non-inverting amp

$$A = 1 + \frac{R_2}{R_1}$$ and if $R_2 = 30\ k$ and $R_1 = 3\ k$, then

$$A = 1 + \frac{30}{3} = 1 + 10 = 11$$

(d) As an Integrator An integrator is basically a low pass filter whose output is proportional to the product of the amplitude and duration of the input. In other words it denotes the area under the voltage-time curve of the output signal.

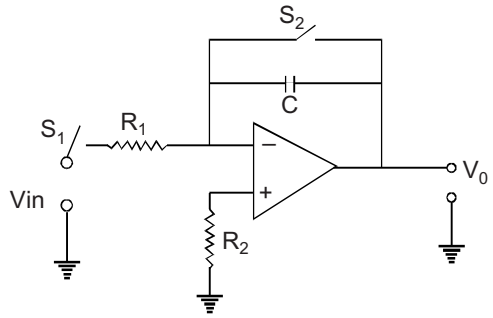


Fig. 7.7 Integrator

In the circuit shown for integrator

$$V_0 = -\frac{1}{RC}\int v_1 \cdot dt$$

(e) As a Differentiator A differentiator circuit shown in Fig. 7.8 produces an output signal which is proportional to the rate of change of the input signal.

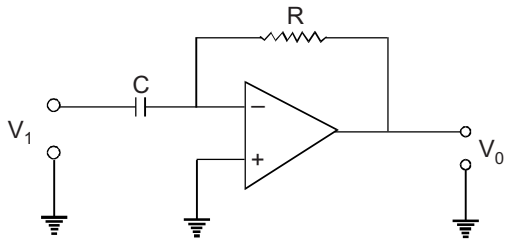


Fig. 7.8 Differentiator

(f) As Active Filter An op-amp is widely used as an active filter element because reasonably valued capacitors and resistors may form filters operating at frequencies as low as 0.01 Hz.

(g) Low Pass Filter Figure 7.9 shows a low pass filter which will start roll-off at 12 dB/octave. The high pass filter can be easily formed by inter changing the capacitors and resistors as shown in Fig. (7.10).

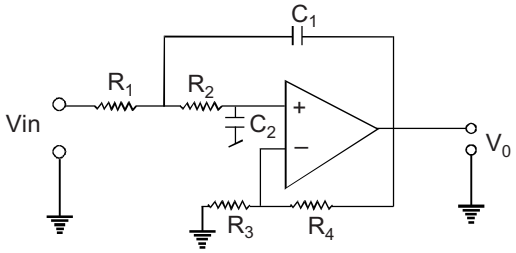


Fig. 7.9 Low Pass Filter

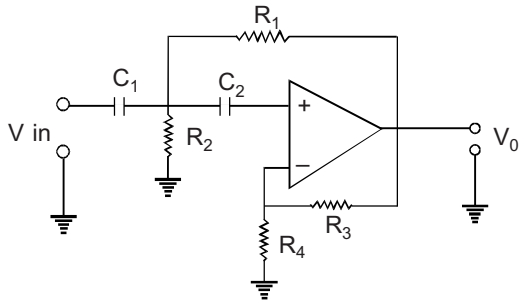


Fig. 7.10 High Pass Filter

Figure 7.11 shows a multiple feedback band pass filter. Filter bandwidth is a function of Q.

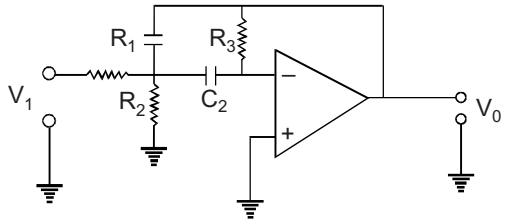


Fig. 7.11 Bandpass Filter

$$\text{BW (bandwidth)} = \frac{f_0}{Q}$$

Where f_0 = resonant frequency
BW = Bandwidth at 3dB point

(h) As a Voltage Regulator The circuit is shown in Fig. 7.12. The regulated output voltage may be varied by varying R_2 or R_1.

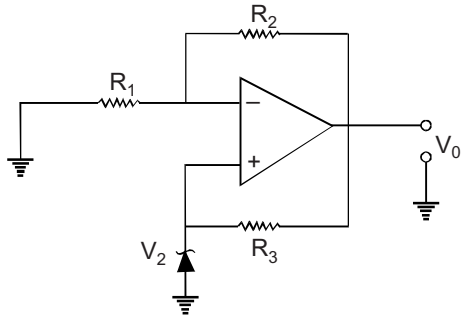


Fig. 7.12 Voltage Regulator

If the load current is more than the output capability of the op-amp, a transistor used as an emitter follower can be connected at the output.

(i) As Comparator The comparator circuit of Fig. 7.(13) is used to switch at any reference level. The output can be made to switch from plus saturation to minus saturation and vice versa with less than one mV change across its input.

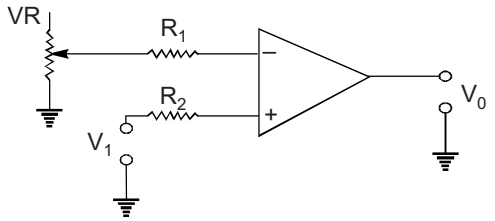


Fig. 7.13 Comparator

(j) As a Divider op-amp as a divider is shown in Fig. 7.14.

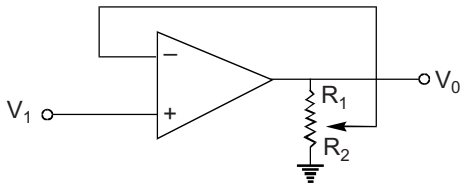


Fig. 7.14 Divider Circuit

$$V_0 = \frac{V_1}{\text{Divider ratio}} \text{ where}$$

$$\text{divider ratio} = \frac{R_2}{R_1 + R_2}$$

(k) As a Square Wave Generator The circuit shown in Fig. 7.15 generates an output voltage of square waveform.

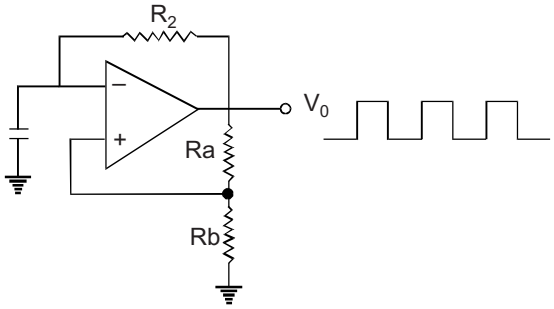


Fig. 7.15 Square Wave Generator

The frequency f of the square wave is given by

$$f = \frac{1}{2R_2C} \text{ for } \beta = 0.473$$

where $$\beta = \frac{R_b}{R_a + R_b}$$

(l) As a Sine-Wave Generator Figure 7.16 is simple Wien Bridge circuit that produces a sine wave whose output amplitude is controlled by the potentiometer R_a with $R_1 = R_2$ and $C_1 = C_2$, then frequency of oscillation f is given by

$$f = \frac{1}{2\pi R_1 C_1}$$

Diodes can be used to reduce distortion in the output.

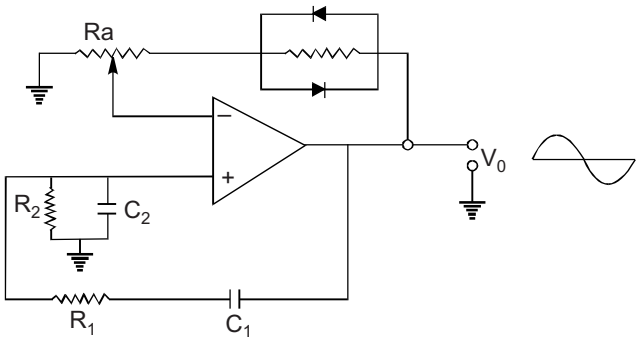


Fig. 7.16 Sine Wave Generator

7.3 DIFFERENTIAL AMPLIFIER (DA)

A Differential Amplifier is a direct coupled linear IC amplifier which finds wide applications in measuring instruments such as oscilloscopes and recording instruments where flat response from dc to megahertz range is required. It is called a differential amplifier or difference amplifier because any of its available outputs is essentially proportional to the difference between the two input signal voltages.

Figure 7.17 gives the basic schematic diagram of a differential amplifier.

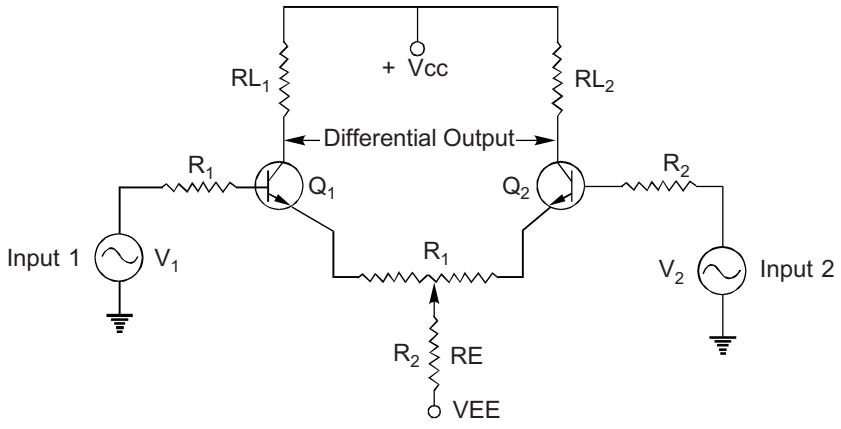


Fig. 7.17 Differential Amplifier

The basic circuit consists of two NPN transistors Q_1 and Q_2 with two inputs and a single output. The circuit is symmetrical, i.e. the two transistors Q_1 and Q_2 have identical characteristics.

The collector load resistors R_{L1} and R_{L2} are equal and the emitter resistor R_E is common to both transistor. The two input circuits are also identical i.e. $v_1 = v_2$ and $R_1 = R_2$.

There are several modes or parallel combinations of input and output signals but two most common modes are (i) common mode operation (ii) differential mode operation. These two modes of operation will be discussed now.

(i) Common Mode Operation The output signal is proportional to the difference between the two input signals i.e.

$$v_{out} = A\ (v_1 - v_2) \tag{7.2}$$

Where A is the gain of each transistor and v_1 and v_2 are the input signal voltages relative to ground. When the two input signals are in the same phase and the amplitudes are equal, then

$$v_1 - v_2 = 0 \tag{7.3}$$

and

$$v_{out} = A\ (v_1 - v_2)$$
$$= A\ (0) = 0 \tag{7.4}$$

This means in the common mode the differential amplifier rejects the common mode signal and the output as common mode is zero.

(ii) Differential Mode Operation In this mode the two input signals are equal in amplitude but opposite in phase (180°) i.e.

$$v_1 = -\ v_2 \text{ or } v_2 = -\ v_1$$
$$v_{out} = A\,[(v_1 - (-v_2)] = A\,(2v_1) \tag{7.5}$$

Thus, in the differential or non common mode of operation the DA gives an output which is equal to twice the gain (A) times the input signal.

Common-Mode Rejection Ratio (CMMR) The advantage of the DA is that it has no output for the input signals applied to Q_1 and Q_2 in the same phase i.e. in common mode. As an example any variations in the supply voltage will be applied in the common mode and will be rejected. In other words, the circuit amplifies differential signals but rejects common mode signals.

The effectiveness of a DA (i.e. gain for differential signals and rejection for common mode signal) is expressed by a factor called Common-Mode Rejection Ratio (CMMR).

$$\text{CMMR} = \frac{A_d}{A_c} = \frac{\text{Differential gain}}{\text{Common mode gain}}$$

when

A_d = Differential gain

A_c = Common mode gain

The CMMR value is therefore an index, the effectiveness of a DA, the higher the ratio the better the DA.

Common-Mode Rejection Ratio (CMRR) is sometimes expressed in dBs. Typical values of CMRR for op-amps. are 80-100 dBs.

The degree of common mode rejection depends on the balance of two stages.

Off-Set Voltages This is the amount of differential voltage output with no input. The off-set should ideally be zero but if the stages are not perfectly balanced, the variations of temperature and pick up of stray signals may also cause off set.

7.4 APPLICATIONS OF LINEAR ICs

Linear ICs are mostly used in analog circuits. An analog signal has continuous variations corresponding to the desired information. Sine wave is an example

of an analog signal. On the other hand pulse or digital circuits operate between the two discrete states of 'off' and 'on'.

Some of the important applications of linear ICs are given below.

(i) Audio amplifiers—These are used in pairs in stereo amplifiers.
(ii) IF amplifier (Intermediate Frequency Amplifiers) used in AM, FM and television receivers using envelope detection for 4.43 MHz or 3.58 MHz). As sub carrier regeneration amplifiers or demodulators for colour TV receivers.
(iii) Operational amplifiers—this has been described in detail in this chapter.
(iv) Voltage regulators
(v) Choppers—To chop a dc waveform into ac segments for easier amplification and then convert back into the dc form.

In addition separate IC chips are available for electronic calculators, electronic watches, video game units and for TV receivers etc.

7.5 FABRICATION OF A MONOLITHIC IC

Let us consider the construction of a simple circuit of the type shown in Fig. 7.17 (a) on a single chip of silicon as shown in Fig. 7.18 (b).

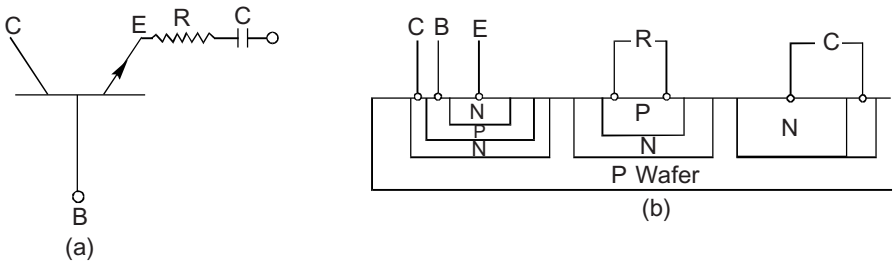


Fig. 7.18 (a) Circuit (b) Construction with Transistor (NPN) and R and C

The steps taken in the construction of the above IC are briefly given below.

(a) The silicon crystal is first cut into thin wafers about 10 ml. thick and having a diameter of about 1.5 to 2 in as shown in Fig. 7.19.

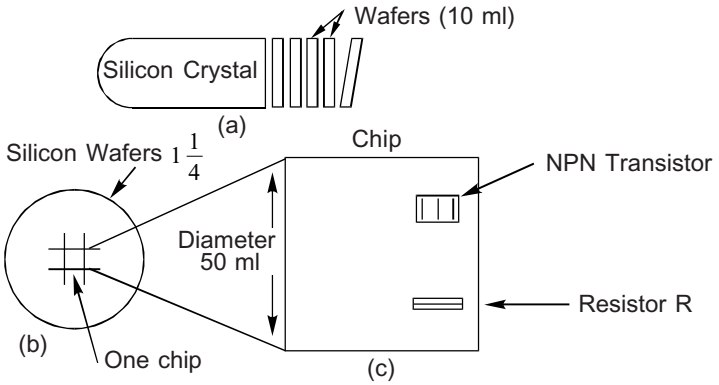


Fig. 7.19 (a) Silicon Crystal, (b) Slice, (c) Chip Manufactured with Components

The steps described are not for just one chip, but these steps simultaneously produce as many as 625 chips into which the wafer is cut and each chip is mounted in a separate package.

(a) Each wafer is polished to a mirror finish by acid elching and a thin layer of silicon dioxide (SiO_2) is formed on the wafer by oxidizing the wafer in dry oxygen.

(b) The thin glass protects the silicon surface and serves as a barrier to the doping of semi-conductor junctions.

(c) A thin uniform coating of a chemical called photoresist is now deposited on the SiO_2 layer. This is a liquid that will harden when exposed to ultraviolet light through a mask placed over the photoresist. The oxide coating is opened in a window pattern by photochemical techniques to allow 'doping' when necessary. The wafers are now ready for diffusion process.

(d) The silicon wafers are now heated in a furnace containing a gaseous boron (acceptor) atmosphere. The P impurities diffuse into silicon turning the N-type material into a P-type channel. This will result in the N-type material resisting on P-type substrate under the SiO_2 layer.

(e) The P-type base of the transistor is now diffused into the collector using the same masking process as described above. This also results in the diffusion of the resistor taking place in the adjoining island.

(f) The N-type emitter is now diffused into the base following the same diffused photoresists and masking process.

(g) Finally, metal contacts with different diffused areas are provided again following the same SiO_2 layer, photoresist and marking process to provide access to different diffused layers. Final shape of the wafer is shown in Fig. 7.20.

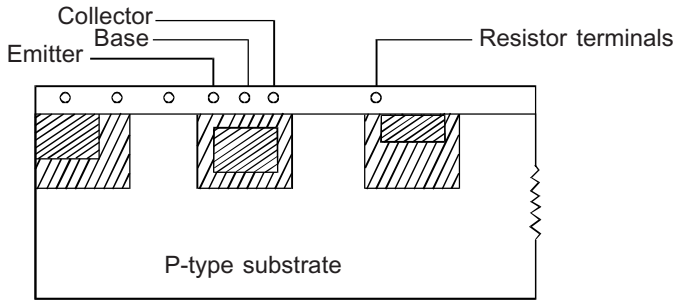


Fig. 7.20 The Cross-section of the Wafer before Metalization

Metallic paths needed to inter-connect components in the IC are provided by aluminium printed wiring. These terminals are at the edges of the chip where the wires are bonded for connection to the external leads. These chips are finally tested by a computer controlled process and defective chips rejected.

The chips are separated by cutting the wafer with a thin diamond point similar to the one used for cutting glass.

This process of separating the wafer into individual chips is called "DICING".

Each chip is then mounted in a plastic case or package which may be in the form a round TO-5 case with 8 terminals or Dual-in-line form (DIP) with its terminal as shown in Fig. 7.21.

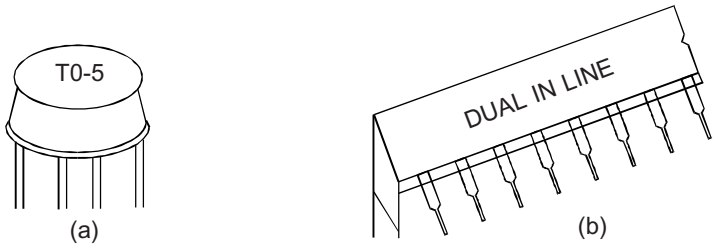


Fig. 7.21 (a) TO-5 8-terminal Package, (b) Dual-in-line Sixteen Terminals Package

The terminal leads on the package are connected to the internal electrodes on the chip with 1.5 ml wire of either aluminium or gold. Connections are made either by ultrasonic bonding or thermo-compression bonding.

7.5.1 Fabrication of Integrated Circuit Components

The integrated components fabricated on the IC chips include transistors, diodes, resistors and capacitors. The advantages of integrated components are miniaturization and matched characteristics for the components.

Integrated Transistors (Bipolar) NPN construction as described above is the most common for bipolar transistors. The P substrate forms a reverse diode with the P collector. Also the P substrate could be the collector for PNP transistor with the adjacent N and P layers. The combination, then could form a complementary pair of NPN and PNP transistors.

Integrated Diodes An N-type diffusion into the P substrate can provide diode junction where required. This is essentially the same as the emitter-base junction on a transistor. Therefore, the diode junction can also be provided by joining transistor collector to base, serving as the anode.

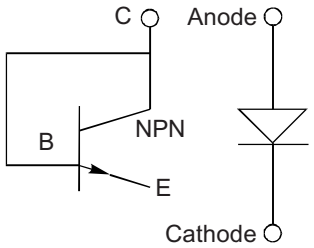


Fig. 7.22 NPN Transistor Connected as a Diode

Integrated Resistors Resistors are obtained utilizing the resistivity of the diffused films of metal laid on a substrate. The values obtained are describe by the term "sheet resistance". For a given thickness we can refer to "ohms per square" of the sheet. Typical values run from 50 to 250 ohms per square" of sheet depending on the thickness of P type diffused material and from 50 to 10 K ohms per square for thin films. The tolerance on monolithic resistors is ± 30 to ± 50 per cent.

Integrated Capacitors There are two types of integrated capacitors, the junction type and the MOS type. In the junction type the capacitance is across the depletion zone of a PN junction with reverse bias. Typical values are 0.1 to 0.1 pf per square ml with reverse bias of 5V. If an area of 100 square ml is used out of 2500 sq ml on a chip the C value will be 10 to 40 pF.

For the MOS type of capacitor the N^+ layer forms the botton plate, the SiO_2 insulating layer is the dielectric and the aluminium metalizer serves as the opposite plate as in Fig. 7.23.

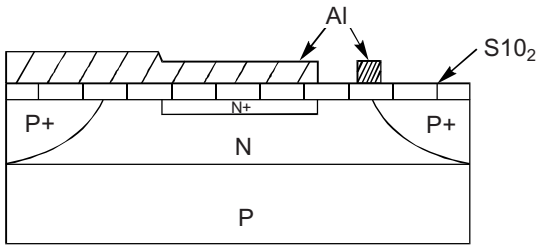


Fig. 7.23

The capacitance value of 3 to 30 pF depends just on the dielectric of the plate area. The + sign or N+ or P+ indicates heavy doping.

Supply Voltage for ICs The linear IC units use 15 to 40 V as typical value of voltage with 3 to 15 volts for digital IC units with monolithic construction. Usually, positive polarity is required for collector voltage on NPN transistors.

SUMMARY

Integrated Circuits or ICs are small packages containing dozens of components such as diodes, transistors, resistors and capacitors which are mounted or integrated in specific circuits by a special manufacturing process on the surface of small wafers called 'chips'. ICs are constructed as monothilic ICs and or thin film ICs. Hybrid ICs consist of components which are both of monothilic and film type.

ICs can be classified as linear and digital ICs. Linear ICs are mostly used in linear audio and RF circuits, and digital ICs are employed in pulse circuits.

Typical example of linear IC is an operational amplifier (op-amp). Its important applications are as a summaring amplifier, inverting amplifier, Non-inverting amplifier, integrator, differentiator, active filter, low pass filter, voltage regulator, comparator, devider, square wave generator and sine wave generator.

A differential amplifier (DA) is a direct coupled linear IC amplifier which finds wide applications in measuring instruments.

Fabrication of monolithic ICs consists of a number of steps like cutting the silicon crystal into the wafer, polishing, coating with photoresist material, doping, healing where necessary. Finally these are mounted into plastic packages of different shapes and sizes.

Review Questions

1. What is an IC? How are ICs useful in electronic Circuits.
2. Distinguish between Monolithic and thin or thick film ICs.
3. How are ICs classified, give examples of each class.
4. What is an Operational Amplifier? Give its main characteristics.
5. Described an Op Amp as an (i) Invertor (ii) Differentiator and (iii) Integrator.
6. Describe a Differential Amplifier. What are its two common modes of operation.
7. Give briefly the various steps taken in the construction of a monolithic IC.
8. Describe briefly the process of fabrication of circuit components on IC Chips.
9. Answer True or False.
   (a) Digital ICs are used in computers.
   (b) Differential Amplifier is an RC coupled amplifier for use in measuring instruments.
   (c) CMMR is an indication of the effectiveness of a DA.
   (d) A linear IC is better suited for pulse circuits than a digital IC.
   (e) Digital ICs use higher voltage than linear ICs.

   Ans. (a) True (b) False (c) True (d) False (e) False.

chapter 8

Digital Electronics

8.1 ANALOG AND DIGITAL SIGNALS

There are two types of signals in electronics and communication system. The Analog signals and the Digital signals.

In analog signals the value of the signal varies continuously over its full range from its minimum to its maximum value. The word analog indicates a similarity (analogy) between a real thing like a voltage, current and frequency and the representation of the real thing. Such quantities are generally read by the swing of a needle or a pointer on an analog meter. Figure 8.1(a) indicates the waveform of an analog signal.

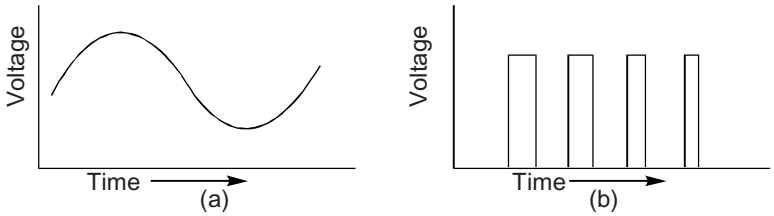


Fig. 8.1 (a) Analog Signal, (b) A Typical Digital Signal

A digital signal is represented by discrete numbers and has only two levels 0 and 1 and also sometimes known as low and high levels. The word digital has its origin from the digits of fingers which have been used by man from early times for making simple arithmetical calculations like addition and subtraction etc. A digital signal is shown in Fig. 8.1(b) and can be represented by an electronic switch which has either an on position or an off position.

Digital meters provided with read out devices are much simpler compared to analog meters which always introduce an element of personnel error in the value read by different persons.

Digital signals have the advantage that these can be used for communication over long distances without being distorted by noise and other interference. Morse signals were used originally to avoid noise and distortion.

A common use of digital circuits is in counting devices. These applications include the important fields of computers, electronic calculators, clocks and various types of test equipments.

The reduced cost of integrated circuits (ICs) which are widely used in digital electronic circuits explains the wide use of digital voltmeters, digital panel meters and other read out devices in recent years.

The important principles involved in the use of panel digital circuits will be discussed in this chapter.

8.2 NUMBER SYSTEMS

Digital techniques are based on counting system using discrete units. Various number systems are employed in digital electronics of which the most important are the Decimal and the Binary systems.

8.2.1 Decimal System

The digits used by various systems are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The number of digits used by any system is called the base or the radix. For example the decimal system makes use of all the ten digits from 0 to 9 and it has a radix 10. The base or the radix is generally indicated by a subscript as in $(324)_{10}$. Although the decimal system has only 10 digits, it can be used to represent any number or any magnitude by taking into account the positional value of each digit.

$$(324)_{10} = 300 + 20 + 4$$
$$= 3 \times 10^2 + 2 \times 10^1 + 4 \times 10^0$$

It will be seen that for numbers greater than 1 the first place to the left of the decimal point is for count of digits alone, from 0 to 9. The second place is for count of 10, the next place is for count of 100 (10^2) and further next for 1000 $(10)^3$. Places to the right of the decimal point can be used to represent fractions less than 1, then it-will decrease in multiples of 1/10 or 10^{-1}.

For example $(3472.53)_{10}$ is represented as

$$(3472.53)_{10} = 3000 + 400 + 70 + 2 + 05 + 03$$
$$= 3 \times 10^3 + 4 \times 10^2 + 7 \times 10^1 + 2 \times 10^0 + 5 \times 10^{-1} + 3 \times 10^{-2}$$

8.2.2 Binary System

This system makes use of only two digits 0 and 1 which represent the two levels of a binary system. The radix or base is 2 as in $(1100)_2$. As in the decimal system, the binary system can be used to represent any number. The first place to the left of the binary point can be only 0 or 1. The second place is in counts of 2^1, the third place in counts of 2^2 with successive places for counts of 2^3, 2^4 and so on.

$$(1100)_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0$$
$$= 8 + 4 + 0 + 0$$
$$= (12)_{10}$$

As in decimal system, places to the right of the binary point can be used for fractions less than 1. The values, however decrease in multiple of 1/2 instead of 1/10.

For example $(101.11)_2$ can be represented as

$$(101.11)_2 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 1 \times 2^{-2}$$
$$= 4 + 0 + 1 + 0.5 + 0.25$$
$$= (5.57)_{10}$$

Thus the binary number $(101.11)_2$ is the equivalent of decimal number $(5.75)_{10.}$

8.2.3 Binary to Decimal Conversion

The digits in decimal as well as binary system have positional values or weights. In the decimal system the weights are units (10^0), tens (10^1), hundreds (10^2), thousands (10^3) and so on. The sum of all digits multiplied by the positional weights gives the total amount being represented.

In the binary system only two digits 0, 1 are used and the positional weights are powers of 2 instead of powers of 10 as in $2^{0.}$ (units), 2^1 (twos), (2^2) fours, (2^3) eights and (2^4) sixteens and so on.

The decimal equivalent of binary number is the sum of all binary digits multiplied by the weights (2^0, 2^1, 2^3 etc).

Table 8.1(a) shows the weights for a few decimal numbers and (b) binary numbers.

Table 8.1

5	7	0	3	4	digits
10^4	10^3	10^2	10^1	10^0	weights

(a) Decimal weight

1	1	0	0	1	digits
2^4	2^3	2^2	2^1	2^0	weights

(b) Binary weight

As in Table 8.1(a).

$$5 \times 10^4 + 7 \times 10^3 + 0 \times 10^2 + 3 \times 10^1 + 4 \times 10^0 = 57034$$

In the same way from (b)

$$1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 16 + 8 + 0 + 0 + 1 = 25$$

So binary $(11001)_2$ is equivalent to decimal 25. For conversion of binary to decimal the following steps may be followed.

1. Write the binary number.
2. Write the binary weights 1, 2, 4 and 8 under the binary digits.
3. Cross out any weight under a zero.
4. Add the remaining weights.

EXAMPLE 1: Convert binary $(1101)_2$ into decimal

1101 ← binary number
8421 ← weight
8401 ← Multiply
841 ← Cross out zero
8 + 4 + 1 = 13 ← Add

so $(1101)_2$ is equivalent of 13.

EXAMPLE 2: Convert $(111)_2$ into decimal

111 ← binary
421 ← weight
421 ← multiply
4 + 2 + 1 = 7 Add
so $(111)_2 = 7$

EXAMPLE 3: Convert $(1100)_2$ into decimal

1100 ← binary
8400 ← weight
8 + 4 + 0 + 0 = 12 Add

Binary Fractions: The digits to the right of the binary point have the positional weight of $\frac{1}{2}(2^{-1})$, $\frac{1}{4}(2^{-2})$, $\frac{1}{8}(2^{-3})$ and so on

Binary point.

1/2	1/4	1/8	1/16	–1.32
2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}

Thus 0.101 has a decimal equivalent

0.101
0.1/2 + 1/4 + 1/16 = 0.5 + 0 + 0.125 = 0.625

EXAMPLE 4: Convert binary $(1011.11)_2$ into decimal

1011.11
8421.1/2 1/4
$8 + 0 + 2 + 1 + 0.5 + 0.25 = (11.75)_{10}$

8.2.4 Decimal to Binary Conversion

The simplest method to convert decimal numbers into binary is by successive divisions by 2 and writing down the quotient and its remainder. The remainders when read from bottom to top gives the binary equivalent of the decimal number. The method is explained below.

EXAMPLE 1: Convert $(13)_{10}$ into its binary equivalent

2 | 13		
2 | 6	remainder 1	↑
2 | 3	remainder 0	↑
2 | 1	remainder 1	↑
0	remainder 1	

$(13)_{10} = (1101)_2$

for check up

1101 = 8 + 4 + 0 + 1 = 13

EXAMPLE 2: Convert $(43)_{10}$ into its binary equivalent

2	43	
2	21	remainder 1
2	10	remainder 1
2	5	remainder 0
2	2	remainder 1
2	1	remainder 0
	0	remainder 1

$$(43)_{10} = (101011)_2$$

Check up. $101011 = 2^5 \times 1 + 2^4 \times 0 + 2^3 \times 1 + 2^2 \times 0 + 2^1 \times 1 + 2^0 \times 1$

$= 32 + 0 + 8 + 0 + 2 + 1 = 43$

The two sides of the equation balance.

As for the fractional part of the binary number, we multiply by 2 and record a carry in the integer portion as below. Carries taken in the downward direction are the binary fraction as shown below:

Convert 0.625 into binary fraction

$0.625 \times 2 = 2.5$ with a carry of 1

$0.25 \times 2 = .50$ with a carry of 0

$0.50 \times 2 = 1.0$ with a carry of 1

By taking the carries downward we have

$$0.625 = .101$$

Also $$.101 = \frac{1}{2} \times 1 + \frac{1}{4} \times 0 + \frac{1}{8} \times 1$$

$$= 0.5 + 0 + .125 = 0.625$$

8.3 BINARY ARITHMETIC

The arithmetic processes of addition, subtraction, multiplication and division are also performed with binary numbers but using only two digits 0 and 1 as explained below.

8.3.1 Binary Addition

The following rules apply to binary addition

Rule 1 $0 + 0 = 0$

Rule 2 $0 + 1 = 1$

Rule 3 $1 + 0 = 1$

Rule 4 $1 + 1 = 10$ (This is to be read as 0 with carry 1)

EXAMPLE 1: Add 101 and 110

$$\begin{array}{r} 101 \\ 110 \\ \hline 1011 \end{array}$$

In the first column $1 + 0 = 1$, in the second column $0 + 1 = 1$, in the third column $1 + 1 = 10$, i.e. 0 with carry 1.

EXAMPLE 2: Add $(1011)_2$ and $(110)_2$

$$\begin{array}{r} (1011)_2 \\ (110)_2 \\ \hline (10001)_2 \end{array}$$

8.3.2 Binary Subtraction

The following rules apply to binary subtraction

Rule 1 $0–0 = 0$

Rule 2 $1–0 = 1$

Rule 3 $1–1 = 0$

Rule 4 $10–1 = 1$

EXAMPLE 1: Subtract $(101)_2$ from $(111)_2$

$$\begin{array}{r} 111 \\ -\ 101 \\ \hline 010 \end{array}$$

This can also be checked up as below

$$\begin{array}{lll} 111 & \text{is} & 7 \\ 101 & \text{is} & -\ 5 \\ 010 & \text{is} & 2 \end{array}$$

The process of subtraction sometimes involves a lot of borrowing from adjacent columns and the results can be negative also, when numbers of larger magnitude are subtracted from numbers of smaller magnitude as in the following example.

EXAMPLE 2: Subtract $(111)_2$ from $(100)_2$

$$\begin{array}{rl} 100 & \text{first col. } 1 - 0 = 1 \\ -111 & \text{second col. } 1 - 0 = 1 \\ -\ 011 & \text{third col } 1 - 1 = 0 \end{array}$$

Here we subtract the number of larger magnitude from the smaller one and prefix the –sign in the result or check up $(110)_2$ is 4 and $(111)_2$ is 7, so the difference is 3 with a minus sign (–3).

8.3.3 Binary Multiplication and Division

Rules for binary multiplication

Rule 1 $0 \times 0 = 0$

Rule 2 $0 \times 1 = 0$

Rule 3 $1 \times 0 = 0$

Rule 4 $1 \times 1 = 1$

EXAMPLE 1: Multiply $(111)_2$ by $(101)_2$

$$\begin{array}{r} 1\ 1\ 1 \\ \times\ 1\ 0\ 1 \\ \hline 1\ 1\ 1 \\ 0\ 0\ 0 \\ \hline 1\ 1\ 1 \\ \hline 1\ 0\ 0\ 0\ 1\ 1 \end{array}$$

Check $(111)_2$ is 7 and $(101)_2$ is 5

$7 \times 5 = 35$ and $(100011)_2$ is $1 + 2 + 0 + 0 + 0 + 32 = 35$

Binary division is similar to division of decimal numbers.

EXAMPLE 2: Divide $(1100)_2$ by $(10)_2$

```
      1 1 0
   ---------
1 0 | 1 100
      1 0
    -------
      1 0
      1 0
    -------
        0 0
```

Check: $(1100)_2$ is 12 and 10 is 2 and hence 12 divided by 2 is 6 i.e. 110 and the remainder is zero.

8.4 LOGIC AND SWITCHING CIRCUITS

A GATE is a circuit with one or more input signals but only one output signal. An example of a logic gate is the circuit shown in Fig. 8.2. This circuit consists of two switches A and B in series which are used to light up a bulb with the help of a battery V. The bulb will light up only when both the switches A and B are closed.

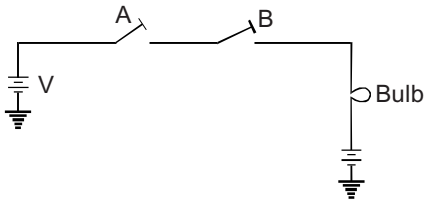


Fig. 8.2 Logic Gate

The bulb will not light up if either A or B or both switches are open. The logic gate formed by these two switches with two inputs and one output, which is the condition for lighting the bulb, is called the AND gate or function. The relationship between input and output is a logic function. When symbols like A and B are used to represent various switching conditions, the system is called symbolic logic. This symbolic logic represents a special type of mathematics called Boolean algebra named after its inventor George Boolean, an English mathematician of eighteenth century. Practical use of the Boolean algebra was, however, made by Claude Shannon of Bell Telephone Laboratories who made use of the symbolic logic for simplifying telephone switching circuits. In Boolean algebra a variable can be either 0 or 1 this means that a signal voltage can be either low or high.

The basic logic functions or gates are AND and OR with their inverted or negative functions NAND and NOR. These will now be described.

8.4.1 AND Gate

A symbolic *representation* of AND gate is given in Fig. 8.3.

This represents a symbolic function with two inputs A and B and the output represented by $A \cdot B = A \times B = AB$ which in Boolean algebra indicates multiplication of $A \times B$ and not addition of signals representing A and B.

In actual practice diodes and transistors are used to represent these symbolic functions. Figure 8.4 shows the electronic circuit of an AND gate.

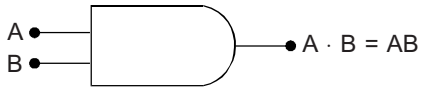


Fig. 8.3 Symbol for AND Gate

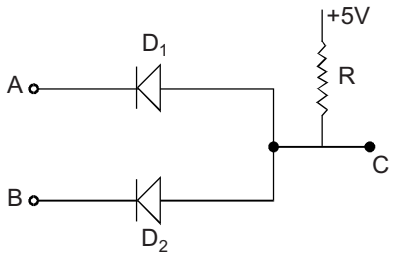


Fig. 8.4 Electronic Circuit of AND Gate

This circuit with two inputs and one output makes use of two diodes D_1 and D_2 which act as switches which are ON when forward biased and off when reverse biased.

In this circuit the input can be either low (ground) or high (+5V). When both inputs are low both diodes conduct and the output voltage is low due to current flowing through R. If one of the inputs is low and the other high, the diode with the low input conducts and this brings the output to a low value. The diode with the high input is reverse biased or cut off.

When both inputs are high both diodes are cut off, there is no current in the resistor and the output C is high. More than two inputs can also be used for one output. This will require the use of 3-diodes.

The action of a logic gate can be represented by a TRUTH Table which summarises the various possible combinations of input signals to a logic gate with the resulting output for each combination. In these tables binary 0 stands for low voltage and binary 1 stands for high input. A TRUTH Table for AND gate is given below.

Table 8.2 Truth Table for AND Gate

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

All possible combinations must be listed in the truth table. For two variables with two input symbols, the combinations are $2^2 = 4$, with three variables there are $2^3 = 8$ combinations and so on.

8.4.2 The OR GATE

The OR gate has two or more input signals but only one output signal. Figure 8.5(a) shows a representation of OR gate with switches A and B in parallel. The bulb will light up when either or both of the switches are ON. The bulb will be OFF only when both the switches A and B are OFF.

Figure 8.5(b) is the symbol for 2 input OR gate. This also shows that OR gate performs a logic addition.

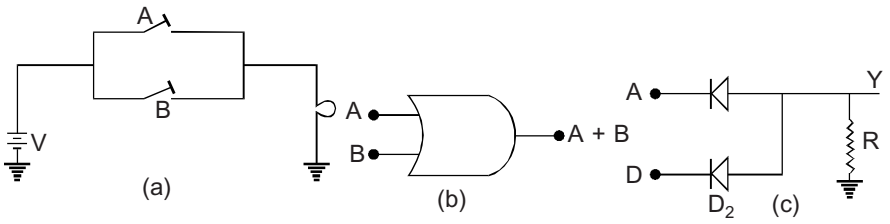


Fig. 8.5 (a) Representation (b) Symbol (c) Diode Circuit

Table 8.3 Truth Table for OR Gate

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

The + sign does not mean addition as in arithmetic. In symbolic logic the + sign means an OR function. This means all inputs must be off to turn OFF the output.

Figure 8.5(c) shows how an OR gate can be built with-two diodes. If both inputs are low, the output is low. If either input is high the diode with the high input conducts and the output is high.

Table 8.3 is the Truth Table for the OR gate. Binary 0 stands for low voltage and binary 1 for high voltage. Notice that one or more inputs produce a high output, this is why the circuit is called an OR gate.

An OR gate can have as many inputs as desired with one diode added for each additional input.

8.4.3 Inverters

An inverter is a gate with only one input signal and one output signal, the output state is always the opposite of the input state.

An NPN transistor in CE configuration, is an inverter as shown in Fig. 8.6 (a). Figure 8.6(b) is the logic symbol for an inverter.

The inverter function is also called negation. The logic symbol for inversion or negation is a bar over the function to indicate opposite state. For instance $\overline{A}$

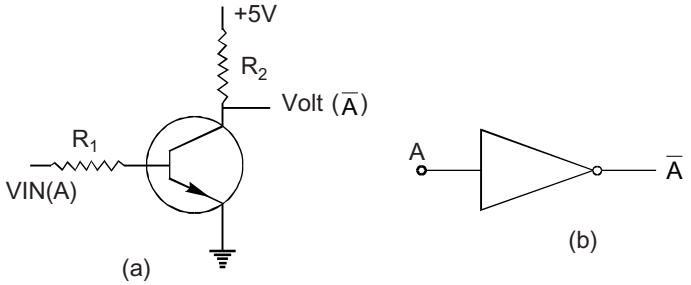


Fig. 8.6 Transistor Inverter (b) Inverter Symbol

means not A. In binary numbers inversion of 0 is 1 and 1 can be inverted to 0. The inversion also applies to ON and OFF states. In Fig. 8.6(b) input A to the amplifier with 0 circle is $\overline{A}$ in the output.

The two negative functions NAND and NOR gates will now be described.

8.4.4 NAND Gate

Figure 8.7 (a) shows the symbol for a NAND gate which is nothing but an AND gate with inverted output. The circle in the output of the AND symbol shows inversion. In logical terms a NAND gate is an AND gate followed by an inverter as shown in Fig. 8.7(b).

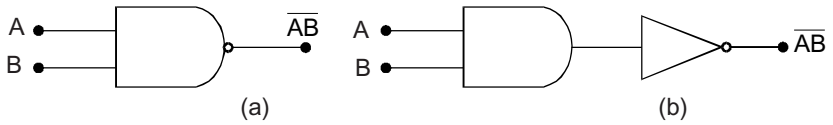


Fig. 8.7 (a) Symbol Nand Gate (b) Logical Meaning (c) Truth Table

Table 8.4 Truth Table of NAND Gate

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

(c)

The NAND gate has two or more input signals but only one output signal. All input signals must be high to get a low output—indicated by the truth table of a NAND gate in Table 8.4. In terms of the circuit the NAND gate must have both inputs ON to have the output in the OFF state.

8.4.5 NOR Gate

The standard symbol for a NOR gate is shown in Fig. 8.8(a). The logical structure in Fig. 8.8(b) shows that it is only an OR gate followed by an inverter.

Table 8.5 Truth Table-NOR Gate

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

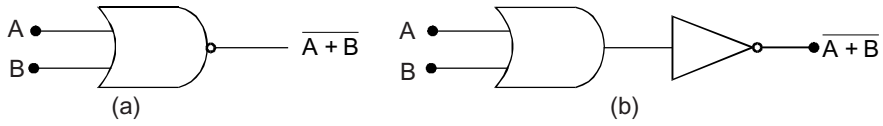


Fig. 8.8 Nor Gate (a) Standard Symbol (b) Structure

The NOR gate has two or more inputs but only one output. All inputs must be low to get a high output as indicated by the truth table of a 2-input, NOR gate in Table 8.5.

8.5 BOOLEAN ALGEBRA

As stated earlier, the Boolean algebra was invented by George Boolean to solve certain logic problems and the practical use of this mathematics was made by Shanon in analysing the operation of telephone relays which have two states either ON or OFF. Today the Boolean algebra plays an important role in computer circuit analysis and designs. Digital circuits mainly consist of the function AND, OR.INVERTER and their combinations. How Boolean algebra helps to analyse and realise various digital circuits will now be described.

In Boolean algebra letters like A, B, C etc. are used to describe inputs and outputs for logic functions and the equations formed with these letters are known as word equations. For example the inverter circuit in Fig. 8.9 has an input A with output Y.

The word equation for this circuit is,

$$Y = \text{Not } A = \overline{A}$$

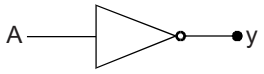


Fig. 8.9 Inverter

because $\overline{A}$ is for not A.

Since the variable A can be 0 or 1 (low or high), If $A = 0$, then $Y = \text{NOT } 0 = 1$. Similarly if $A = 1$, $Y = \text{NOT } 1 = 0$.

The word equation $Y = \overline{A}$ is read as Y equals NOT A. $\overline{A}$ is also called the complement of A.

$$Y = \overline{A} \tag{8.1}$$

is the standard equation for an inverter. Similarly we can write word equation for OR gate and AND gate as explained below.

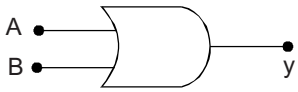


Fig. 8.10 OR Gate

Word equation for the two input OR gate is shown in Fig 8.10.

$$Y = A \text{ OR } B \tag{8.2}$$

Given the values of A and B we can solve the equation. If $A = 0$ and $B = 0$ then $Y = 0 \text{ OR } 0 = 0$.

This means that the output from an OR gate is 0 when both inputs are 0.

Now take the case where both inputs to the OR gate are high or 1. i.e.

$$A = 1 \text{ and } B = 1, \text{ then}$$

$$Y = 1 \text{ OR } 1 = 1$$

In Boolean algebra the + sign stands for OR operator and so equation can be written as

$$Y = A + B$$

This is the standard way of writing the output of an OR gate and the equation will be read as "Y equals A OR B".

Thus we find that the Boolean algebra helps us to find the output of a logic OR gate when the inputs to this gate are known.

EXAMPLE: Find the output of the OR gate if
(i) $A = 0$ and $B = 0$ (ii) $A = 1$ and $B = 1$

(i) $$Y = A + B = 0 + 0 = 0$$

Which will be read as O OR 0 O gives 0.

(ii) $$Y = A + B = 1 + 1 = 1$$

is 1 OR 1 gives out 1 OR one gives out one

In boolean algebra + sign does not mean ordinary addition as in decimal number but + sign stands for OR addition in logic circuit.

AND gate

The word equation for the AND gate shown in Fig. 8.11 is

$$Y = A \text{ AND } B \qquad (8.3)$$

In boolean algebra this equation can be written as

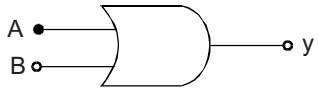


Fig. 8.11 AND Gate

$$Y = A \cdot B = AB$$

because the AND operation stands for multiplication. This is the standard way of writing the AND equation and this will read "Y equals A AND B.

Given the values of A and B, the output of AND gate can be formed with the help of boolean algebra as below:

Case 1 Both inputs low i.e. $A = 0$, $B = 0$

$\therefore$ $$Y = AB = 0.0 = 0$$

or 0 ANDED *w*ith 0 gives 0 or low output.

Case 2 If A is low (0) and B is high (1)

$$Y = AB = 0.1 = 0, \text{ the output is low}$$

Case 3 If A = high (1) and B is low (0) then $Y = AB = 1.0 = 0$, output is low.

Case 4 If both inputs are high $A = 1$, $B = 1$ then $Y = A \cdot B - 1.1 = 1$ which means 1 ANDED with 1 gives 1 which is a high output.

These results are the same as indicated by the truth table for AND gate shown under 8.4.1.

The three logic gates viz. the Inverter, OR gate and the AND gate are called decision making elements because they can recognize some input words while disregarding others. A logic gate recognizes a word when its output is high, it disregards a word when its output is low.

Positive And Negative Logic In *positive logic* 1 stands for the more positive of the two voltages whereas in negative logic 1 stands for more negative of the two voltages. For example, if the two voltage levels are 0 and – 5V, the positive logic will have binary 1 stand for 0V and 0 for – 5V. In negative logic however, binary 1 will represent –5V and binary 0 will stand for 0V.

Normally positive logic is used with positive supply voltages and negative logic is used with negative supply voltages.

8.6 De MORGAN'S THEOREMS

Laws of Boolean algebra pertaining to various logic gates have been generalized by Augustus De Morgan in the form of two identities known as DC Morgan's theorems.

1. De Morgan's First Theorem The theorem states that

$$\overline{A + B} = \overline{A}\ \overline{B} \tag{8.4}$$

In words the theorem says that the complement of a sum equals the product of the complements. A graphic representation of the theorem is given in Fig. 8.12.

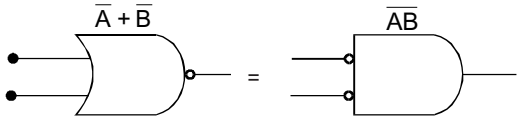


Fig. 8.12 De Morgan's First Theorem

The left side in Fig 8.12 is an OR gate and the right hand side is an AND gate with inverted inputs also known as the bubbled AND gate in which the inverter triangles have been replaced by two bubbles at the input of the AND gate. The bubbles are a reminder of the inversion that takes place before ANDing.

Table 8.6 is a Truth Table for the De Morgans first theorem.

Table 8.6

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

The last two columns in the table are identical for all possible combinations of A and B which proves the theorem.

If both inputs are low (0), the AND gate has high inputs and therefore, the final output is high (1). If one or more inputs are high, one or more AND gate inputs must be low and the final output is low.

The above theorem is equally applicable to more than two inputs.

De-Morgans Second Theorem De-Morgans second theorem for two inputs states that

$$\overline{AB} = \overline{A} + \overline{B} \tag{8.5}$$

In word this means that the complement of a product equals the sum of the complements. Left hand side of identity represents a NAND (Fig. 8.13 (a)) and the right hand side stands for an OR gate with inverted inputs.

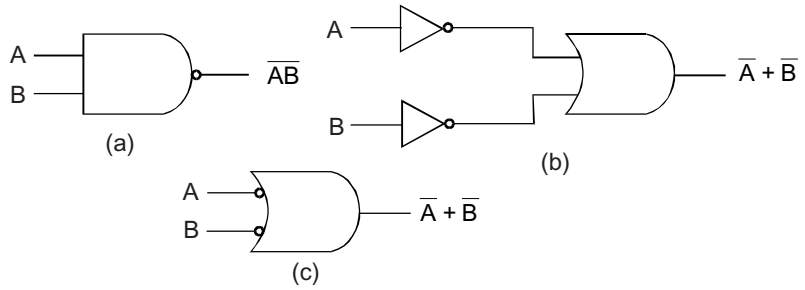


Fig. 8.13

Figure 8.13 (b) can be reduced to Fig. 8.13 (c) which is a bubbled OR gate, the bubbles indicate the inversion that takes place before ORing. De-Morgan's second theorem boils down to the fact that (a) and (b) in Fig. 8.14 are equivalent.

This indicates that a NAND gate and a bubbled OR gate are equivalent.

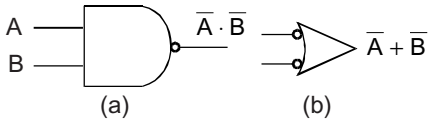


Fig. 8.14 De-Morgans Second Theorem

To Prove De-Morgans second theorem we can take all possible combinations of the values for A and B, find the values of $\overline{AB}$ and $\overline{A} + \overline{B}$ and show that these are equal.

(i) When $A = 0$ and $B = 0$

$$\overline{A \cdot B} = \overline{0 \cdot 0} = \overline{0} = 1$$

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

(ii) When $A = 0$ and $B = 1$

$$\overline{A \cdot B} = \overline{0 \cdot 1} = \overline{0} = 1$$

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

(iii) When $A = 1$ and $B = 0$

$$\overline{A \cdot B} = \overline{1 \cdot 0} = \overline{0} = 1$$

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

(iv) When $A = 1$ and $B = 1$

$$\overline{A \cdot B} = \overline{1 \cdot 1} = \overline{1} = 0$$

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

Since there are no other input combinations the second De-Morgans theorem stands proved.

The above results are also summarised in the truth table in Table 8.7.

Table 8.7 Truth Table for De-Morgan Second Theorem

AB	$\overline{A \cdot B}$	$\overline{A} + \overline{B}$
00	1	1
01	1	1
10	1	1
11	0	0

The last two columns of the table are the same.

De-Morgans theorems are useful in changing Boolean expressions to equivalent forms. For applications of these theorems change plus signs to multiplication signs and vice versa and take the complements of individual terms rather than the entire expression.

For $\overline{A + B}$, change the + sign to a – sign to get A. And to get $\overline{A} \cdot \overline{B}$ take the complement of each term.

EXAMPLE: Apply De-Morgans second theorem to the following identity.

$$\overline{\overline{A} \cdot B} = A + \overline{B}$$

Solution:

De-Morgans second theorem states that complement of a product equals the sum of the complements. $\overline{\overline{A} \cdot B}$ is the complement of $\overline{\overline{A} \cdot B}$ which is the product of $\overline{A}$ and B. So the complement of $\overline{\overline{A} \cdot B}$ is the sum of the complements of $\overline{A}$ and B.

$$\overline{\overline{A} \cdot B} = \overline{A} + \overline{B}$$

$\overline{\overline{A}}$ is called the double complement A and the double complement of a variable is equal to the variable itself.

$$\overline{\overline{A}} = A \quad \text{if} \quad A = 0 \quad \text{then}$$
$$\overline{\overline{A}} = \overline{\overline{0}} = \overline{1} = 0$$

If $A = 1$

$$\overline{\overline{1}} = \overline{0} = 1$$

So $$\overline{\overline{A} \cdot B} = \overline{\overline{A}} + \overline{B} = \overline{A} + \overline{B}$$

and this proves the identity.

8.7 EXCLUSIVE—OR GATE

Figure 8.15 (a) shows one method of building an exclusive OR gate, abbreviated as XOR gate.

The upper AND gate forms the product $\overline{A} \cdot B$ and the lower AND gate $A \cdot \overline{B}$, therefore the boolean equation for the output is given by.

$$Y = \overline{A}B + A\overline{B}$$

Figure 8.15 (b) is the standard symbol for an exclusive OR gate.

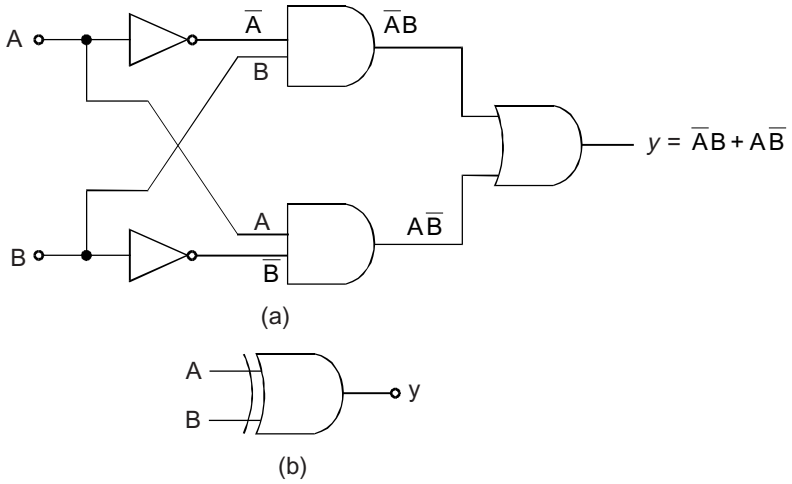


Fig. 8.15 (a) Exclusive OR Gate (b) Symbol for 2-input Exclusive OR Gate

The truth table for the exclusive OR gate is given in Table 8.8.

Table 8.8 Truth Table for Input XOR gate

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

The output is high when A or B is high but not both. This is why the circuit is known as an Exclusive OR gate because output is 1 only when the inputs are different.

Logical Symbol and Boolean sign

Figure 8.15(b) is the standard symbol for a 2-input XOR gate with equation $y = A$ XOR B which can also be written as $y = A + B$ where + sign stands for XOR addition. This is read as "y equals A XOR B".

8.8 EXCLUSIVE NOR GATE

The exclusive NOR gate abbreviated XNOR is logically an XOR gate followed by an inverter. Figure 8.16 (a) is the circuit and (b) is the symbol for a XNOR gate, because of the inversion on the output side the truth table of an XNOR gate is the complement of an XOR gate truth table as shown in Table 8.9.

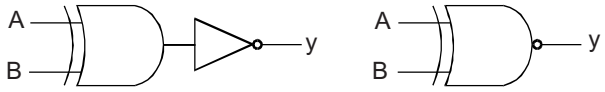


Fig. 8.16 XNOR Gate (a) Circuit (b) Symbol

Table 8.9 Truth Table for XNOR Gate

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

8.9 THE UNIVERSAL BUILDING BLOCK

Boolean expressions can be used to build logic circuits and logic circuits can be reduced to Boolean expressions. To build a logic circuit from any Boolean expressions we can use OR gates for the + sign, AND gates for – sign and NOT circuits for the overhead bars. This means that OR, AND and NOT circuits are the basic building blocks for all logic circuits. The real building block is, however, the NAND gate because it can be used to build all the three gates mentioned above i.e. OR, AND and NOT gate as shown below.

(i) Building NOT circuit from NAND gate

This can be done by connecting all inputs together as shown in Fig. 8.17 (a). The simplified symbol for the NOT gate is shown in Fig. 8.17 (b).

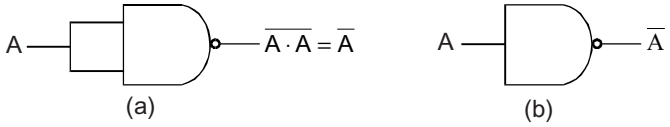


Fig. 8.17 NAND Gate as an Inverter

EXAMPLE:

If $A = 0$

$$\overline{A} \cdot \overline{A} = \overline{0} \cdot \overline{0} = 1 = \overline{A}$$

If $A = 1$

$$\overline{A} \cdot \overline{A} = \overline{1} \cdot \overline{1} = 1 = 0$$

(ii) Converting a NAND gate to AND gate. This is shown in Fig. 8.18 (a).

The output of the first NAND gate is $\overline{AB}$ which is complemented by the second NAND gate to produce $\overline{\overline{AB}}$. This double complement of AB is equal to AB as mentioned earlier in the chapter.

(iii) Converting a NAND gate to OR gate.

This is shown in Fig. 8.18 (b). The first two NAND gates invert A and B to produce $\overline{A}$ and $\overline{B}$. The two input NAND gate produces an output of $\overline{A} \cdot \overline{B}$, then

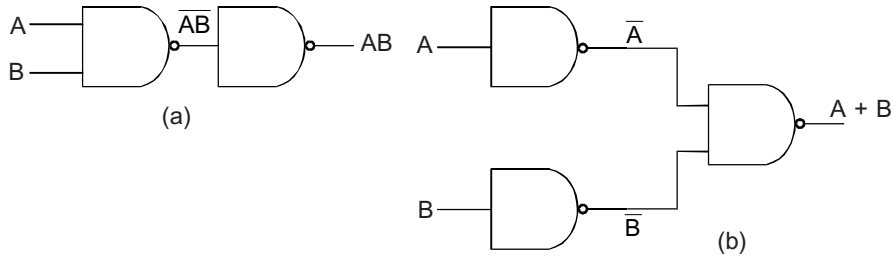


Fig. 8.18

$$\overline{\overline{A} \cdot \overline{B}} = \overline{\overline{A}} + \overline{\overline{B}} = A + B$$

as per De-Morgans second theorem.

The above examples prove that any logic circuit can be built using only NAND gates.

8.10 LOGIC CIRCUITS

The logic gates described earlier can be realized and built with the help of components like resistors, capacitors, diodes and transistors, both junction type and FET. A diode is switched on by applying forward bias. The conducting diode is a short circuit with no resistance and internal voltage drop of less than 1 V. Applying reverse bias to a conducting diode cuts it off. It is an open circuit with very high resistance.

Similarly a transistor can be switched on by applying enough forward bias to produce saturation. The transistor output voltage is very low with the collector saturation voltage less than 1V.

When the transistor is cut off without any output current, output voltage is high and this voltage is equal to supply voltage as there is no IR drop in the load R_L. A transistor in the CE configuration is also an inverter. It negates a logic function from the input at the base to the output at the collector.

All these components are generally combined on a chip to form an Integrated Circuit (IC). Most logic functions are available in the form of IC chips.

8.10.1 Types of Logic Circuits

Depending on the type of components used for their manufacturer, the various logic circuits can be divided into two main categories or families known as Bipolar and MOS families. The first category fabricates bipolar transistors, in a chip and the second MOSFETs. Bipolar technology is preferred for SSl and MSI because it is faster. MOS technology is prevalent in the LSI field because more MOSFETs can be accommodated in the chip of the same area.

8.10.2 Compatibility

Two logic devices are said to be compatible when the output of one device can be connected to the input of another device belonging to the same logic family. Compatibility permits a large number of different combinations.

A list of logical circuits that fall under the two main categories i.e. Bipolar and MOS families is given below.

8.11 BIPOLAR FAMILIES

RTL — Resistor-Transistor logic
DTL — Diode-Transistor logic
TTL — Transistor Transistor logic
ECL — Emitter Coupled logic.

RTL using resistors and transistors family was the first to be introduced. DTL uses diodes and transistors, once popular is now obsolute. TTL uses transistors almost exclusively and is perhaps, the most popular family of SSI and MSI chips. ECL is the fastest logical family and used in high speed technology.

8.12 MOS FAMILIES

This comprises the following families:-

P MOS — p-channel MOSFETs
N MOS — n-channel MOSFET
C MOS — Complementary MOSFET.

P MOS is becoming obsolete whereas N MOS is largely used in LSI field, which includes microprocessors and memories. C MOS a push pull arrangement of n and p-channel MOSFETS is extensively used where low power consumption is required as in pocket-calculators, digital wrist watches etc.

8.13 DEVICE NUMBERS

7400 Series In the line of TTL circuits is the most widely used of all bipolar ICs. This TTL family contains a variety of SSI and MSI chips that allows you to build all kinds of digital circuits and systems.

By slightly varying the design of the circuit the manufacturer can alter the number of inputs and the logic function. For example 7400 with four 2 input NAND gates is one package and 7402 has four 2-input NOR gates, the 7404 has six inverters and so on.

5400 Series Any device in the 7400 series works over a temperature range of 0° to 70°C and over a range of 4.75 to 5.25 V. This is adequate for commercial applications. However, the 5400 series, primarily meant for military applications, has the same logic function as the 7400 series except that it works over a temp. range of – 55 to 125°C and over a supply range of 4.5 to 5.5 V. The 5400 series are rarely used commercially because of the much higher cost.

8.14 SPECIFICATIONS OF LOGIC FAMILIES

The specifications provided for the various logic families by the manufacturers contain many terms which are unique to digital electronics. Some of the important specifications must be defined before describing the characteristics of individual logic families.

FAN-IN and FAN-OUT The fan-in of a logic circuit is the number of inputs that the logic circuit can handle. For example, an eight input gate requires one unit load per input and hence its fan-in is 8.

The fan-out of a circuit is the number of unit inputs that can be driven by a logic element. If a gate has a fan-out of 6, this means it can drive six unit inputs and still maintain its logical 1 and logical 0 output voltage specifications.

Noise Immunity This is defined as the max. induced noise voltage a TTL device can withstand without a false change in the output state.

8.14.1 TTL Device

Figure 8.19 (a) shows two TTL devices connected together with the dividing device having an output of 0.4 V in the low state or worst case. If no more voltage is induced into the connecting wire, the input to the second device is 0.4V. If, however, an induced voltage of 0.4V is developed in the wire with a polarity that it adds to the input of the second device making this input the higher than low state input voltage as per specifications of the device, the second TTL device could undergo a false change output voltage under the worst case conditions.

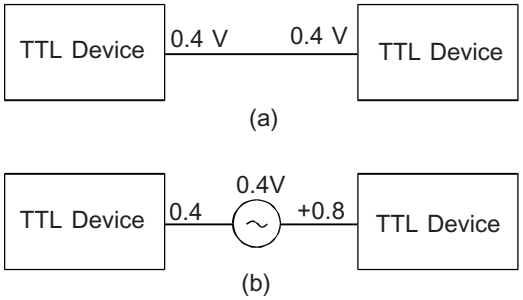


Fig. 8.19

Similarly false triggering can also take place in the high state as shown in Fig. 8.20.

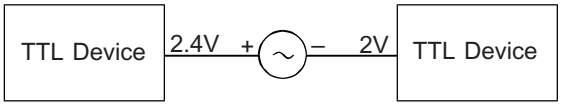


Fig. 8.20 False Triggering to Low State

If the worst case TTL output of first device is 2.4 and if 0.4 V of more voltage is induced on the line which is of opposite polarity to the output voltage of first device, then the input voltage to the second device is 2V which is less than the min. voltage required for high state input. The second TTL device will be on the verge of a false change in output state. The first voltage is more than 0.4 V, the second TTL could falsely trigger to the opposite state under worst conditions.

8.15 PROPAGATION DELAY

Propagation delay time is the time that elapses between the change in input state and the resulting change in output state. The propagation delay line for the TTL gate is of the order of 10 ns (10×10^{-9}s). If a number of TTL devices are cascaded, the total propagation delay time is equal to the sum of the individual propagation times. Thus if three TTL devices, each with a propagation delay time of 10 ns are cascaded, the total propagation delay time will equal 30 ns.

8.16 LOGIC CIRCUITS

8.16.1 Resistor-Transistor Logic (RTL)

RTL family was the first to be introduced. An RTL NOR gate is shown in Fig. 8.21(a). The symbol of RTL is shown in Fig. 8.21(b)

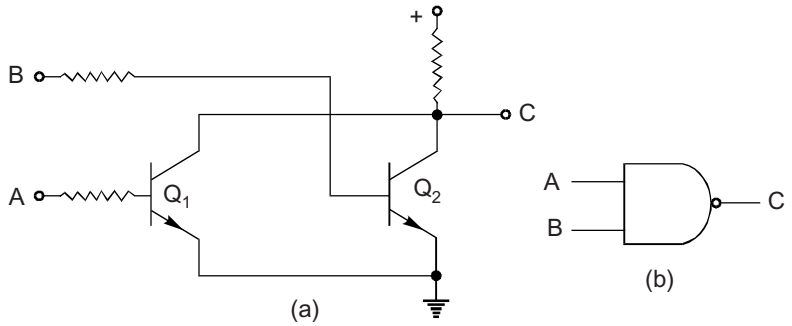


Fig. 8.21

If input A is high, Q_1, will conduct, pulling pin C, the output to ground. Similarly, if B is high C will be pulled to ground through Q_2. The circuit has a propagation delay of 12 ns and a fan out of 4.

The RTL circuit can be used for NAND and NOR gates.

8.16.2 Diode-Transistor Logic (DTL)

These circuits use diode gates with a CE transistor amplifier as an inverter stage.

The DTL circuit in Fig. 8.22 (a) is a NAND gate.

The diodes D_1 and D_2 with R_A form an AND gate. Both inputs V_1 and V_2 must be high with positive voltage at the cathode to cut off the diodes to allow point C to rise to V_A. When only one diode is cut off, the other conducting diode keeps the gate output low, close to ground potential.

R_1 isolates the loading effect of this slow series from base to emitter. C_1 is the speed up capacitor. Fast variations in gate voltage can be compiled through C_1.

The output voltage of the diode AND gate is the input V_B to the base of the NPN transistor Q_1. Without any input voltage from the gate, Q_1 is cut off by

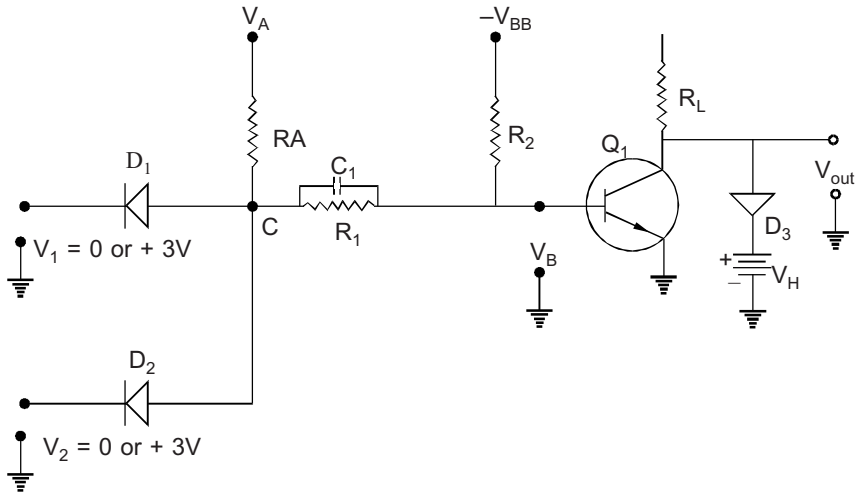


Fig. 8.22 NAND Gate in DTL CIRCUIT

negative base voltage from—V_{BB} through R_2. Then the collector voltage is high at $+V_{CC}$ without any collector current. V_{out} therefore is high at the state 1. However, positive drive at the base makes Q_1 conduct and as a result V_{out} drops to the O state.

D_3 is a diode clamp which holds the high state of V_C at the clamping level V_A. Any V_C greater than V_A makes D_3 conduct and the collector is effectively connected to V_H.

Figure 8.23 is a NOR gate in DTL circuit.

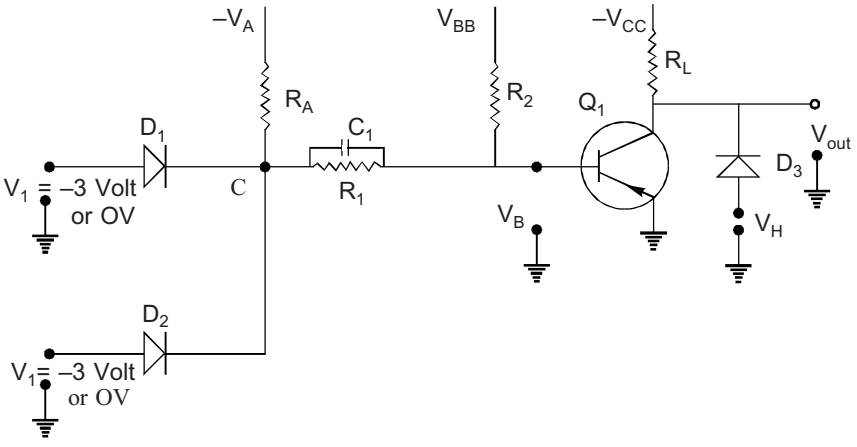


Fig. 8.23 NOR Gate in DTL Circuit

In this the diodes of the OR gate are reversed from the AND gate of Fig. 8.22, the transistor is a PNP transistor instead of NPN.

Note that with the positive logic ground potentials is the 1 state compared with $-3V$ for the 0 state.

Without any input voltage to the diodes point C is the ground potential through the conducting diodes. When both diodes pass $-3V$ to point C it is at 0 state. When only one diode has $-3V$ input the short circuit through the other conducting diode without input can ground point C for 1 state.

V_{out} is low when Q_1 conducts and high when it is cut off.

V_{out} for the NOR gate is inverted from an inclusive OR gate. The output is at 1 when both inputs are at 0.

8.16.3 WIRE OR Circuits

Both RTL and DTL have a capability called wire OR. Since the output is effectively a transistor, two outputs can be wired together, producing a low true OR function. This saves on the total package needed for the design Fig. 8.24(a).

8.16.4 Standard TTL NAND Gate

A standard TTL NAND gate is shown in Fig. 8.24(b). The multi-input transistor is a typical of all 7400 series gates and circuits.

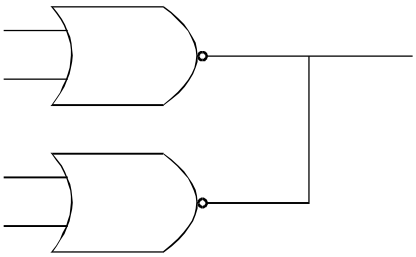


Fig. 8.24(a) Wired OR (RTL Logic)

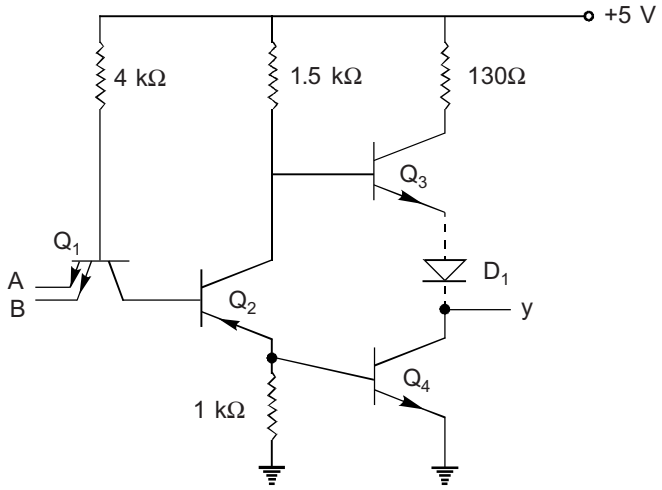


Fig. 8.24(b) Standard TTL NAND Gate

Each emitter acts like a diode and therefore Q_1 and 4 kV resistor act like a 2-input AND gate. The rest of the circuit inverts the signal and the entire circuit acts like a 2-input NAND gate.

TOTEM Pole Connection The output transistors Q_3 and Q_4 form a totem-pole connection, typical of most TTL devices. Either Q_3 or Q_4 is on. When Q_3 is on, output is high, when Q_4 is on the output is low. When Q_3 is on, it acts like an emitter follower (high output) and when Q_4 is saturated, the output is low. In either case the output impedance is very low. This is important because it reduces the switching speed. In other words, when the output changes from low to high or vice versa, the low output impedance means a short RC time constant. This short time constant means that the output voltage can change quickly from one state to the other.

8.17 MULTIVIBRATORS

A multivibrator is also a relaxation oscillator like the blocking oscillator but without the use of a transformer for providing the feedback. A multivibrator consists of two RC coupled amplifier stages in which the output of one stage is applied as input to the other stage, as shown in Fig. 8.25. Since each amplifier stage produces a 180° phase-shift, the signal applied to the input of each stage is in phase with the original input and this helps produce oscillations.

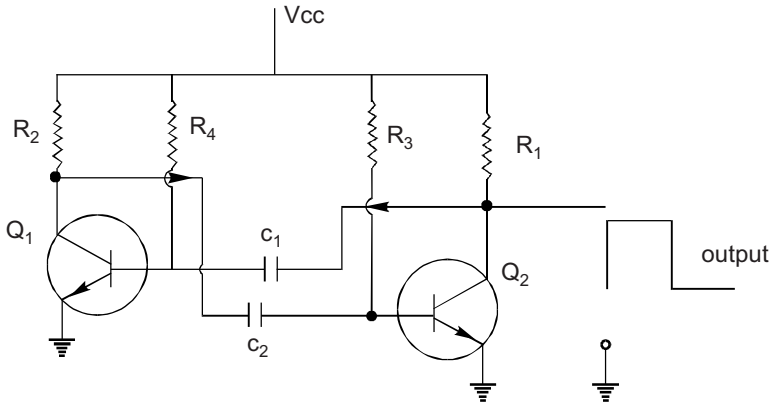


Fig. 8.25 Basic Multivibrator

Multivibrators are useful devices for generating square waves and pulses where the frequency can be controlled by voltages injected from outside. The waves generated by multivibrators are also very rich in harmonics.

Multivibrators are classified either according to the system of feedback employed or according to the stability conditions obtaining in its operation. The classification according to the feedback method includes.

8.17.1 Collector Coupled Multivibrator

As shown in Fig. 8.25, the collector of Q_1 drives the base of Q_2 and the collector Q_2 drives the base of Q_1 through RC coupling.

8.17.2 Emitter Coupled Multivibrator

In this type the collector of Q_1 drives the base of Q_2 but is coupled back to Q_1 only through an emitter resistor which is common to both stages.

The classification of multivibrators according to stability conditions includes the following:

1. Free-running or Astable Multivibrator

When neither of the two stages is stable but remains alternately cut off and conducting at the MV repetition rate, the MV is a free running or astable type. The frequency of an astable multivibrator can be controlled by external synchronising pulses. The multivibrator tends to adjust the frequency of its oscillations so that the ratio of the external synchronising frequency and the multivibrator frequency is an integral ratio which can be either unity or less or greater than unity.

2. Monostable or "One-shot" Multivibrator

This type of multivibrator has only one stable state. An external pulse is required to initiate action when the multivibrator goes through one cycle of operation and returns to its original stable state, when the external pulse is no longer present. In this type of multivibrator, one of the MV transistors is biased to a voltage more negative than cut-off bias for the collector supply employed.

3. Bistable Multivibrator or Flip-flop Circuit

This type of circuit which has two stable conditions is also known as the Eccles-Jordan trigger. This is a direct coupled two-stage amplifier in which the output of the second stage is connected to the input of the first stage. When an external pulse is applied to the base of the non-conducting transistor, the system jumps almost instantly from one stable state to the other resulting in a trigger or flip-flop action.

Both the monostable and bistable multivibrators are driven oscillators or trigger circuits which need an external driving pulse to start the operation and maintain it.

8.18 FLIP-FLOP

A flip-flop circuit is only a bistable type of multivibrator in which each stage can change abruptly between cut off and conduction. The circuit can stay in either state indefinitely till a trigger pulse arrives to change that state.

When the flip-flop is in either state it remains that way till a trigger pulse changes that state. This condition is one stable state. It flips to another state on arrival of the input pulse and this is also a stable state. The current behaves like a toggle switch, which remains in one state till it is toggled to the opposite state.

Figure 8.26 shows a general form of flip-flop with input and output terminals. The input terminals are labelled set (S) and Re-Set (RS) for set and reset pulses. The output terminals are labelled Q and its negative $\overline{Q}$.

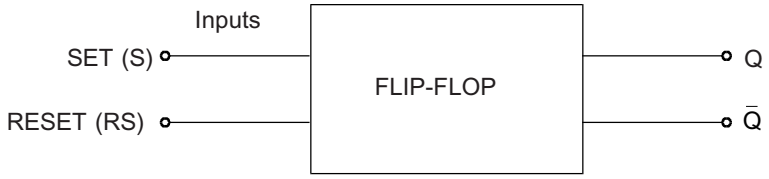


Fig. 8.26 Flip-flop with Reset-Set (RS) Inputs

8.18.1 RS-Flip-flop from Logic Gates

Figure 8.27 shows the formation of the RS-Set flip flop formed by the combination of two Nand gates. Two NOR gates can also be combined to form a flip-flop.

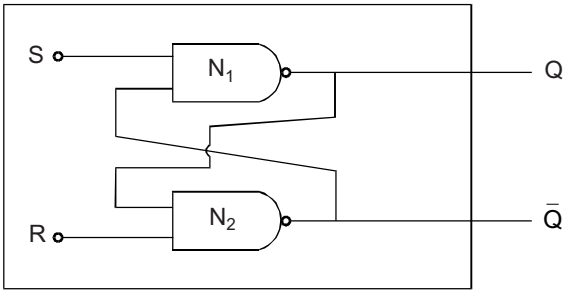


Fig. 8.27 Formation of Reset-Set Flop from Nand Gates

8.18.2 Clocked RS Flip-flop

The flip-flop cannot change position unless a clock-pulse is provided. It is, therefore, necessary to provide an additional terminal for clock pulses as shown in Fig. 8.28.

Fig. 8.28

The clocked type is a synchronous FF as the operation is timed by the clock pulses.

8.18.3 J K Type Flip-flop

Various types of clocked flip-flops are available but J K type flip-flop is the most common type because it has no ambiguous state.

Figure 8.29 shows 7476 type of J K flip flop.

In a clocked-flip flop either the S or R input can toggle the output, as long as both are not same. The condition of both S and R low is not permitted. The condition of both S and R high allows J and K clock inputs to toggle the outputs.

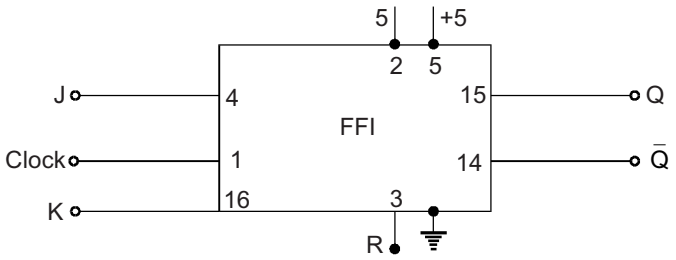


Fig. 8.29 J K Type Flip Flop Type 7476

For the clock-pulses every two changes result in one cycle of change in the FF output. Therefore the circuit is a divide by two or binary flip flop.

8.19 REGISTER

A register is a memory device for storing digital data. There are two types of registers.

Buffer Register A buffer register is the simplest kind of register. All it does is to store a digital word.

Shift Register A shift register is a memory in which information is shifted on position at a time when one clock pulse is applied. The data can be shifted in either direction i.e. left or right. When the data is shifted to left it is called a left shift register and when shifted to right it is a right shift register. This data shifting or both shifting is essential for certain arithmetic and logic operations in computer technology. A counter is a form of register.

8.20 COUNTERS

A counter or binary counter is a chain of flip-flops in which each divides by 2. The successive flip-flops divide in multiples of two. The division is in the frequency of the clock-pulse input. Since each flip-flop divides the frequency by factor of 2, the flip-flop is called a divide-by-two circuit. Since each flip-flop divides the clock frequency by 2, n flip-flops will divide the clock frequency by 2^n. Figure 8.30 shows a counter with three flip flops and it is a divide by three counter.

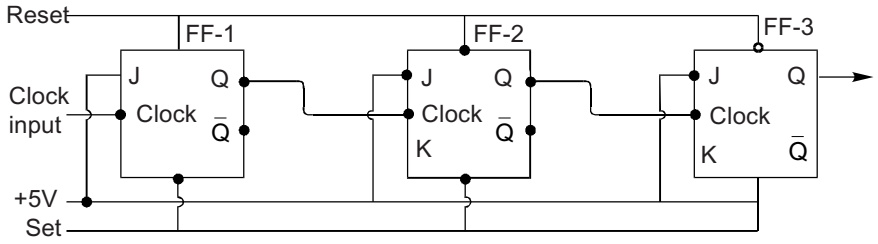


Fig. 8.30 Ripple Counter (Divide by Eight)

There are two types of counters, a ripple counter and a synchronous counter.

8.20.1 Ripple Counter

In a ripple counter the carry moves through flip-flop like a ripple on water. Figure 8.30 is a ripple counter.

However where the carry moves through a chain of *n*-flip-flops, the overall delay time is *n*, t_p. As such, the ripple counters are too slow for certain applications. To get over the problem of ripple-delay, use is made of a synchronous counter.

8.20.2 Synchronous Counter

In this type of counter the clock pulses drive all the flip-flops in parallel. Because of this simultaneous clocking, the correct binary word appears after only one propagation delay time rather than three as in Fig. 8.30 above.

Summary

The electronic communication signals can be divided into two types—the analog signals and the digital signals. In analog signals the values of the signal varies continuously over its full range from its minimum value to its maximum. value over its full range of values. A digital signal, however, is represented by discrete numbers, and has only two limits 0 and l. Digital signals are less prone to noise and interference than analog signals. A common use of digital circuits is in numerical counting. These applications include the field of computers, electronic calculators, digital clocks and various types of measuring equipment. Digital electronics makes use of binary numbers, binary arithmetic, logic gates and switching circuits, other topics discussed in this chapter include Truth tables, Boolean Algebra, Basic logic circuits, Diode gate circuits, Diode Transistor logic (DTL), Transfer-Transistor logic (TTL) Multivibration, flip-flop circuits and counters.

Review Questions

1. Explain the difference between an analog and a digitial signal. Give examples.
2. What is a binary system? Convert $(13)_{10}$ into its binary form.
3. What is a logic gate? Define fan-in and fan out factors for a logic gate.
4. Compare the following pairs of logic gates (a) AND with OR (b) AND with NAND (c) OR with NOR.
5. What is a Truth Table? Use a Truth Table to show that $(\overline{A})\ (\overline{B}) = \overline{A + B}$.
6. Explain what is meant by negation or inversion. Describe the two negative functions NAND gate and NOR gate.
7. State De-Morgans theorems. Give the truth tables for these theorems.
8. Explain what is meant by a Universal Building Block. Give examples.
9. What is meant by compatibility in logic circuits. Give the names of two main categories of logic circuits.

10. What is Diode-Transistor Logic (DTL). Explain with a diagram the working of DTL.
11. What is a multivibrator? Describe briefly two types of multivibrators.
12. What is a counter? Describe two types of counters and give their practical applications.

chapter 9

Amplifiers

9.1 AMPLIFIER—DEFINITION

Vacuum tubes and transistors are *active* components which can be combined with *passive* components such as resistors, capacitors and inductors to form what are known as *electronic circuits*. An *amplifier* is an electronic circuit that produces an enlarged version of a small signal fed into the circuit. The terminals where the small signal is applied form the *input* and the terminals where the enlarged signal appears is the output. Figure 9.1 shows the block schematic of an amplifier with the input and output terminals marked. One terminal is generally common to both input and output and is grounded to the chassis. Any of the three electrodes in a transistor could form the common or grounded terminal, giving rise to the common-base (CB), common emitter (CE) or common collector (CC) configurations. The characteristics of the amplifier will depend on the configuration used in a particular circuit.

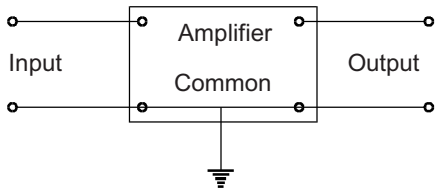


Fig. 9.1 Block Schematic of an Amplifier

Amplifiers of different types form the building blocks for a complete circuit of a radio or television receiver.

9.2 CLASSIFICATION

9.2.1 Classification According to Power-voltage and Power Amplifiers

In a voltage amplifier, the output voltage is many times greater than the input voltage. The components in a voltage amplifier are selected to give high *gain*

or amplification which is the ratio of the output voltage to input voltage. Voltage amplifiers form the earlier stages of complete amplifier circuits in radio and TV receivers where the voltages received are extremely small and need amplification before being really useful or effective. Voltage amplifiers are also sometimes called pre-amplifiers. Power amplifiers are similar to voltage amplifiers but the power developed in the output circuit or the load is very high compared to input power. A power amplifier may have some voltage gain also but the main consideration is the power output which is the product of the current and voltage. The vacuum tubes or transistors used in power amplifiers are capable of drawing large currents and generally get hot during operation and need cooling. Power amplifiers are used in the output stages of Radio, TV and other electronic equipments. Power amplifiers may develop only few milliwatts of power as in small transistor receivers or hundreds of kilowatts of power as in the case of high power broadcast transmitters.

9.2.2 Audio Frequency Amplifiers

The word 'audio' pertains to the sense of hearing in human beings. Audio frequencies are pressure variations in air which are within the hearing capability of a human ear. The range of audio frequencies extends from 20 to 20,000 Hz. This range, however, varies with age and the individual. The sound frequencies, as such, cannot be amplified by electronic amplifiers. These sound frequencies must be converted into electrical frequencies or variations before these can be amplified. This conversion from sound pressure variations to equivalent electrical variations is effected by means of a *microphone.*

Audio amplifiers are, therefore, amplifiers which are suitable for the amplification of electrical frequencies lying within the audio frequency range of 20 to 20,000 Hz. Audio frequency (AF) amplifiers are mainly used in the output stages of radio receivers and sound sections of TV receivers for driving loudspeakers to produce sound. AF amplifiers are also used in record players, PA systems, tape recorders, stereo system and other Hi-Fi equipments.

9.2.3 Classification of Amplifiers—According to Bias

The mode of operation of an amplifier is decided by the fixed bias (V_{BB}) applied for a particular value of the collector voltage (V_{CC}). Depending upon the mode of operation, the amplifiers are classified as Class A, Class B, Class AB and Class C amplifiers.

Class A Amplifiers A Class A amplifier is so biased in relation to the input voltage that the collector current flows throughout the complete cycle (360°) of the input signal voltage. The amplifier operates on the linear portion of the characteristic and output waveform is a true replica of the input waveform. The transistor neither cuts off nor saturates during the complete range of the input signal. Figure 9.2 is a schematic diagram of a Class A amplifier.

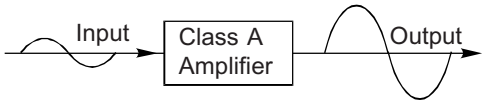


Fig. 9.2 Class A Amplifier

Because of their linear operation, Class A amplifiers have a high quality output with low distortion. However, Class A amplifiers have low power output and the efficiency is also low (25-35 per cent). The efficiency is the ratio of the power output to power input. Class A amplifiers are, therefore, used mostly as voltage amplifiers or drivers in the earlier stages of an audio amplifier. Class A amplifiers are also used as output stages in audio amplifiers where the quality of the output is the main consideration.

Class B Amplifiers In a Class B amplifier, the base bias is adjusted to cut off, so that the collector current flows only for half the cycle (180°) of the input signal voltage and remains cut off during the other half cycle of the input. When there is no input signal, the transistor does not conduct and there is no current flow in the amplifier. Figure 9.3 is a representation of a Class B amplifier.

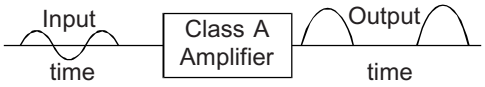


Fig. 9.3 Class B Amplifier

It will be noted that amplification takes place only during the positive half of the input signal and the output waveform is not an exact replica of the input waveform. The amplifier produces high distortion although the efficiency is comparatively high (50 to 60 per cent). These amplifiers are not suitable as audio amplifiers because of the high distortion.

In order to make a Class B amplifier suitable for use in audio amplifiers, two Class B amplifiers are so connected that their outputs, when combined, form a complete replica of the input waveform as shown in Fig. 9.4. Two Class B amplifiers operated in this fashion are said to operate as a *push-pull amplifier.*

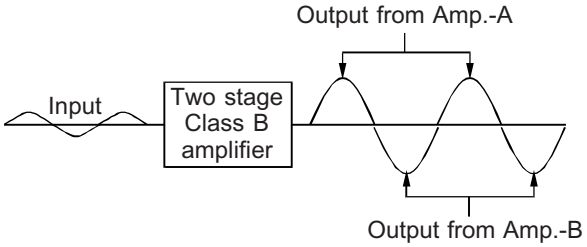


Fig. 9.4 Push-pull Amplifier

In push-pull operation one Class B amplifier amplifies the positive half and the other Class B amplifier amplifies the negative half of the input signal. The

two outputs are so combined that the output waveform exactly corresponds to the input waveform. For successful operation of a push-pull amplifier, the inputs to the two Class B amplifiers must be 180° out of phase with each other. There are several methods of arranging out of phase inputs to push-pull amplifiers. Also the two transistors used in the two branches of the amplifier must have identical characteristics.

Class B push-pull operation combines the advantages of high output and low distortion. It draws current only when the input signal is applied. This means less drain on the power supply compared to a Class A amplifier which draws a constant current even without signal. In view of the economy in power consumption, push-pull Class B amplifiers are mostly used in battery operated transistor radio receivers where power consumption is a major consideration.

Class AB Amplifiers The mode of operation of a Class AB amplifier is such that the collector current flows for more than half but less than the complete cycle of the input signal voltage. At low signal levels these amplifiers automatically operate as Class A amplifiers with low distortion and at high signal levels they provide higher output with increased efficiency. Class AB amplifiers, therefore, combine the advantages of Class A and Class B amplifiers.

Class C Amplifiers In a Class C-amplifier the transistor is biased much below cut off, so that the transistor conducts during less than half of each cycle of the input (Fig. 9.5). When there is no signal the transistor remains completely cut off. There is no resemblance between the input and output waveforms. As such, the amplifier output is highly distorted and not suitable for use as an audio amplifier. A Class C amplifier, however, provides high output and extremely high efficiency (70 to 75 per cent). It is, therefore, used in oscillators and in the output stages of high power transmitters where efficiency is a major consideration.

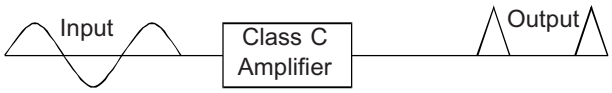


Fig. 9.5 Class C Amplifier

9.3 TRANSISTOR AMPLIFIERS

A transistor can perform almost the same functions as a triode vacuum tube. Transistors possess many advantages over vacuum tubes and most amplifiers these days are transistor amplifiers particularly the audio amplifiers.

A transistor is a current controlled device and the basic principle underlying the operation of a transistor as an amplifier is that the base current (I_B) controls the collector current (I_C). For satisfactory operation of a transistor the bias conditions have to be such that the emitter-base junction is forward-biased and the collector-base junction is reverse-biased. Of the three configuration, CB, CE and CC, in which a transistor can be connected, the CE (common-emitter)

configuration is the one that is most suitable for the audio amplifiers. Compared to a CB amplifier, the CE amplifier possesses many advantages like higher input impedance, lower output impedance and ability to provide current, voltage and power gain. In view of these advantages the CE amplifier is the most widely used amplifier in audio, TV and other applications.

The amplifying action of a CE amplifier is mainly due to the fact that a small increase in base current (ΔI_B) can produce a much higher increase in collector current (ΔI_C) and the current gain β is given by the relation:

$$\beta = \frac{\Delta I_C}{\Delta I_B} (E_{cc} \text{ constant}) \tag{9.1}$$

Small changes in base current produced by the input signal will result in bigger changes in the collector current and correspondingly bigger voltage drops in the load resistor R_L placed in the collector circuit to provide voltage amplification. Figure 9.6 shows the circuit of an NPN transistor connected as a CE amplifier. The input signal is applied between the base and emitter through the coupling capacitor C_1 whereas the output signal is taken from the collector via another coupling capacitor C_2. The battery E_{bb} provides the forward bias for the base emitter circuit and R_1 controls the base current to fix the operating point for the amplifier. The reverse bias for the collector-emitter junction is provided by the battery E_{cc} and R_L is the load resistance in the collector circuit across which the output voltage E_{out} is developed.

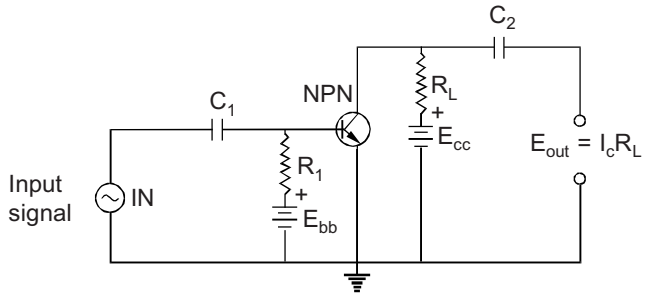


Fig. 9.6 Transistor as a Common-emitter Amplifier

When the operating point of the amplifier is fixed by the base current, a sinusoidal signal current imposed on the steady base current causes the base current to increase and decrease according to the variations of the input AC signal current. The resulting base current has a sinusoidal waveform. The changes in base current produce correspondingly bigger changes in the collector current. The current flowing through the collector circuit also has a sinusoidal waveform, provided the transistor is operated on the linear portion of the characteristics. The sinusoidal base current and the sinusoidal collector current are in phase. The sinusoidal collector current i_c flowing through the load resistance R_L will produce a sinusoidal voltage $i_c \times R_L$, which is the output voltage produced by the CE amplifier. The output voltage E_{out} will, however, be 180° out of phase with the base current i_B as in the case of a triode amplifier.

If the CE amplifier is overdriven to nonlinear portions of its characteristics due to a bigger input signal swing, the resulting output waveform will not be sinusoidal and the amplifier output will be distorted as in Fig. 9.7.

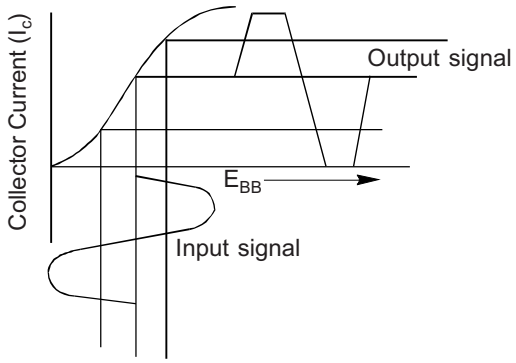


Fig. 9.7 Overdriven CE Amplifier

If a PNP transistor is used for the amplifier instead of the NPN transistor in Fig. 9.6, the bias polarities will be reversed and the effect of signal polarities will be just the opposite. However, the amplifying properties of the circuit will remain the same.

Biasing Methods for CE Amplifiers An important property of a CE amplifier is that it is quite suitable for base biasing from a single power supply source rather than the use of two bias batteries as in Fig. 9.6. This is made possible because in a CE amplifier the base and collector bias polarities are the same. Figure 9.8 shows two practical methods of biasing a CE amplifier from a common power supply source.

In the arrangement of Fig. 9.8(a), the base bias is determined by the voltage divider formed by R_1 and the base-emitter resistance (R_{BE}) drawing current from battery E_{cc}. R_3 is a stabilising resistor which prevents thermal runaway by

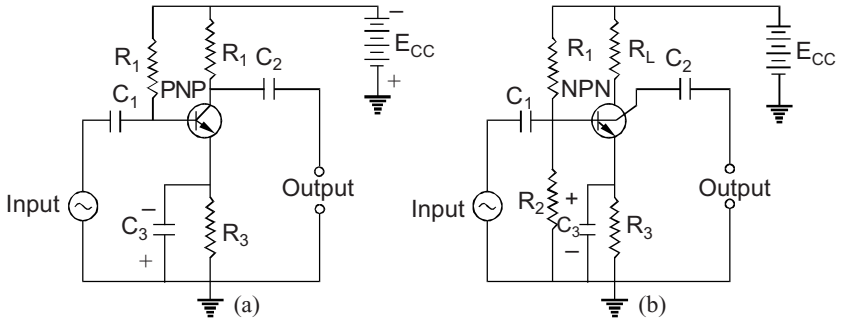


Fig. 9.8 Biasing a CE Amplifier with a Single Supply Source
(a) Common Supply with DC Stabilisation
(b) Potential Divider with DC Stabilisation

developing a voltage whose polarity is opposite to the forward bias applied to the base. Any increase in the collector current causes a larger voltage drop across R_3 which reduces the base current thereby bringing down the collector current. The bypass capacitor C_3 prevents any AC signal voltage developing across R_3 which would otherwise oppose the input signal voltage and reduce the gain of the amplifier. Figure 9.8(b) is the potential divider method of biasing the base emitter junction in an NPN transistor amplifier. With the polarity of voltage developed across R_3 as shown, the difference between the emitter voltage and the base voltage provide the bias voltage for the base bias current. Resistor R_3 compensates for the temperature changes and any variation of β with different transistors of the same type number.

The voltage gain of a CE amplifier is the ratio of the output voltage e_{out} to the input voltage e_{in}

$$\text{Voltage gain} = \frac{e_{out}}{e_{in}}$$

This voltage gain can be measured experimentally with a VTVM and an audio oscillator. The audio oscillator, set for 1000 Hz, is connected at the input of the amplifier and its voltage adjusted for a value of say 0.5 V with the help of the VTVM connected across the input. The VTVM is then connected across the output terminals and the output voltage measured on the appropriate scale. If the output voltage is say 10 V, then the voltage gain of the amplifier at 1000 Hz is equal to 10 V/0.5 V = 20.

9.4 MULTISTAGE AMPLIFIERS

Amplifier Coupling Methods Where the amplification provided by a single amplifier is not sufficient to be applied directly to an output device such as a loudspeaker or the recording head of a tape recorder, it becomes necessary to increase the amplification by connecting together two or more stages of amplifiers in such a way that the output signal of one amplifier serves as the input signal for the second amplifier, and so on. Two or more amplifiers connected in this fashion are said to be connected in *cascade*. For connecting two amplifiers in cascade certain coupling methods are required to be used. There are four coupling methods in general use which are applicable to both the vacuum tube amplifier and transistor amplifiers. These methods are (i) resistance-capacitance (*RC*) coupling, (ii) impedance-coupling, (iii) transformer coupling, and (iv) direct coupling, *RC* coupling and transformer coupling are the most popular methods of coupling in audio amplifiers.

9.4.1 RC Coupling

In the *RC* coupled amplifier shown in Fig. 9.9, the signal voltage developed across the load resistor R_L of Q_1 is coupled into the base of Q_2 through the coupling capacitor C_c which also isolates the base circuit of Q_2 from the collector of Q_1. When two or more stages in an amplifier are connected together, the

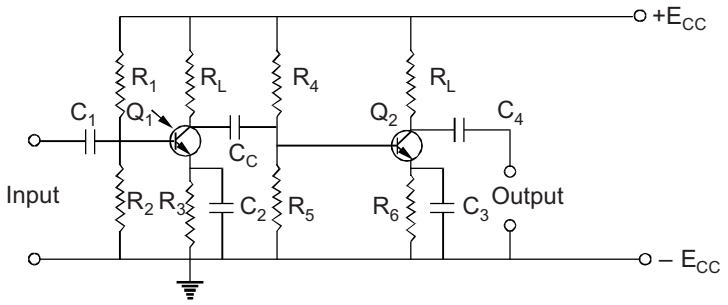


Fig. 9.9 RC Coupled Transistor Amplifier

transistors in the first, second and subsequent stages are represented by Q_1, Q_2, Q_3, respectively. The two *RC* coupled stages in the above amplifier are identical. Each stage has a potential divider method of base biasing with bias stabilisation provided by suitable resistors (R_3, R_6) and bypass capacitors (C_2, C_3) in the emitter circuit. C_4 couples the output signal from Q_2 to its output load.

If the reactance of the coupling capacitor C_c is negligibly small at the frequencies of the input signal, the base resistance R_5 comes in parallel with the load resistance R_L of Q_1. As a result the actual load resistance for Q_1 which supplies the voltage that determines the value of signal current flow in the input base circuit of Q_2 is reduced. This means that the *RC* coupling changes the output voltage of Q_1 from its original value before coupling.

RC coupling has the advantage of providing excellent audio quality over a wide range of frequencies and is also cheap compared to other methods of coupling. This method of coupling is widely used in audio amplifiers.

9.4.2 Impedance Coupling

Impedance coupling derives its name from the fact that in the collector circuit of the amplifier, an inductor coil or a choke having impedance is introduced in place of the load resistance as shown in Fig. 9.10. The output from the first stage is coupled on to the base of the next stage in the usual way through a coupling capacitor. The coil L has a low DC resistance and its inductive reactance $2\pi fL$ varies with frequency. It is, therefore, possible to get a large value of load impedance and correspondingly higher output AC voltage without any excessive DC voltage drop that occurs in the case of RC coupling. However, the amplification is rather non-uniform being high at higher frequencies and low at lower frequencies. This type of coupling is rarely used.

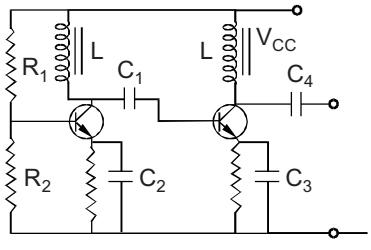


Fig. 9.10 Impedance Coupling

9.4.3 Transformer Coupling

Of all the coupling methods used in amplifiers, transformer coupling is by far the most popular coupling method. The *interstage transformer* not only provides a coupling method but also provides a means for matching the output impedance for the first stage to the input impedance of the next stage. Proper impedance matching ensures the maximum transfer of power from one stage to the next. Interstage transformer also provides the required isolation between the DC collector voltage of one stage and the base of the next stage without coupling capacitors. Moreover, the primary reactance has a high impedance and low DC resistance as in the case of impedance coupling. This permits the use of the higher output voltage from the transformer. This output voltage from the first stage can even be stepped up with a suitable step up transformer.

Figure 9.11 illustrates a typical transformer coupled transistor amplifier. Transformer T_1 is the interstage coupling transformer where T_2 is the output transformer which enables proper matching conditions to be provided between the output impedance of the amplifier and the impedance of the load which may be a loudspeaker or any other device. The two transistor amplifier stages Q_1 and Q_2 are conventional CE amplifiers using PNP transistors and a potential divider biasing arrangement with bias stabilisation. NPN transistors can be used with suitable circuit modifications.

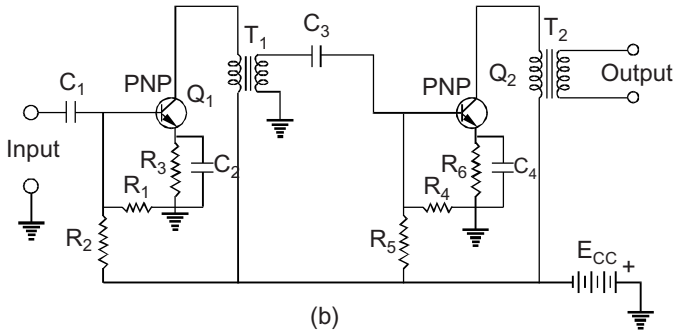


Fig. 9.11 Transformer Coupled Amplifier

Transformer coupling will normally provide a non-uniform amplifier response with peaks at high frequencies due to resonance between the secondary reactance and the inter-electrode capacitances and the shunt capacitances of the windings. Unless suitably designed, the transformer coupled amplifier will not provide the type of response expected of a high quality audio amplifier. High quality with a transformer coupled amplifier will mean larger size, more weight and high cost of equipment.

9.4.4 Direct Coupling

Figure 9.12 is the circuit of a *direct coupled* transistor amplifier. The collector of Q_1 is directly coupled to the base of Q_2. Direct coupling saves the cost of

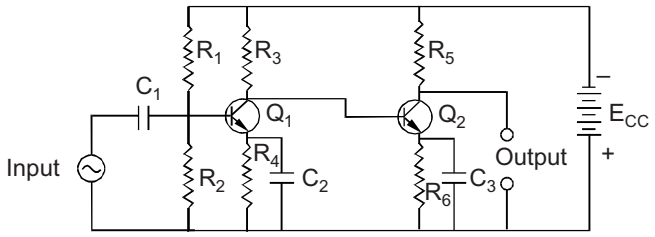


Fig. 9.12 A Direct Coupled Amplifier

components used for coupling the amplifier stages. Elimination of the frequency sensitive components like transformers and capacitors results in much improved frequency response in direct coupled amplifiers.

Direct-coupled amplifiers are useful for the amplification of pulse signals.

9.5 CHARACTERISTICS OF AN AMPLIFIER

9.5.1 Frequency Response of an Amplifier

The frequency response of an audio amplifier is the ability of the amplifier to *equally* amplify the signal voltages in the audio range of frequencies (20 to 20,000 Hz). The frequency response is generally expressed as a graph plotted between the frequency on the *x*-axis and the voltage gain or output along the *y*-axis. A typical frequency response curve for an audio amplifier is shown in Fig. 9.13.

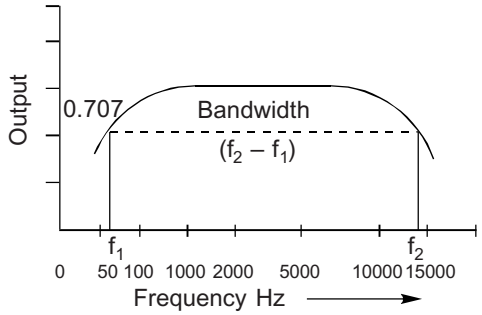


Fig. 9.13 Typical Frequency Response for an Audio Amplifier

The *bandwidth* of the amplifier is represented by the flat portion of the frequency response curve indicating a uniform response. The bandwidth is defined as the difference between the frequencies f_2 and f_1, where the gain of the amplifier is 0.707 of the maximum gain at any frequency. Thus, if the gain of the amplifier drops down to 0.707 of the maximum gain (at 1000 Hz) at f_1 = 50 Hz, and f_2 = 12,000 Hz, the bandwidth of the amplifier is

$$f_2 - f_1 = 12000 \text{ Hz} - 50 \text{ Hz} = 11950 \text{ Hz}$$

The larger the bandwidth of an audio amplifier, the better is quality of the sound reproduction with this amplifier. Audio amplifiers used in Hi-Fi equip-

ment have a much larger bandwidth than the audio amplifiers in radio or TV receivers.

Frequency response is a very important characteristic of an audio amplifier. It can be measured experimentally with the help of the set up shown in Fig. 9.14. A voltage input of say 1 V is applied at the input of the amplifier from an AF signal generator and this voltage is kept constant at 1 V with the help of a

Fig. 9.14 Measuring the Frequency Response of an Amplifier

VTVM, for the entire range of frequencies over which the frequency response of the amplifier is measured. Starting from a low frequency of 30 Hz, a voltage of 1 V is applied at the input at each frequency and the output voltage is measured across the output terminals with the help of the same VTVM or another VTVM using an appropriate range. The output for the same constant input voltage is taken at shorter frequency intervals of about 50 Hz up to say 500 Hz and then at bigger intervals of about 1 kHz at frequencies above 500 Hz up to 10 to 12 kHz. A graph of output voltage versus frequency is then plotted as in Fig. 9.15.

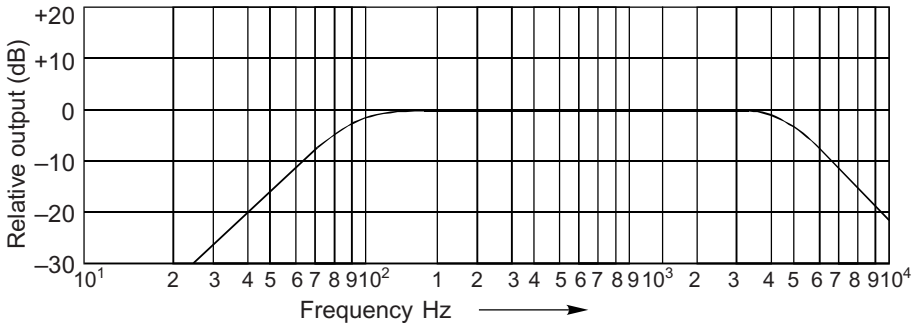


Fig. 9.15 Frequency Response Curve on a Semi-log Paper

The frequency range being very large (10 to 10,000 Hz), cannot be easily indicated on linear graph paper. For plotting frequency response semi-log graph papers are used. In these graph papers the scale along the *x*-axis is logarithmic whereas the scale along the *y*-axis is linear. Frequencies like 10, 10^2, 10^3, 10^4, can be conveniently represented on the log scale whereas the output voltages can be represented on the linear scale along the *y*-axis. A frequency response curve plotted on a semi-log graph paper is shown in Fig. 9.15.

An examination of the shape of the frequency response curve indicates that, compared to the middle range of frequencies, the gain of the amplifier shows a downward trend both at the higher frequencies and at the lower frequencies. The loss in gain at high frequencies is due to the shunting effect produced by

the inter-electrode capacitances of the tubes (Transistors), and the stray wiring and other distributed capacitances. The response at low frequencies (below 100 Hz) is limited by the coupling capacitance and the value of the bypass capacitors.

9.5.2 Distortion and Noise

An amplifier is said to possess distortion when the output waveform produced by the amplifier is different from the input waveform. The quality of output is different from the input quality and does not sound natural. Amplifier distortion is of three types: (a) frequency distortion, (b) phase distortion, and (c) amplitude distortion.

(a) Frequency Distortion When the amplifier is not able to amplify a complete range of frequencies, the distortion is known as *frequency distortion*. If the output is deficient in HF, the sound will appear to be boomy or bassy. If the low frequencies are not properly amplified the sound is tinny.

(b) Phase Distortion The phase shift produced by the various capacitive coupling elements in an amplifier is not the same for all frequencies and as a result the output waveform is different from the input waveform. This is called *phase distortion*. Phase distortion is not very important in audio amplifiers where it cannot produce any audible effects but this type of distortion is important in video amplifiers used in TV receivers.

The response of an amplifier to *transients*, which are sharply rising and suddenly changing waveforms, depends on its frequency response and its phase-shift characteristics. An amplifier with a poor *transient response* cannot reproduce satisfactorily the sound of drums and other percussion instruments.

(c) Amplitude Distortion Amplitude distortion is the result of operating a tube or a transistor on the non-linear portion of its characteristics. In this case, the relative amplification produced in the two halves of a sine wave is not the same, and the sine wave gets flattened and clipped either at the top or at the bottom. Such a distorted waveform can be shown to be the combination of a fundamental sine wave frequency and its multiple frequencies, called harmonics. The amount of distortion depends on the number and the amplitudes of the harmonics. This type of distortion is also known as harmonic distortion.

9.5.3 Noise

Besides the desired voltages applied to the input of an amplifier for amplification, there are always some undesirable voltages which are either produced internally by random movement of electrons in tubes, transistors, resistors and other components, or are induced into the circuit externally from components like transformers and chokes. These unwanted voltages called noise also get amplified with the actual signal voltages and can reach alarming proportions to completely *mar* the quality of output from the amplifier. The amount of noise or noise level is kept very low by design methods, including choice of

components and their placement in circuit. *Signal-to-noise ratio* is another important characteristic by which the quality of audio amplifiers is judged.

One of the methods used to improve the frequency response and signal-to-noise ratio in audio amplifiers and to reduce distortion is the use of *negative feedback.*

9.6 FEEDBACK IN AMPLIFIERS

Feedback is the process by which a portion of the output energy (current or voltage) is transferred to the input of the amplifier. When the feedback voltage or current is in phase with the input signal and aids it, the feedback is called *positive or regenerative* feedback. If, however, the energy fed back is opposite in phase to the input signal, the feedback is *negative* or *degenerative* feedback.

Positive feedback increases the gain of the amplifier which starts oscillating when the positive feedback is sufficiently large. Positive feedback forms the basis of an important type of circuits called *oscillators* which will be described later. Negative feedback reduces the gain of the amplifier and in the process makes the amplifier stable. Negative feedback plays an important part in audio amplifiers because it improves the frequency response of the amplifier and reduces noise and distortion.

The operation of a feedback amplifier can be explained by the block diagram shown in Fig. 9.16. In this case, the feedback is applied by feeding a fraction β of the output voltage back to the input signal voltage. The path over

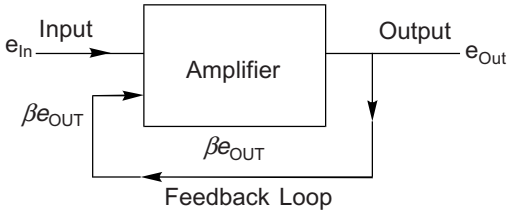


Fig. 9.16 Feedback Amplifier

which the feedback is applied is the feedback loop. With feedback, the actual input to the amplifier becomes $e_{in} + \beta_{out}$. If A is the voltage gain of the amplifier without feedback the new input with feedback when amplified by the gain A will become the output voltage e_{out}, or in mathematical terms,

$$(e_{in} + \beta e_{out})A = e_{out} \tag{9.2}$$

Dividing both sides of the equation by e_{in}

$$\left(1 + \beta \frac{e_{out}}{e_{in}}\right) A = \frac{e_{out}}{e_{in}} \tag{9.3}$$

If the voltage gain with feedback is represented by A', then

$$A' = \frac{e_{out}}{e_{in}} \tag{9.4}$$

Substituting this value in Eq. (9.3), we get

$(1 + \beta A')\, A = A'$, which when simplified gives

$$A' = \text{voltage gain with feedback} = A/(1 - \beta A) \tag{9.5}$$

In the case of negative feedback, fraction β is negative, and the expression for the voltage gain with feedback becomes

$$A' = \frac{A}{1-(-\beta A)} = \frac{A}{1+\beta A} \tag{9.6}$$

This indicates a decrease in the gain of the amplifier.

When a large amount of feedback is applied, βA will be large compared to 1 and Eq. 9.6 will reduce to

$$A' = \frac{1}{\beta} \tag{9.7}$$

This is a very significant relation. It means that with large negative feedback, the voltage gain of the amplifier becomes independent of the actual gain A of the amplifier. As a result, the amplifier gain becomes stable and is practically independent of supply voltage fluctuations, ageing of tubes and transistors and differences in tubes and transistors due to replacement. A large amount of negative feedback also improves the frequency response because the gain is controlled only by the nature of β, and with a resistive feedback loop, the gain will not vary with frequency as long as the feedback is negative.

The feedback can be made selective by designing the feedback loop with capacitive and inductive components and any desired frequency response obtained. Feedback circuits can, therefore, provide the tone control in audio amplifiers. Feedback circuits are also used in record players and tape recorders for providing necessary compensation or equalisation.

9.6.1 Negative Feedback Circuits

There are two practical methods of applying feedback in audio amplifiers. These are the *voltage feedback* and the *current feedback.*

(a) Voltage Feedback In a single amplifier, the output is 180° out of phase with the input voltage and a portion of the output voltage can be tapped and applied directly to the base of the transistor as in Fig. 9.17. This is *voltage feedback* because the feedback is derived from the output voltage. Resistances R_1 and R_2 form a voltage divider across the output voltage and capacitor C

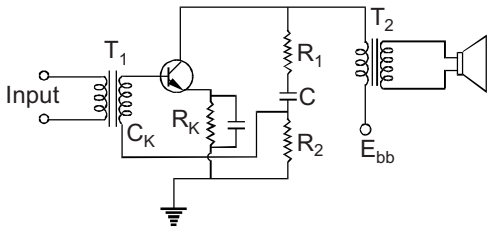


Fig. 9.17 Voltage Feedback

blocks the DC voltage from the base. The AC voltage developed across the resistor R_2 is applied to the base via the secondary of the input transformer. The fraction of the output voltage actually applied as negative feedback is $R_2 l(R_1 + R_2)$ and this is the fraction mentioned earlier. This type of feedback is commonly employed in the output stages of audio amplifiers particularly those using high power transistors. The reduction in gain caused by negative feedback has to be made up either by increasing the input signal voltage or by introducing extra stage of amplification.

(b) Current Feedback In *current feedback* the negative feedback applied to the base is developed across a resistor R_k in the emitter circuit due to the plate current flowing through this resistor as in Fig. 9.18. This resistor R_k is similar to the self-bias resistor but without the *bypass* capacitor. Figure 9.18 is a transistor amplifier using an unbypassed resistor in the emitter circuit to provide current feedback. With negative feedback the effective input signal is the difference between the applied input signal and the signal developed across the emitter resistor R_k. The current feedback results in the increased input and output impedance and lower stage gain.

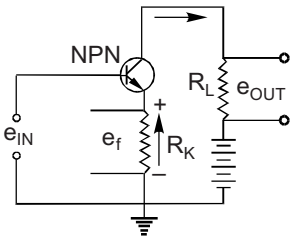


Fig. 9.18 Current Feedback in a Transistor Amplifier

In a two-stage amplifier, the output voltage in the second stage is in phase with the input signal and hence the degenerative feedback cannot be applied by a portion of the output signal fed back directly into the base (or grid) of the first stage. In this case a portion of the signal from output stage (Q_2) is coupled back to the emitter of input stage (Q_1) as shown in Fig. 9.19. The unbypassed emitter resistor R_2 develops its own negative feedback and the signal feedback via R_3C_3 is in phase with this voltage. The total input signal is the difference between the input signal e_{IN} and the sum of the in phase feedback voltages developed across R_2. The fraction of the output voltage fed back (β) can be calculated from the voltage divider consisting of R_3, C_3 and R_2.

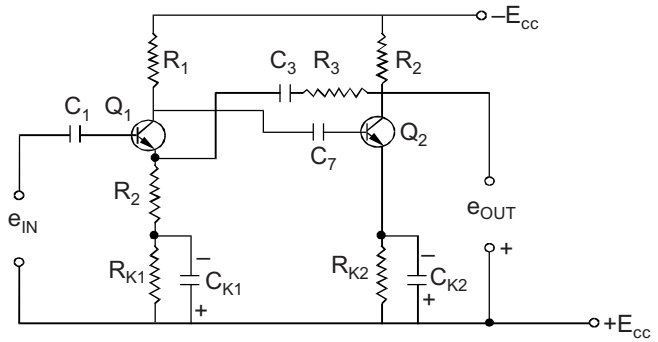


Fig. 9.19 Two-stage Negative Feedback Amplifier

(c) Emitter Follower When the entire load of an amplifier is connected in the emitter circuit as R_L in Fig. 9.20 and the output voltage is taken off the resistor R_L, the circuit is known as emitter follower. The name emitter follower is derived from the fact that in this type of circuit the voltage developed across R_L follows the sign (in-phase) of the input signal. As can be seen from Fig. 9.20, a emitter follower works on the principle of current feedback, as the negative feedback voltage developed across R_L is due to the emitter current flowing through this resistor. Moreover, the entire output voltage (e_{out}) developed across R_L is fed back into the base circuit as degenerative feedback. In other words, β in this case is one as can be seen from the expression for gain of the amplifier with feedback, the emitter follower has a gain less than one or it can at the most equal one. Although a emitter follower has no voltage gain, it can have considerable power gain.

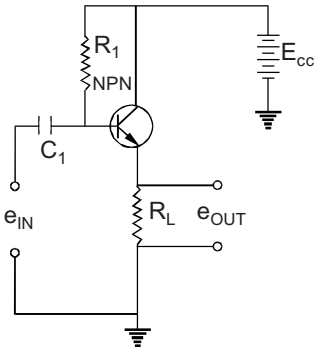


Fig. 9.20 Emitter Follower Circuit

As a result of high degenerative current feedback an emitter follower develops a very high *input* impedance and a low output impedance. This makes the emitter follower a very good impedance matching device like a transformer.

Figure 9.20 is the circuit of an emitter follower. It is actually the common collector (CC) amplifier in transistor as already discussed. The voltage gain is less than unity but it exhibits current and power gain. It has impedance characteristics similar to a cathode follower in a valve circuit and serves as a useful impedance matching device.

9.7 POWER AMPLIFIERS

9.7.1 Difference Between a Voltage Amplifier and a Power Amplifier

A voltage amplifier raises the voltage level of the input signal and does not develop any significant amount of power in the output. On the other hand, a power amplifier is meant to raise the power level of the input signal. It can supply a large amount of output power to the load which is generally a loudspeaker in audio power amplifier. A voltage amplifier normally precedes a power amplifier. The larger the input voltage, the greater is the output power of the power amplifier. This is the reason why power amplifiers are also called large signal amplifiers.

A power amplifier does not amplify power. It merely converts the dc power supply to the amplifier into an ac output signal which supplies power to the load. Its action is controlled by the input ac signal.

9.7.2 Single Ended Power Amplifier

A single ended power amplifier is a power amplifier that makes use of only one transistor compared to a push-pull power amplifier (described later) which uses two transistors. A single-ended power amplifier is shown in Fig. 9.21. The input comes from a voltage amplifier called a pre-amplifier and the output is supplied to the load R_L through the transformer which not only couples the output of the amplifier to the load R_L but also matches the output impedance of the amplifier to the impedance of the load R_L.

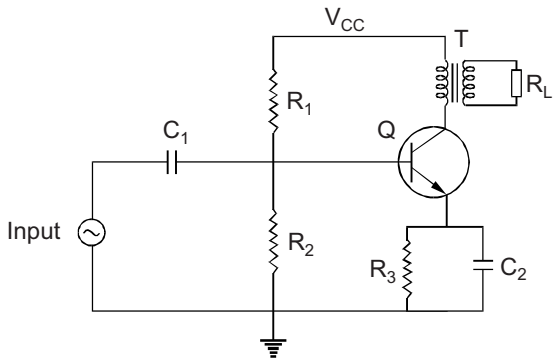


Fig. 9.21 Single Ended Amplifier

Such single ended amplifiers are commonly used in the power output of radio receivers, television receiver, tape recorders to drive a loudspeaker, which has an impedance of about 8 ohm. The transformer T is an impedance matching transformer which matches the output impedance of the amplifier which is a few thousand ohm to the low impedances (8 ohm) of the loudspeaker to ensure maximum transfer of power from the amplifier to the loudspeaker. Impedance matching transformers have already been described in Part I-Chapter 2.

9.7.3 Push-Pull Amplifiers

A push-pull amplifier makes use of two transistors either in Class A or Class B configuration.

A Class B push-pull amplifier consists of two identical transistors which are biased approximately to cut-off and are so operated in circuit that the two bases are excited with two equal but opposite (out of phase) voltages from the input signal. When operated in this manner, one of the transistors amplifies the positive half-cycles of the signal voltage while the other one amplifies the negative half-cycles. The amplified half cycles are then combined in the output transformer in such a manner as to produce an amplified reproduction of the input voltage. A schematic diagram of a simple push-pull amplifier is shown in Fig. 9.22(a) and an equivalent circuit is shown in Fig. 9.22(b).

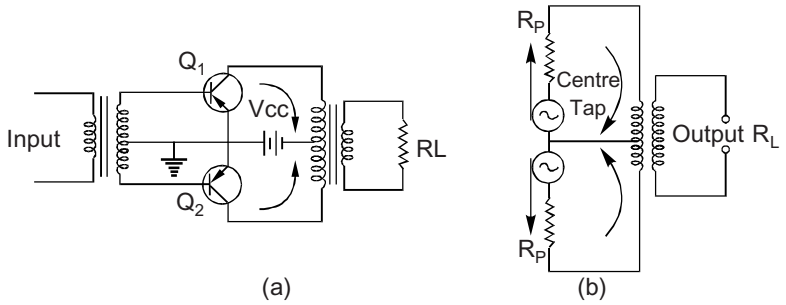


Fig. 9.22 Push-pull Amplifier (a) Schematic Diagram (b) Equivalent Circuit

A push-pull amplifier possesses the following advantages:

1. There is no DC saturation of the core of the output transformer because the DC plate currents flow through the output transformer in opposite directions due to the two transistors. This reduces the size of the output transformer.
2. As the AC signal frequency current does not pass through the common supply, the push-pull amplifier has no regenerative effect on other stages. The common emitter resistor for the two transistors does not require a bypass capacitor.
3. It has less amplitude distortion in the output due to cancellation of all even order (2nd, 4th, etc.) harmonics.
4. Any AC hum or ripple currents from the DC power supply sources balance out in the circuit and there is no hum in the output.
5. A Class B push-pull amplifier without input signal draws very little collector or plate current resulting in great economy in battery power. The system is, therefore, suitable for battery operated equipment including transistor radio receivers.

Push-pull amplifiers can be either Class A or Class B amplifiers. However, push-pull amplifiers are mostly operated as Class B or Class AB in audio circuits because of higher output and increased output efficiency for a given tube or transistor type.

9.7.4 Transistor Push-pull Amplifiers

Two transistors can also be connected in a push-pull circuit. A Class B push-pull amplifier can handle a signal amplitude approximately twice as large as that of a Class A power amplifier and, as a result, can deliver more than double the output power from a single-ended Class A amplifier. Moreover, a Class B push-pull amplifier has a lower power dissipation rating of one-fifth of the total output load power and its no signal current or quiescent current is very low. Because of these advantages, transistor push-pull circuits find extensive use in communication transmitters, high power audio amplifiers, and battery operated devices like transistor radios.

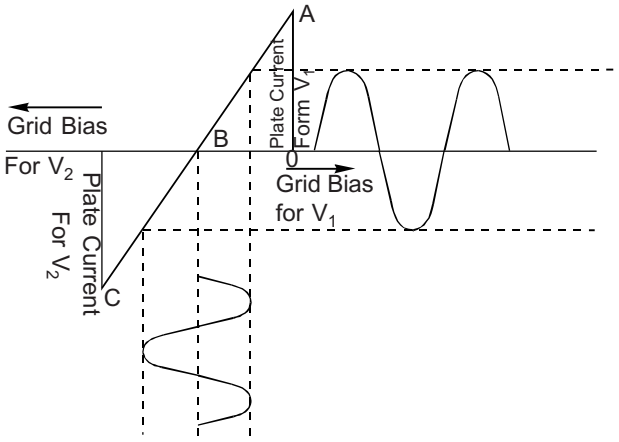


Fig. 9.23 Composite Characteristics of Push-pull Amplifier

9.7.5 Transformer—Coupled, Push-pull Amplifiers

The conventional method of connecting two transistors in a transformer coupled push-pull circuit is shown in Fig. 9.24. A centre-tapped input transformer T_1 is used to supply two signals equal in amplitude but opposite in phase to the bases of the transistors Q_1 and Q_2 which is a matched pair of NPN type transistors. The outputs from the two transistors are combined in an output transformer T_2 which also serves as a matching transformer for matching the load R_L to the output impedance of the push-pull amplifier. The load R_L is generally a loudspeaker with 3-5 Ω impedance.

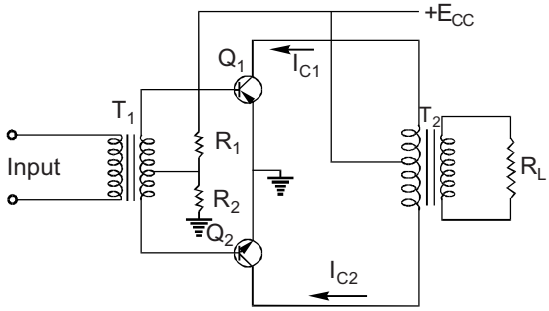


Fig. 9.24 Transformer-coupled Push-pull Amplifier

The voltage divider R_1 and R_2 provides the forward bias for the two NPN transistors Q_1 and Q_2 and the values of R_1 and R_2 are selected for Class B operation. When an AC input signal is applied to transformer T_1, transistor Q_1 conducts during the positive half-cycle of the input signal voltage and a current I_{C1} flows through one half of the primary of T_2. Q_2 remains cut off when Q_1 conducts. During the negative half cycle of the input voltage, Q_2 conducts and Q_1 remains cut off. A current I_{C2} flows through the other half of the primary of T_2 but in the opposite direction. However, when the current I_{C1} is increasing,

the current I_{C2} is decreasing and vice versa. As a result, the magnetic fields induced by the two collector currents are in the same direction and the voltage induced in the secondary of T_2 is much higher than that induced by each current alone.

In actual practice, the two transistors are not biased to complete cut off but the biasing voltages are so adjusted that a small collector idling current (quiescent current) flows through Q_1 and Q_2 even without signal. This is to prevent what is known as *crossover distortion*.

Crossover Distortion In a Class B push-pull amplifier using silicon transistors, if the two transistors are biased to cut off with zero bias on the base-emitter junction, the incoming signal voltage must rise to the barrier potential voltage of about 0.7 V before the transistor Q_1 can start conducting. In the same way the action of Q_2 on the negative half cycle is complimentary and it will also not conduct till the negative voltage reaches a value of approximately –0.7 V. There is, therefore, a period of time T between the positive and negative half cycles of the input voltage when no current flows through the circuit. There is thus a gap between the time when one transistor shuts off and the other is turned on and this produces a clipping effect on the output voltage waveform as shown in Fig. 9.25. The output voltage waveform does not resemble the input waveform and the output is distorted. This type of distortion is called *crossover distortion*.

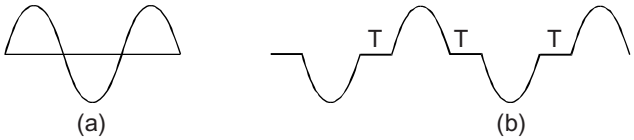


Fig. 9.25 Crossover Distortion in Class B Push-pull Amplifiers (a) Input Waveform (b) Output Waveform

To eliminate crossover distortion, the transistors in a Class B push-pull amplifier are not biased to complete cut off but are slightly forward-biased to allow a small amount of collector current to flow even in the absence of signal. Although the operation of the amplifier does not remain strictly Class B it is still known as a Class B push-pull operation.

9.7.6 Complementary Symmetry Push-pull Amplifiers

A PNP and NPN transistor with identical electrical characteristics is said to form a complementary pair. A push-pull amplifier with a complementary symmetry can be arranged as in Fig. 9.26.

Initially, the two transistors Q_1 (NPN) and Q_2 (PNP) are equally biased so that each draws the same amount of current in the absence of the signal. When an input signal with a sinusoidal waveform is applied to the common base, the NPN transistor Q_1 is turned on during the positive alternation while the PNP transistor Q_2 is cut off and remains reverse-biased during the positive alternation. Only the positive half of the output current waveform is reproduced by

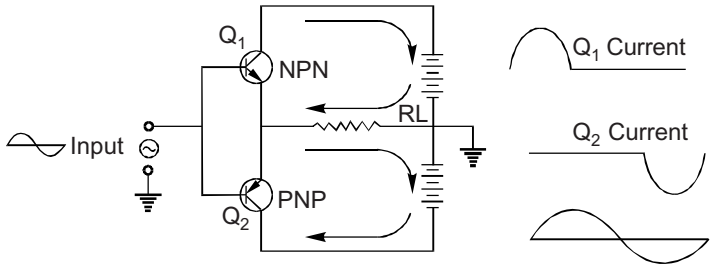


Fig. 9.26 Complementary Symmetry Push-pull Amplifiers with Current Waveforms

Q_1. During the negative alternation, however, the NPN transistor Q_2 is tuned on when it is forward-biased while the NPN transistor Q_2 is reverse biased and remains cut off during the negative alternation. Q_2, therefore, reproduces the negative half of the output current waveform. The two halves of the output current wave flowing through the common emitter load R_L in opposite directions produce an output voltage which has the same waveform as the input signal. Transistors Q_1 and Q_2 complement each other and the arrangement is called *complementary symmetry*. Moreover, complementary symmetry push-pull amplifiers can be either Class A or Class B type.

The complementary symmetry circuit described above makes use of two power supplies but the arrangement shown in Fig. 9.27 uses only a single power supply source. Symmetrical voltage dividers R_1 and R_2 are used for the two transistors so that the top dividers provide a forward bias to Q_1 and the bottom dividers a forward bias to Q_2 which allows a small idling current to flow through the circuit to avoid distortion.

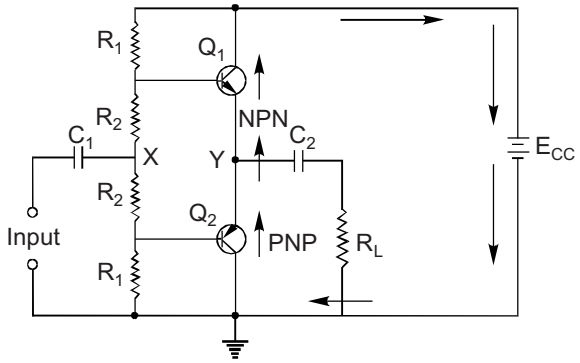


Fig. 9.27 Complementary Symmetry Push-pull Amplifier with Single Power Supply Source

Due to the symmetry of the entire circuit, the points X and Y form the DC voltage mid-points for the battery, and whereas the base of Q_1 is positive with respect to its emitter Y, the base of the Q_2 is negative with respect to Y which is also the emitter of Q_2. When the input signal is applied at X through C_1, it is

equally divided between the bases of Q_1 and Q_2. The transistor Q_1 is turned on during the positive alternation while Q_2 is cut off, and the transistor Q_2 is turned on during the negative alternation when Q_1 is cut off. The rest of the action is the same as in the case of the circuit with two power supplies. The output voltage is developed across the load R_L which in the case of complementary symmetry audio amplifiers is the voice coil of the loudspeaker. The emitter-follower design of the circuit permits low impedance loudspeaker to be connected directly between the emitter and the ground and the capacitor C_2 can be eliminated by proper design of the circuit. Complementary-symmetry Class B push-pull amplifiers are often used as output stages in audio amplifiers. One great advantage of this type of amplifiers is that it permits the operation of Class B amplifiers without the use of input and output transformers. This reduces cost and improves the quality of the audio equipment.

It may be mentioned that complementary symmetry circuits have to be carefully designed to prevent thermal runaway. Unless special stabilisation methods like diode stabilisation are adopted, any imbalance or leakage in the power transistors can result in repeated failure of the output power transistors.

9.8 RF AMPLIFIERS

Audio amplifiers are amplifiers which are meant to equally amplify all frequencies lying within a wide range of frequencies called audio frequencies (20 Hz to 20 kHz). There is no selection of any particular frequency or rejection of other frequencies. When used at higher frequencies, above the audio range, these AF amplifiers become less efficient. However, there is another type of amplifiers called RF (radio frequency) amplifiers which select and amplify a particular high frequency or a narrow band of high frequencies above the normal audio frequency range. RF amplifiers are commonly used in radio and TV circuits which operate on specific radio frequencies assigned to their respective transmitters. Both transistors and vacuum tubes can be used in RF amplifiers with only slight modifications in design.

9.8.1 Tuned RF Amplifiers

Tuned circuits are used for the selection and efficient amplification of a specific radio frequency or a narrow band of frequencies in conjunction with transistor or vacuum tube amplifiers. A tuned circuit is only a resonant circuit formed by an inductor and a capacitor connected in parallel or in series as shown in Fig. 9.28.

In the circuit in Fig. 9.28(c), the voltage induced in the coupled secondary is in series with L and C and the circuit is a series-tuned circuit.

Properties of resonant circuits have already been described but for the sake of continuity it may be mentioned again that at resonance a series resonant circuit has minimum impedance and maximum current and a parallel resonant circuit exhibits maximum impedance and minimum line current at the frequency of resonance.

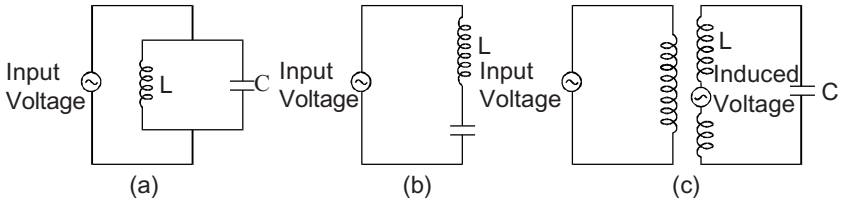


Fig. 9.28 Tuned Circuits (a) Parallel Resonant Circuit (b) Series Resonant Circuit (c) Inductively Coupled Series Resonant Circuit

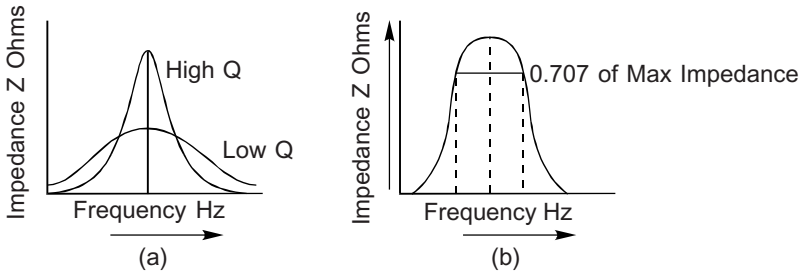


Fig. 9.29 Selectivity and Bandwidth in a Parallel Tuned Circuit
(a) Effect of *Q* on Selectivity
(b) Bandwidth of a Parallel Tuned Circuit

A parallel resonant circuit also known as a *tank circuit* is often used in RF amplifiers and it develops a high voltage across it at the applied resonant frequency. At frequencies above and below the resonant frequency, the impedance and hence the voltage developed across the resonant circuit decreases. The extent to which these conditions change at frequencies above and below resonance, determines the ability of the circuit to select certain frequencies in preference to other frequencies. This ability to discriminate between frequencies is also known as selectivity and depends on the Q-factor (X_L/R) of the circuit. The larger the Q of the circuit (less resistance) the greater the selectivity of the circuit as explained in Fig. 9.29. In a parallel tuned circuit, the bandwidth or bandpass is the frequency range ($f_2 - f_1$) between two frequencies f_1 and f_2 at which the circuit impedance is 0.707 times the value of impedance at the resonant frequency f_r as seen in Fig. 9.29(b).

The bandpass or bandwidth can be calculated from the formula

$$\text{Bandwidth } (f_2 - f_1) \text{ in Hz} = \frac{f_r}{Q}$$

where f_r is the resonant frequency in Hz.

Whenever a resonant circuit is tuned to different frequencies as in radio and television, L or C can be made variable as in Fig. 9.30.

The frequency to which the circuit is tuned in a parallel position of C or L is given by the formula

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

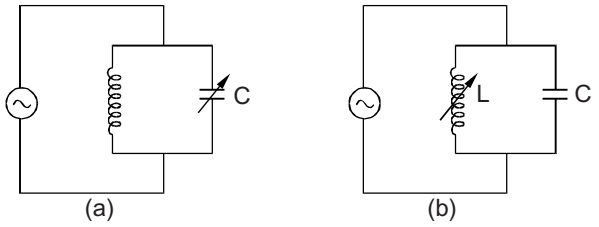


Fig. 9.30 Variable Tuned Circuit (a) Variable *C* (b) Variable *L*

A sharply tuned circuit with a large Q will generally have small bandwidth. In radio and TV circuits, the Q of a tuned circuit can be lowered and bandwidth increased by putting a shunt resistor of suitable value across the tank circuit as in Fig. 9.31.

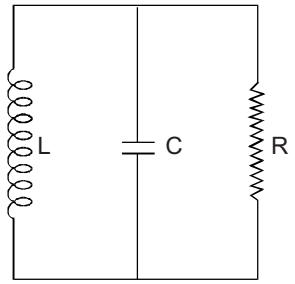


Fig. 9.31 Lowering the Q of a Tuned Circuit by Shunt Resistance *R*

9.8.2 Practical RF Amplifiers

Practical RF amplifiers used in radio and TV receivers generally have a tuned or resonant circuit both in the input and the output. Both the input and output tuned circuits can be made variable, if necessary. Furthermore, the tuned circuits in the input and the output of the RF amplifier must both be tuned to the same frequency. When a new band of frequencies is to be amplified, the two circuits are retuned to the new frequencies. Figure 9.32 gives the simplified circuit of an RF amplifier in a radio receiver.

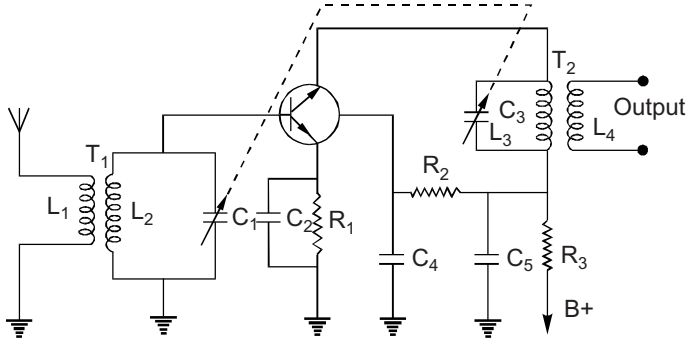


Fig. 9.32 RF Amplifier with Transformer Coupling

The input from the antenna is inductively coupled to the tuned circuit L_2C_1, which develops the maximum input signal for the base at the tuned frequency. Other frequencies develop practically no signal voltage at the input. The output signal voltage is developed across the tuned circuit L_3C_3 and again maximum signal output voltage is developed at the tuned frequency. C_1 and C_3 are

variable capacitors which are normally ganged to keep the input and output tuned circuits always tuned to the same frequency. R_3 and C_5 in the collector circuit are meant for *RF* decoupling. The output is obtained by transformer coupling at the secondary of the transformer T_2 and this output can be applied as input for the next stage of amplification, if any. In certain cases the output signal is capacitively coupled to the next stage as in Fig. 9.33.

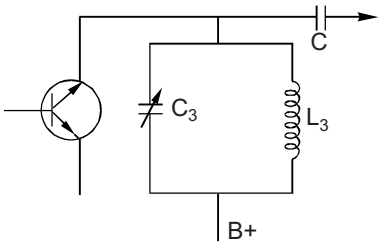


Fig. 9.33 Capacitor Coupling of RF Amplifier

The output load impedance is provided by the parallel resonant circuit (L_3C_3) in the output circuit which offers a high impedance to the AC component of the plate current whereas it offers a low impedance to signal currents that are not of the resonant frequency. Thus, only a particular radio frequency or a narrow band of radio frequencies is amplified.

At very high frequencies on which TV circuits work, the capacitance between the windings of the coil is generally enough to produce resonance at these high frequencies and a separate capacitor is not necessary. The coil is permeability tuned and can be tuned by moving into or out of it a ferrite or a brass slug.

9.8.3 Types of RF Amplifiers

Tuned RF amplifiers are of two types: (a) voltage amplifiers, and (b) power amplifiers.

Voltage Amplifiers In radio and TV circuits where very small RF voltages have to be amplified, the RF amplifiers work as Class A voltage amplifiers which means that the amplifiers are so biased that current flows throughout the complete cycle of signal voltages.

In transmitters where large voltages at RF have to be amplified, the RF amplifiers are power amplifiers in Class C operation. In Class C operation, relatively large power outputs and efficiency can be obtained by biasing the tube (transistor) to cut-off so that current flows for less than half the cycle of operation. Any distortion in the Class C operation is mostly eliminated by the resonant tank circuit by its fly wheel action.

9.8.4 Transistorised RF Amplifiers

Transistor RF amplifiers work more or less in the same way as tube amplifiers by making use of tuned circuits both for the input and the output of the amplifier. The only problem that arises in the case of transistor RF amplifiers is the matching of the input and output impedances of the transistors to the impedances of the tuned circuits. Transistors, particularly when used in the grounded-emitter and grounded-base configurations, have low input impedances and cannot be easily matched to the very high impedances of the tuned

circuits, as is done in the case of vacuum tubes which themselves have high input impedance. This difficulty is overcome by using tapped inductances (or capacitances) in transistor RF amplifiers as in Fig. 9.34. The tapping makes it possible to use the small impedance of the lower portion of the coil for matching the low input and output impedances of the transistors.

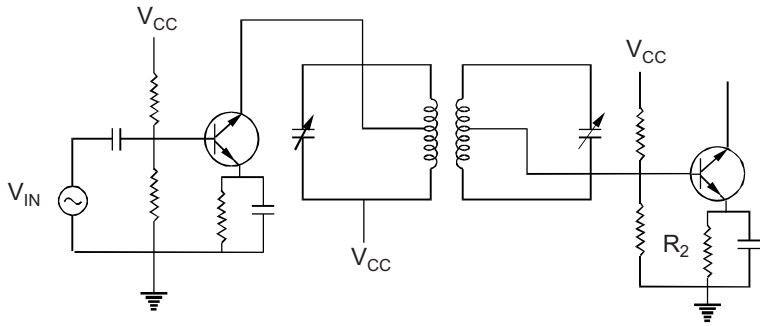


Fig. 9.34 Transistor RF Amplifier

9.8.5 Single -tuned RF Amplifier

Figure 9.35(a) shows the circuit of a single tuned RF amplifier using a NPN type bipolar transistor.

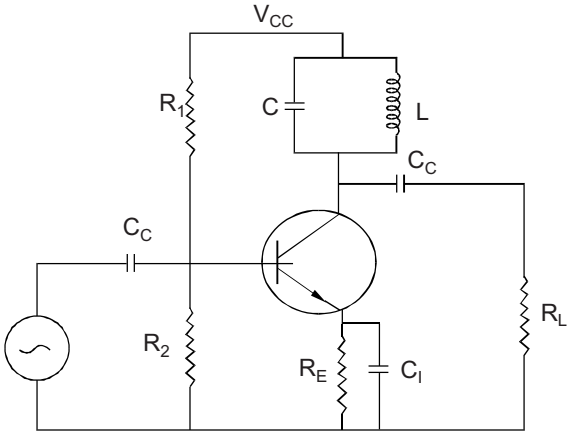


Fig. 9.35 Single-Tuned RF Amplifier

The output is taken with the help of capacitor coupling but transformer (induction) coupling can also be used.

Because of the advantage like very high input impedance and reduced inter electrode capacitance, modern circuits make use of FETs and MOSFETs, in these circuits. Either C or L can be made variable for adjusting the frequency of the circuits.

9.8.6 Double-tuned RF Amplifiers

This amplifier makes use of two tuned circuits as shown in Fig. 9.35. Induction coupling is used.

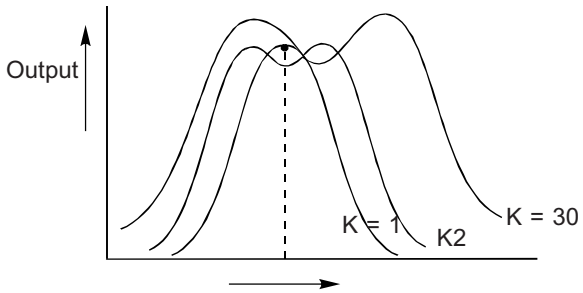


Fig. 9.35(b) Double-tuned RF Amplifier—Frequency Responses for Different Coefficients of Coupling.

The primary and secondary coils of the transformer are shunted by capacitors which can be made variable. A bipolar transistor has been used in the circuit but because of the low input impedance of such transistor, serious limitations in performance are noticed. To avoid this, in the modern solid state circuits, MOSFETs are generally used.

Advantages of a Double-tuned Amplifier A single tuned circuit possesses a high Q and is very selective but its bandwidth is reduced and this results in very poor quality of reproduction.

In the double-tuned amplifier there are two tuned circuits inductively coupled. A change in the coupling of two tuned circuits results in a change in the frequency response. If the coupling between the two coils of the tuned circuits are properly adjusted, the required selectivity gain and bandwidth can be obtained.

Figure 9.35(b) shows the response of the circuit for different co-efficients of coupling K. Most suitable response curves are obtained when optimum coupling exists-between the two tuned circuits. Such circuits are generally used in RF stages of wireless communication sets (Fig. 9.36).

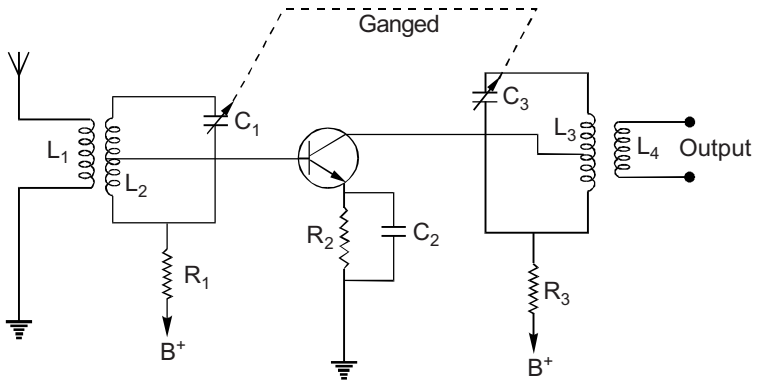


Fig. 9.36 Double-tuned Circuit for RF Stages of Communication Receivers

Summary

An amplifier is an electronic circuit that produces an enlarged version of a small signal fed into the circuit. Amplifiers of different types form the building blocks for a complete circuit of a radio or television receiver.

An audio amplifier is an amplifier that is suitable for the amplification of electrical frequencies that lie within the AF range of 20 Hz to 20 kHz. Audio amplifiers work as voltage amplifiers or power amplifiers depending upon whether they are required to produce high voltage gain or high output power.

The operating conditions of an audio amplifier have to be so adjusted as to give a distortion free output. When the operating conditions permit the flow of current during the entire cycle of the input signal, the amplifier is a Class A amplifier; if the current flow is only for half the cycle the amplifier operates as a Class B amplifier. In a Class C amplifier the current flows for less than half the cycle. Two Class B amplifiers connected together to produce a complete replica of the input voltage are said to form a push-pull Class B amplifier.

Transistors can also be used as amplifiers like vacuum tubes. For satisfactory operation of a transistor as an amplifier the bias conditions have to be such that the emitter-base junction is forward-biased and the collector-base junction is reverse-biased. Of the three configurations, CB, CE and CC in which the transistor can be connected as an amplifier, the CE configuration is the one that is most commonly used in audio amplifiers.

Two or more amplification stages are often coupled together to increase the gain of the amplifier for feeding a loudspeaker. Coupling methods generally used are RC coupling, impedance coupling, transformer coupling and direct coupling. RC coupling and transformer coupling are the most popular methods used in audio amplifiers.

Frequency response, distortion and noise level are important characteristics of an audio amplifier which should have wide band frequency response, low distortion and noise level. Frequency response, distortion and noise figures in audio amplifiers can be improved by the application of a negative feedback. Cathode followers in vacuum tube amplifiers and emitter-followers in transistor circuits are amplifiers with a high percentage of negative feedback. Cathode followers have a high input impedance and low output impedance and are useful as impedance matching devices.

A push-pull amplifier is a combination of two Class A or Class B amplifiers so connected that one amplifier amplifies the positive half and the other amplifies the negative half of the input signal and the amplified halves are then combined in the output transformer to produce an amplified replica of the input voltage. Push-pull amplifiers give higher output with less distortion than a single ended amplifier. In transistors, push-pull amplifiers having complementary symmetry with one PNP and the other NPN transistor are used which do not need any transformers.

Review Questions

1. What is an amplifier?
   What is the difference between a voltage amplifier and a power amplifier?
2. What is an audio amplifier?
   Why is it necessary to convert sound waves into electrical frequencies before amplification?
3. Describe the operation of a transistor as an amplifier. What is the phase relationship between the input waveform and the output waveform?
4. What is the difference between a Class A and a Class B audio amplifier?
   How can Class B amplifiers be used as audio amplifiers?
5. What is a push-pull Class B amplifier?
   Describe the circuit of a Class B transistor push-pull amplifier (a) with transformers, and (b) without transformers.
6. Explain three types of distortion in audio amplifiers.
7. What is crossover distortion in push-pull amplifiers and how is it overcome?
8. What is negative feedback in audio amplifiers?
   Show that, with a large negative feedback, the gain of the amplifier becomes independent of the actual gain without feedback.
9. What is a cathode follower? Describe the use of a cathode follower as a matching device.
10. What is meant by the frequency response of an audio amplifier? How is the frequency response measured and plotted as a graph?
11. What is an RF amplifier? How does an RF amplifier differ from an audio amplifier?
12. Describe the use of tuned circuits in RF amplifiers.
13. Define the bandwidth of an RF amplifier. How does the bandwidth depend on the Q of the tuned circuit?

chapter 10

Oscillators

10.1 WHAT IS AN OSCILLATOR?

A body that performs a back and forth motion in a regular and uniform manner is called an oscillator. The pendulum of a clock is a good example of a mechanical oscillator. Once the pendulum is started from its mean position at *A*, it swings to an extreme position *B* at the left, comes back to its mean position at *A*, swings to another extreme position at *C* and again returns to the mean position at *A* (Fig. 10.1). The pendulum is said to have performed one complete cycle of oscillation. The number of such complete cycles of oscillations made in one second is the frequency of oscillations and the maximum distance the pendulum swings to one side from its mean position *A* is the amplitude of the oscillations. If the energy lost due to air resistance is made up by the main spring of the clock, the pendulum will keep on swinging indefinitely with the same frequency and amplitude and thereby regulate the time of the clock. If, however, no energy is supplied to the pendulum from the clock, the pendulum will come to rest after performing a few oscillations. The motion of the pendulum is a periodic motion and can be represented by a sine wave as in Fig. 10.1 (b).

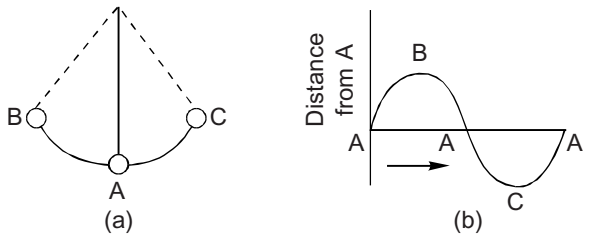


Fig. 10.1 Pendulum as an Oscillator, (a) Pendulum as a Mechanical Oscillator, (b) Periodic Motion of the Pendulum Represented by a Sine Wave

In electronic oscillators used in radio and TV, the electrons in a circuit are made to perform an oscillatory motion similar to the motion of the pendulum and such electronic devices can generate AC signals of any frequency starting from a low frequency of a few hertz to very high frequency of several megahertz. The frequency of the generated signal depends on the circuit constants.

To understand the principle of an electronic oscillator consider the arrangement in Fig. 10.2 (a) where the capacitor C can be charged to the battery voltage E by connecting the switch S to P_1. When the capacitor discharges through L, Fig. 10.2(b), the current flowing through the coil builds an expanding magnetic field around it. After C has completely discharged, the magnetic

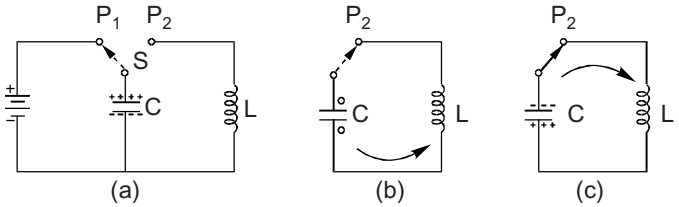


Fig. 10.2 Principle of an Electronic Oscillator

field collapses and as per Faraday's laws, induces an emf in the coil L which tries to maintain the flow of current through L in the original direction. This electron flow charges the capacitor to opposite polarity as in Fig. 10.3(c). Capacitor C again tries to discharge itself through L but the electron flow is now in the opposite direction and another magnetic field in the opposite direction is built around L. This back and forth motion of electrons in the circuit constitutes oscillations and the process continues till all the energy given to the circuit by the battery during the initial charging of the capacitor is dissipated as heat due to I^2R losses in the resistance of the circuit. The energy which is alternately stored in L and C diminishes gradually after each oscillation, making the amplitude of oscillations smaller and smaller till the oscillations completely die down as shown in Fig. 10.3(b). Such oscillations are known as *damped oscillations*.

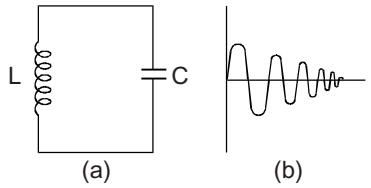


Fig. 10.3 Damped Oscillations
(a) LC Circuit $f = \dfrac{1}{2\pi\sqrt{LC}}$
(b) Damped Oscillations

The period of oscillations, however, remains the same. The frequency f at which the LC circuit (also known as the tank circuit) oscillates is given by the formula:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where f is in hertz and L in henrys and C in farads.

In order to produce sustained continuous oscillations with constant amplitude, it is necessary that the energy lost as heat during oscillations must be

made up or restored to the circuit by external means. This can be done by charging the capacitor at regular intervals by means of the switch *S*, but a better method is to connect the tuned *LC* circuit to the input of an amplifier as in Fig. 10.4. A portion of the amplified voltage is fed back inductively to the *LC* tank circuit to maintain the oscillations.

10.2 TYPES OF OSCILLATORS

10.2.1 Armstrong Oscillator or Tickler Coil Oscillator

In the *Armstrong oscillator*, a coil L_1 called the tickler coil, is introduced in the collector circuit of the transistor as shown in Fig. 10.4. The tickler coil is inductively coupled to the tank coil L_2 (usually both coils are wound on the same coil (form) so that when the switch *S* is closed, a surge of collector current flowing through the coil L_1 induces a voltage in L_2 and the tank circuit L_2C_1 is shock-excited to generate oscillations at a frequency determined by the constants of the tank circuit. These oscillations are applied to the input of the transistor amplifier and the amplified voltage through L_1 produces further feedback in the correct phase to the tank circuit to sustain the oscillations produced. This process continues till conditions stabilise to produce sine waves at the desired frequency. If the capacitor C_2 is made variable, oscillations at different frequencies can be generated by the oscillator circuit. L_2C_2 are the main frequency-determining components. R_2C_1 combination provides enough self-bias to drive the transistor to cut off so that the transistor operates as a class C amplifier. Current pulses are provided to the tank circuit during positive peaks of oscillations through mutual coupling between L_1 and L_2 to sustain the oscillations.

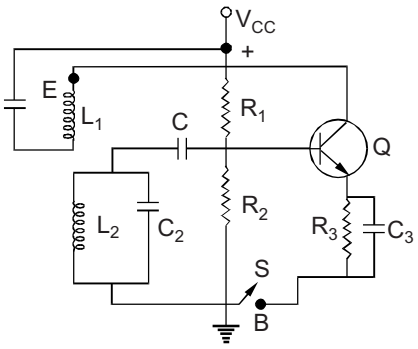


Fig. 10.4 Armstrong Oscillator or Tickler Oscillator

10.2.2 Hartley Oscillator

The *Hartley* oscillator is only a modification of the Armstrong or tickler-coil oscillator. In the Hartley oscillator, no separate tickler coil is used but the tickler coil is formed out of a part of the coil *L* in the *LC* tank circuit as in Fig. 10.5.

The single tank circuit coil is tapped at a suitable point so that the emitter current flowing through L_1 induces a kick voltage in L_2 to provide the starting feedback voltage. The

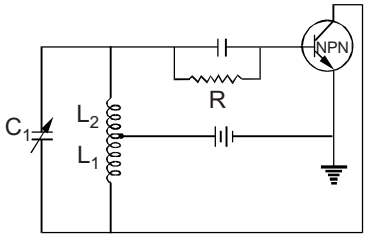


Fig. 10.5 Transistor Hartley Oscillator

amount of feedback voltage can be adjusted by moving the tap. Once the oscillations start, the cycle is very much the same as in the Armstrong oscillator. The Hartley oscillator is widely used in radio and TV circuits because it is easy to tune and is adaptable to a wide range of frequencies.

10.2.3 Colpitt's Oscillator

The Colpitt's oscillator is similar to the Hartley oscillator except that the tapped coil is replaced by two variable capacitors C_1 and C_2 as shown in Fig. 10.6. These two variable capacitors form a simple AC voltage divider. The amount of feedback depends on the ratio of C_1 and C_2. Capacitor C_b feeds back the RF voltage in the proper phase to the lower end of the tank; the amount of feedback is adjustable by the capacitor C_2. The radio frequency choke (RFC) prevents the RF currents from flowing through the collector supply E_B. The frequency of oscillations is determined by the tuned circuit formed by the coil L and the capacitance of the two capacitors C_1 and C_2 in series as in Fig. 10.6.

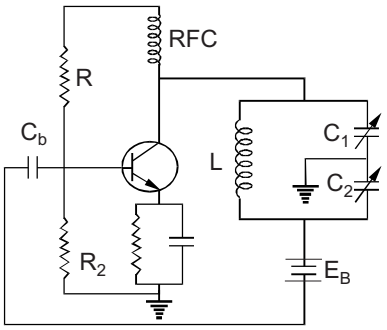


Fig. 10.6 Colpitt's Oscillator

Colpitt's oscillator is simple to operate and can be adapted to a wide range of frequencies. It has better frequency stability than the Hartley Oscillator.

10.2.4 Crystal Oscillator

A *crystal controlled oscillator*, a crystal oscillator as it is called, makes use of a property, known as piezoelectricity, possessed by certain crystals like quartz, tourmaline and Rochelle salt. When subjected to mechanical pressure, the piezoelectric crystals develop a potential difference between their faces, the polarity of the charges developed depending on whether the pressure compresses or stretches the crystal. Conversely, if an electric potential is applied to the faces of the crystal, it will expand or get compressed depending on the polarity of the applied electric potential. If the applied electric potential is a sinusoidal AC voltage, the crystal will also start oscillating sinusoidaly and develop sinusoidal voltages in its faces due to the piezoelectric effect. The voltages so developed can be applied to the base of a transistor to control the amount of feedback.

A crystal with its natural frequency of vibrations resembles a series resonant circuit as shown in Fig. 10.7(a). However, when used in an oscillatory circuit, the crystal slice is generally enclosed between two conducting plates constituting a crystal holder. With the capacitance of the conducting plates the crystal behaves like a parallel resonant circuit as shown in Fig. 10.7(b).

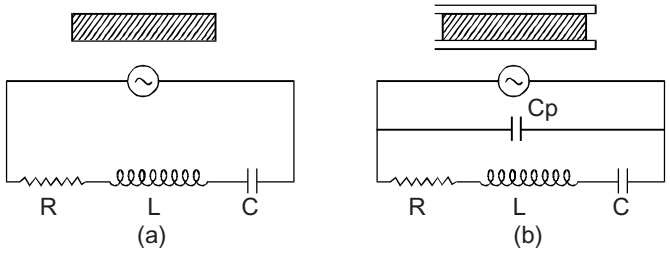


Fig. 10.7 Crystal and its Equivalent Circuits (a) Crystal as a Series Resonant Circuit, (b) Crystal with Conducting Plates Equivalent to a Parallel Resonant Circuit

The crystal, as a mechanical circuit resonator, can replace the *LC* resonant circuit in a normal Hartley or Colpitt's oscillator as shown in Fig. 10.8. This circuit is known as the Pierce crystal oscillator. A portion of the energy from the output is being fed back to excite the crystal.

A crystal oscillator has a sharply resonant circuit with a high *Q* and hence it oscillates at one fixed frequency which is extremely stable. Crystal controlled oscillators are mostly used in broadcast transmitters and other communication equipments where very close frequency tolerances have to be observed. This property of fixed frequency oscillations becomes a disadvantage where the frequency changes are frequent and a different crystal has to be used for every desired frequency. A variable frequency oscillator is found more useful in such applications. A crystal oscillator also suffers from another limitation, that of low output power. Hence a crystal oscillator is most useful where frequency stability and not the power output is the prime consideration.

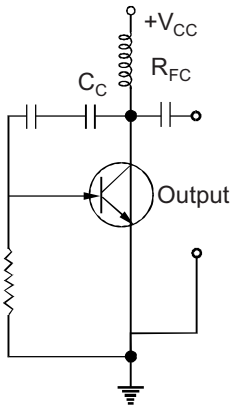


Fig. 10.8 Circuit of a Crystal Oscillator

10.2.5 RC-Oscillators

The LC oscillators discussed so far produce only high frequencies. For the production of low frequencies (say audio frequencies) we make use of RC oscillators. There are two methods to provide positive feedback required for converting an amplifier into an oscillator. The two methods are (1) Phase-shift oscillator method and the (ii) Wien Bridge oscillator method.

1. Phase Shift Oscillator Method A single stage amplifier not only amplifies the input signal but it also shifts the phase of the signal through 180°. If a part of this output signal is shifted through another 180° by an RC network, the total phase shift amounts to 180° + 180° = 360° which is equivalent to a phase shift of 0°. This is a positive feedback and the amplifier builds up oscillations.

Figure 10.9 shows the circuit of a phase shift oscillator in which the output of the transistor amplifier is fed back to the input of the amplifier through a phase shift network consisting of three identical RC networks each providing a phase shift of 60°, the total phase shift being 60 × 3 = 180°. The output of this network is now in the same phase as the input to the amplifier. If the condition

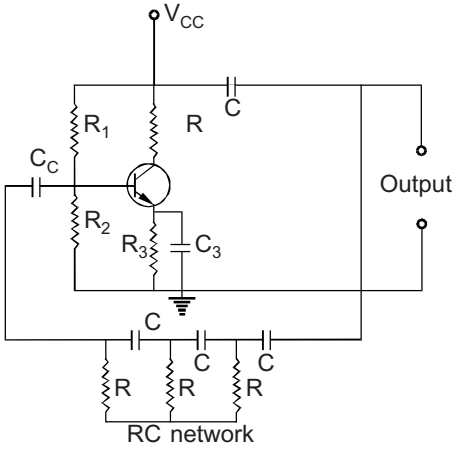


Fig. 10.9 Phase-shift Oscillator

$A\beta = 1$ is satisfied the oscillation will be maintained. It can also be proved that the frequency of oscillations in this type of oscillator is given by

$$f_0 = \frac{1}{2\pi RC\sqrt{6}}$$

Further it can be shown that the feedback factor of the *RC* network is given by

$$\beta = \frac{1}{29}$$

The above equation also signifies that for self-starting the oscillations $A\beta$ must be greater than 1 ($A\beta > 1$). This implies that for the oscillations to start the gain of the amplifier must be greater than 29.

By changing the values of *R* or *C* simultaneously in each section, the frequency of the oscillations can be changed from 20 Hz to as high as 300 kHz.

2. Wien Bridge Oscillator Another method of getting a phase-shift of 360° is to use two amplifier stages each giving a phase shift of 180°. A part of

this output is fed back to the input through feedback network which does not produce any further feedback. This method is used in the Wien Bridge oscillator which can produce low frequencies in the range of 10 Hz to 1MHz. A block diagram of the arrangement is shown in Fig. 10.10.

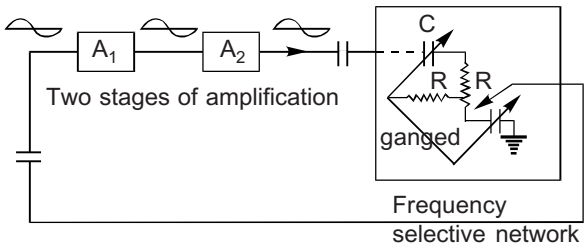


Fig. 10.10 Block Diagram of a Wien Bridge Oscillator

Variable frequency operation requires the simultaneous operation of two capacitors C_1 and C_2 which are shown as ganged in the frequency selective network.

To improve the stability of the oscillator a certain amount of negative feedback is also provided. The circuit can easily be arranged in the form of a bridge (Fig. 10.11) and hence the name Wien bridge. Resistances R_1 and R_2 provide the desired negative feedback. A single block A represents the two stages of amplification A_1 and A_2.

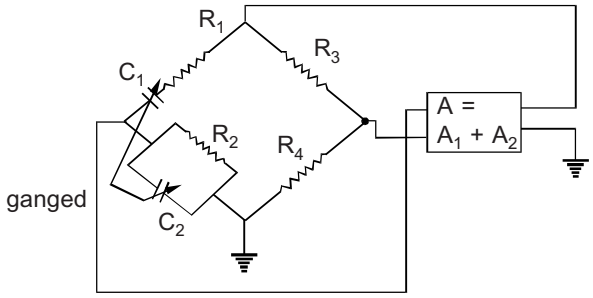


Fig. 10.11 Wien Bridge Oscillator

The ganged capacitors C_1 and C_2 are to change the frequency. The frequency range of the oscillator can also be changed by using the different values of R_1 and R_2.

10.3 FREQUENCY MULTIPLIERS

A frequency multiplier is an RF amplifier which receives as its input any basic sinusoidal frequency from an oscillator, like a crystal oscillator, but delivers at the output, a frequency which is a multiple of the input frequency. A frequency multiplier is a doubler, tripler, quadrupler and so on, depending upon whether the output frequency is double, three times or four times respectively of the

input frequency. Thus, if the input frequency is 1 MHz, a doubler will deliver 2 MHz, a tripler 3 MHz and a quadrupler 4 MHz as the output frequency.

Frequency multiplier stages are often used in AM transmitters operating in the VHF range because it is difficult to design a stable oscillator that will operate at such high frequencies. Frequency multipliers are often used in FM transmitters which also operate in the VHF range.

A frequency multiplier is normally a Class C, RF amplifier which has a parallel tuned circuit in the input and a tuned tank circuit in the output. Plate current flows only during the positive peaks of the input signal, but the flywheel action of the tank circuit produces sinusoidal output signals. The plate current pulses for the fundamental or basic frequency reinforce the tank circuit oscillators at the peak of every alternate sine wave in a doubler, and in a tripler, these pulses will aid the peak of every third sine wave and so on. The frequency of the output signal is determined by the resonant frequency of the tank circuit. A switch *S*, shown in Fig. 10.12, enables the particular tank coil for the desired multiple or harmonic to be selected and the variable capacitor C_2 allows final tuning to be achieved for the output frequency. The output can be coupled to any succeeding stage either by inductive coupling or by RC coupling. Push-pull amplifiers which can produce even order harmonics, are used as doublers and quadruplers. Whenever greater power output and higher collector efficiencies are required that can be made available by a single stage.

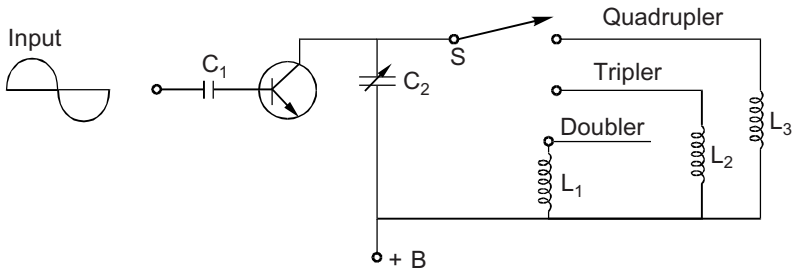


Fig. 10.12 A Frequency Multiplier Stage

Multivibrators A multivibrator is also a relaxation oscillator like the blocking oscillator but without the use of a transformer for providing the feedback. A multivibrator consist of two RC coupled amplifier stages in which the output of one stage is applied as input to the other stage, as shown in Fig. 10.13. Since each amplifier stage produces a 180° phase-shift, the signal applied to the input of each stage is in phase with the original input and this helps to produce oscillations.

Multivibrators are useful devices for generating square waves and pulses where the frequency can be controlled by voltages injected from outside. The waves generated by multivibrators are also very rich in harmonics.

Multivibrators are classified either according to the system of feedback employed or according to the stability conditions obtained in its operation.

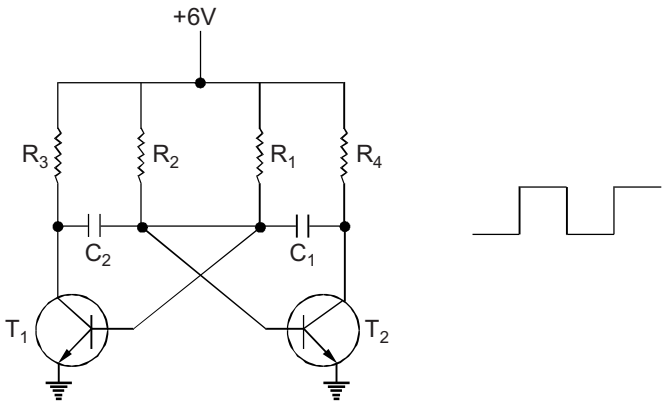


Fig. 10.13 Multivibrator

When transistors are used in the circuits of multivibrators, the types are called the collector-coupled multivibrator and emitter-coupled multivibrator.

The classification of multivibrators according to stability conditions includes:

Free-running or Astable Multivibrators When neither of the stages is stable but remains alternately cut off and conducting at the MV repetition rate, the MV is a free running or astable type. The frequency of an astable multivibrator can be controlled by external synchronising pulses. The multivibrator tends to adjust the frequency of its oscillations so that the ratio of the external synchronising frequency and the multivibrator frequency is an integral ratio which can be either unity or less or greater than unity.

Monostable or "One-shot" Multivibrator This type of multivibrator has only one stable state. An external pulse is required to initiate action when the multivibrator goes through one cycle of operation and returns to its original stable state, when the external pulse is no longer present.

Bistable Multivibrator or Flip-flop Circuit This type of circuit which has two stable conditions is also known as the Eceles-Jordan trigger. This is a direct coupled two-stage amplifier in which the output of the second stage is connected to the input of the first stage. When an external pulse is applied to the base of the non-conducting stage, the system jumps almost instantly from one stable state to the other resulting in a trigger or flip-flop action.

Both the monostable and bistable multivibrators are driven oscillators or trigger circuits which need an external driving pulse to start the operation and maintain it.

SUMMARY

A body, like a clock pendulum, which performs a back and forth motion in a regular and uniform manner is called an oscillator. In an electronic oscillator, the electrons in a circuit are made to perform oscillatory motions similar to that of the pendulum thereby generating sine waves of any frequency depending on

the constants of the circuit. Examples of such sine wave oscillators used in radio and TV circuits are the Armstrong oscillator, Hartley oscillator, the Colpitt's oscillator, and the crystal oscillator. Besides the LC oscillators which are mainly used for the production of high frequency, RC oscillators are used for audio frequencies. These include the phase shift oscillator and the Wien Bridge Oscillator.

Multivibrators are special types of oscillators useful for generating square waves and pulses and are very rich in harmonics. Multivibrators when classified according to stability conditions are called Free running or astable multivibrators, monostable or 'one shot' multivibrators and Bistable multivibrators or flip-flop circuits.

Review Questions

1. What is an oscillator? Describe the principle of operation of an electron oscillator.
2. Draw the circuit of a simple Hartley oscillator and explain its working and also mention its common applications.
3. What is an RC oscillator? Draw the circuit of a Wien Bridge Oscillator and explain its working.
4. Explain the principle of a crystal oscillator and mention its practical applications.
5. What is a multivibrator? Describe the principle of a free running or a stable multivibrator.
6. State whether True or False with brief reasoning.
    1. An oscillators is an amplifier with negative feedback.
    2. The frequency of an oscillator increases with value of the capacitor.
    3. Wien Bridge Oscillator is basically an audio oscillator.
    4. A crystal oscillator is a series reasonant circuit.
    5. A frequency of 20 MHz will need an LC oscillator.

    (*Ans:* 1. False, 2. False, 3. True, 4. True, 5. True)

chapter 11

Power Supplies

11.1 POWER SUPPLY SOURCES

The most common source of power supply these days is the AC mains, because it is more efficient and economical to generate and transmit AC power. However, DC voltages/currents are required for the satisfactory operation of a wide variety of electronic equipment using vacuum tubes and solid state devices, such as transistors and ICs. Vacuum tubes require DC voltages for all electrodes except the filaments, which may be AC or DC operated. It, therefore, becomes necessary to convert AC into DC by means of a device called the *rectifier*. The pulsating DC produced by the rectifier is changed into smooth and steady DC by filter elements. A power supply system for electronic equipment will consist of a transformer, a rectifier, filter elements, voltage dividers and voltage regulators for maintaining the DC output voltage at a relatively constant value under varying conditions.

11.2 BASIC REQUIREMENTS OF A POWER SUPPLY SYSTEM

Figure 11.1 gives a schematic of the basic requirements of a power supply system. These are discussed below.

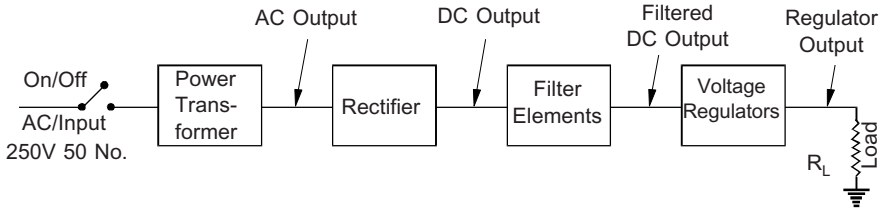


Fig. 11.1 Basic Functions of a Power Supply System

11.2.1 Power Transformer

A step-up or step-down transformer of the required rating is employed as discussed in Part I. Chapter 4 Section 4.25.1.

11.2.2 Rectifiers

A rectifier converts AC into pulsating DC by eliminating the negative half cycle of the AC voltage. A rectifier has to be a device with unidirectional characteristics. An ideal rectifier actually works like a switch that closes with zero resistance on the positive half cycle of the AC voltage and opens with infinite resistance on the negative half cycle. A practical rectifier, however, has a very low resistance during the conducting interval and a very high resistance during the non-conducting interval. A diode possesses unidirectional characteristics of this type and can work as a rectifier irrespective of its type. Various types of diodes used as rectifiers are vaccum diodes*, semiconductor diodes or metal rectifiers such as copper-oxide and selenium rectifiers. Semiconductor diodes and metal rectifiers have become very popular in radio and TV equipment because they require no filament heating. Some of the common methods of rectification will now be discussed.

11.3 HALF-WAVE RECTIFICATION

The process whereby a rectifier conducts during only one alternation of the input cycle is called *half wave rectification*. A single diode which allows current to flow only during positive alternations of the AC voltage will act as a half wave rectifier. The circuit of a half wave rectifier is shown in Fig. 11.2.

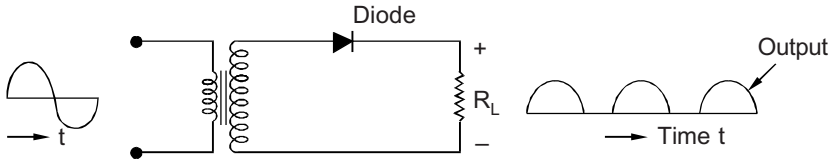


Fig. 11.2 Half-wave Rectification: Semiconductor Diode as a Half-wave Rectifier

For a 50 Hz AC supply the polarity at the input changes every hundredth second. Current flows during the positive alternation when the anode is positive with respect to the cathode. During the negative alternation there is no current because the anode is negative with respect to the cathode. The plate current flowing through the series connected load resistor R_L will develop a voltage across it making the cathode-connected end of R_L positive. The voltage developed across the load R_L is a pulsating DC which has the same waveform as the positive alternations of the input voltage as indicated in Fig. 11.2. The pulsating DC produced in this way has to be filtered by suitable smoothing circuits which are described later.

*Use of vaccum diodes has been retained in view of Hybrid (B/W) TV receivers still in use in many cases.

In half-wave rectification, half the input voltage is lost and the efficiency of the system is low. Moreover, the half wave rectifier system requires elaborate filtering arrangements to remove the low frequency ripple for obtaining a smooth DC.

During the negative half cycle of the input AC voltage, when the diode is not conducting, the actual voltage applied across the diode junction is the sum of the negative AC input voltage at the anode and the DC output voltage at the cathode, because the two voltages are in series aiding. The maximum voltage that can be applied to a diode in the reverse direction without producing a breakdown of the diode is called the *peak inverse voltage* (PIV) and is generally specified by the manufacturers. This is about twice the DC output voltage.

11.4 FULL-WAVE RECTIFICATION

In full-wave rectification, two diodes are connected in such a manner that the plate current flows through the load during the full cycle of the AC supply voltage. The two diodes alternately supply rectified current to the load in the same direction during both halves of the input. AC voltage, thereby filling up the gaps appearing in the waveform of the half wave rectification. The circuit of a full-wave rectifier is shown in Fig. 11.3.

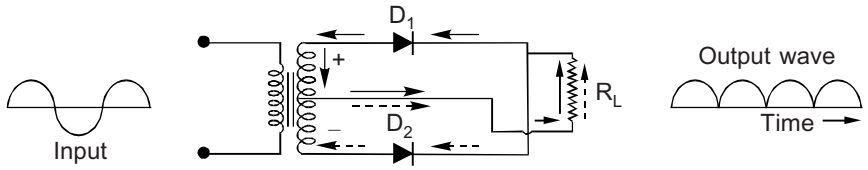


Fig. 11.3 Full-wave Rectification Full-wave Rectifier Using Semiconductor Diodes

In this circuit the cathodes of the two rectifier diodes are connected together and the common junction is connected to one end of the load resistor R_L, the other end of which is connected to the centre tap in the secondary of the transformer. Each diode is connected between one end of the transformer and the centre with the result that only half the transformer secondary voltage is applied between the anode and cathode of each diode. Thus, the secondary winding of the power transformer should be able to supply an end-to-end voltage that is twice the voltage to be rectified.

As regards the actual functioning of the full-wave rectifier, the diodes D_1 and D_2 conduct during successive half cycles of the input AC voltage, each permitting current flow during one half cycle when the anode is positive with respect to the cathode. The direction of electron current flow which is the same through R_L for each diode has been shown in Fig. 11.3. The frequency of DC pulses known as *ripple frequency* is twice the frequency of the AC supply. This makes it easier to design a filter circuit for smoothing than in the case of half-wave rectification. Since both halves of the AC input cycle are rectified, the efficiency of the full-wave rectification is higher than that of half-wave rectification.

Another advantage of full-wave rectification is that there is no DC saturation of the core of the transformer because the DC pulses from the two diodes flow in opposite directions through the secondary of the transformer and thereby cancel each other out. This reduces the size and consequently the cost of the transformer required for full-wave rectification. It may thus be seen that full-wave rectification is a more suitable type of rectification as compared to half-wave rectification and is therefore used as a standard circuits for a wide variety of low power applications in electronics.

11.5 FULL-WAVE BRIDGE RECTIFICATION

The centre-tapped transformer required for the full-wave rectification described earlier can be eliminated by use of four diode rectifiers in a bridge-type rectifier as shown in Fig. 11.4. The secondary of the transformer is connected to diagonally opposite corners *B* and *D* of the bridge, whereas the load resistor R_L is connected to the other two corners. The operation of the bridge rectifier is described below.

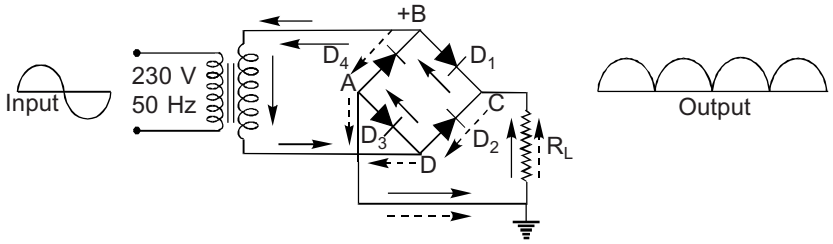


Fig. 11.4 Full-wave Bridge Rectification

During the positive half-cycle of the input AC voltage when *B* is positive with respect to *D*, the diodes D_1 and D_3 are forward-biased and conducting, whereas the diodes D_2 and D_4 are reverse-biased and non-conducting. The electronic current flows through R_L, D_1, the transformer secondary winding and through D_3 as shown by full-line arrows. During the negative half of the AC cycle, however, *D* is positive with respect to *B* and the diodes D_2 and D_4 are forward-biased, whereas D_1 and D_3 are reverse-biased. D_2 and D_4 conduct during the negative half-cycle allowing current to flow through the load resistance R_L in the same direction as shown by dotted arrows. Thus, diodes D_1 and D_3 rectifying in series provide the positive pulses during the positive half-cycle of the input and diodes D_2 and D_4 in series with each other provide the positive pulses during the negative half cycle of the input AC. The bridge rectifier, therefore, acts as a full-wave rectifier with the output waveform as shown in Fig. 11.4.

A bridge rectifier has the advantage that it produces a voltage output that is twice the output voltage obtained from a conventional full-wave rectifier using the same power transformer. This is because in the bridge circuit the full voltage of the secondary is applied to the two conducting diodes in each half

cycle compared to the conventional full-wave rectifier in which the secondary voltage is divided into two halves. The extra cost of using four rectifier diodes in the bridge circuit is offset by the reduction in the cost of the transformer because the entire current in each half cycle flows through the complete winding in opposite directions.

11.6 VOLTAGE DOUBLERS

Voltage doublers are rectifier circuits which are capable of delivering a DC voltage that is twice the peak value of the applied AC input voltage. Voltage-multiplying circuits can also be made to deliver DC output voltages that are several times the peak input AC voltage. Such circuits are particularly useful in equipments where a transformer with a sufficiently high secondary voltage will be expensive and inconvenient.

Figure 11.5 shows the circuit of a full-wave voltage doubler. In this circuit the rectified voltages of two half-wave rectifiers D_1 and D_2 are combined in series so as to produce an output voltage that is twice the peak value of the input AC voltage.

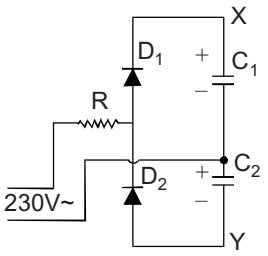


Fig. 11.5 Voltage Doublers Semiconductor

During the positive half cycle of the input AC voltage the anode of D_1 is positive with respect to its cathode and D_1 conducts and charges the capacitor C_1 to the peak input voltage with the polarity as shown. D_2 does not conduct during this half cycle.

During the negative half cycle, D_2 conducts while D_1 does not. The capacitor C_2 is charged to the peak value of the AC input voltage with the polarity shown. The voltages across C_1 and C_2 are in series-aiding and add up to give an output voltage across XY that is twice the peak AC input voltage.

Figure 11.5 uses semiconductor diodes without any power transformer. Such transformerless voltage doublers are quite common in television receivers.

11.7 FILTER CIRCUITS

Various types of rectifier circuits described so far convert AC into a pulsating DC only. The variations in the amplitude of the output voltage must be smoothed out to make the rectified DC voltage suitable for satisfactory operation of vacuum tubes, transistors and other electronic devices. The variations in the output voltage, known as *ripple*, can be eliminated by the smoothing action of certain circuits called *filter circuits* which are networks consisting of capacitors, chokes and resistors.

The filter action of a capacitor depends on its property to oppose any changes in the voltage applied across its terminals by storing up energy whenever the voltage tends to rise and releasing this stored energy back as a voltage or as a

current whenever its voltage tends to fall. A choke coil, on the other hand, stores energy in the form of a magnetic field whenever the current flowing through it increases and restores the same energy to maintain a steady flow of current when the current flow through the coil tends to decrease. A choke coil offers a high impedance to the ripple current while offering only a low resistance of its winding to the flow of DC.

Filter circuits make use of the voltage stabilising action of a shunt capacitor and the current smoothing action of a series choke. Depending upon whether the first component in a filter circuit is a capacitor or choke, the filter network is known as the *capacitor-input filter or choke-input filter*.

11.7.1 Capacitor-input Filter

The capacitor-input filter circuit commonly used in power supply systems is shown in Fig. 11.6. This is known as a π-type filter because its configuration resembles the Greek letter π.

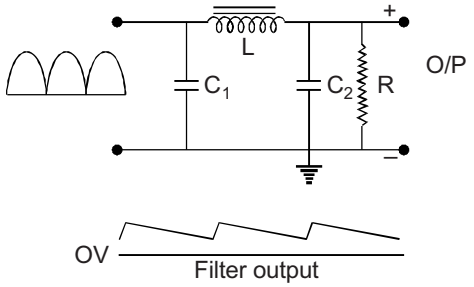


Fig. 11.6 Capacitor-input Filter Circuit

The input capacitor C_1 first gets charged to the peak value of the pulsations from the rectifier output voltage. The capacitor tends to retain the charge between successive pulses but discharges slowly through the choke L and load resistor R. As a result, the output voltage tends to fall off between successive pulsations though remaining substantially near the peak value as indicated by the filter output waveform shown in Fig. 11.6. Any remaining fluctuations in the rectifier output current are impeded by the series choke L and bypassed to the ground by the output capacitor C_2. Additional filter elements or higher values of filter capacitors are necessary for further reduction of the AC ripple. Since large capacitors are needed, electrolytic capacitors are invariably used on these filter circuits. Filter choke values vary from 10 to 15H.

In many cases the choke L is replaced by a resistor R_C as filter element in the CRC type of filter circuits shown in Fig. 11.7.

A resistor is not as effective a filter element as a choke. Therefore, the use of a resistor in

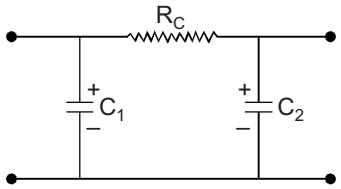


Fig. 11.7 A *m*-filter Circuit Using a Resistor as the Filter Element

place of a choke due to cost considerations or otherwise, will need large-valued capacitors to compensate for the loss of the choke. CRC filters are used in solid-state devices requiring low voltages. For CRC filtered supplies, capacitors with as high values as 500 to 1000 μF are commonly used. For vaccum tube devices requiring much higher voltages than the solid state devices, power supplies of the CLC type are commonly used in which the filter capacitors have a value of about 50 to 100 μF.

A characteristic of the capacitor-input circuit is that it provides maximum voltage output to the load. However, the output voltage falls off rapidly with increasing load current and the voltage regulation of this type of filter circuit is not very satifactory.

11.7.2 Choke-input Filter

In a *choke input filter* circuit, a choke coil is connected in series with the rectifier output as shown in Fig. 11.8.

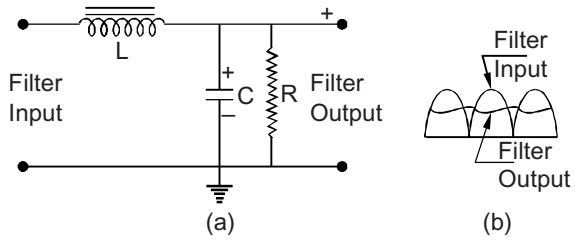


Fig. 11.8 Choke-input Filter Circuit Showing the Input and Output Waveforms

The input series choke *L* offers a high impedance to the pulsations from the rectifier output current but allows the DC current to flow through easily. Any fluctuations that are still present in the current after passing through the choke are bypassed around the load *R* to ground by the shunt capacitor *C*. However, a negligible amount is still present in the output waveform of the filter as shown in Fig. 11.8. It will be seen from the output waveform of the filter circuit that in this type of filter the DC output voltage is not equal to the peak value of the input pulsations as the choke coil does not allow the capacitor to charge to the peak value when the load current is drawn. However, as soon as a small load current is drawn, the voltage drops off to some lower value beyond which the output voltage changes very little with the changes in the load current. Thus a choke-input filter circuit provides a better regulation than a capacity input filter.

11.8 POWER SUPPLY REGULATION

Regulation of a power supply is the variation of the output voltage with changes in the load current. Regulation can be improved by connecting a resistor *R* across the output of the filter circuit. This resistor *R* is known as a *bleeder resistor*. The main purpose of the bleeder is to place a minimum load on the rectifier under all load conditions and thereby improve the regulation of the

power supply. Besides improving regulation, the bleeder resistor also helps in rapid discharge of filter capacitors when the power supply is switched off and this prevents any shock hazards. For good regulation, the bleeder current should be about 15 per cent of the total current.

11.9 POWER SUPPLY SYSTEMS

A complete power supply system should be able to provide all the required voltages for the satisfactory operation of electronic equipment using vacuum tubes, transistors and ICs. These voltages can be DC or AC voltages, high or low voltages and may be supplied by components suitably rated to meet the current load requirements of the particular circuits for which these are intended. Thus, in an equipment using vaccum tubes the power supply system should be able to furnish the *A*, *B* and *C* supply voltages required for lighting the filaments, operating the plate circuit of the tubes and for producing the DC grid bias voltages respectively.

A set of batteries as used in the case of portable or emergency electronic equipment seems to be the easiest solution for the problem of providing a power supply arrangement to produce the *A*, *B* and *C* supply voltages. This type of arrangement is, however, very cumbersome and uneconomical. A very convenient method of devising a power supply stage for any electronic equipment such as radio and TV receivers is by the use of supply mains easily available everywhere. These supply mains may be AC or DC supply mains. In certain areas, both AC and DC supply mains may be available. Accordingly, there are two main types of power supply systems commonly used in Radio and TV equipment. These two system are: the AC power supplies which are meant for use on AC mains only and the AC/DC, systems.

11.10 VOLTAGE REGULATORS

What is Voltage Regulation? The ability of a power supply system to maintain a constant output voltage in spite of changes in input voltage or output circuit conditions is defined as the *voltage regulation of the system*. Power supplies that have a low effective internal resistance are said to have good regulation. An ideal power supply system should have an effective internal resistance of zero value. In practice, however, the internal power supply resistance is considerably higher. It is dependent on such factors as the forward resistance of rectifiers, the transformer winding resistance, the size of the filter capacitors, whether the rectifier is a full wave or half-wave rectifier and the frequency of operation.

The need for voltage regulation is particularly important in the case of solid state transistor TV receivers which operate at relatively low dc voltages (4 to 60 V) but relatively high currents (4 to 1200 mA). The high current demands of the TV receivers are responsible for the need for special regulator circuits that keep the output voltage constant. Regulation also prevents damage to the

circuit transistors from line voltage fluctuations. Moreover, fast acting regulators can provide additional filtering for ripple reduction.

In voltage regulation the percentage of voltage regulation is given by

$$\% \text{ regulation} = \frac{\text{No load voltage} - \text{full load voltage}}{\text{Full load voltage}}$$

11.10.1 Types of Voltage Regulators

The following four types of voltage regulators are used in equipment using sold-state devices particularly the soild state TV receivers.

1. Zener-diode regulator
2. Voltage regulating power transformer
3. Feedback Regulators
4. Switch Mode Power Supply (SMPS)

A brief description of each of the above voltage regulators is given below.

11.10.2 Zener Diode Regulator

Zener diode as a voltage regulator has already been briefly discussed in Chapter 6 of Part 2.

It may be mentioned that with a reverse breakdown, the voltage across the Zener diode is constant for a wide range of current values through the diode. Zener regulators are commonly used as voltage regulators with ratings of 3 to 180 V.

The circuit in Fig. 11.9 is a Zener diode voltage regulator.

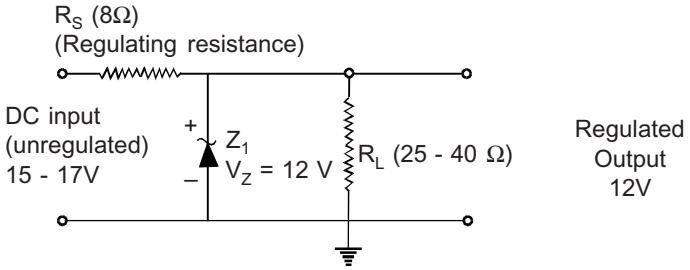


Fig. 11.9 Zener Diode Voltage Regulator

In the circuit shown, reverse bias is used with positive voltage at the cathode. The Zener diode is rated for 12 V breakdown voltage with a max-diode current of 150 mA and 10W power dissipation. The series regulating resistance R_S has the function of providing a voltage drop that varies with the load current. However, there must be at least 12 V across the diode to provide current through Z_1 so that it can operate in its breakdown mode.

11.10.3 Voltage Regulating Power Transformers

Constant Voltage transformers (CVT) are used in many circuits particularly in colour TV receivers to provide voltage regulation. These transformers can be

designed to regulate low voltage power supply for both line voltage and load variation. A CVT also helps to suppress transients pulses if these appear on the AC line.

The constant voltage tansformer is able to keep output voltage constant by making the secondary winding a resonant circuit. Moreover, the primary and secondary windings of a CVT are wound on a specially constructed core that allows the primary to operate normally while the secondary is operated in magnetic saturation. The secondary winding is forced into saturation by the large series resonant circuit current flowing through the winding. An external 3.5 μf oil filled capacitor tunes the secondary to 50 or 60 Hz as the case may be. In practice, TV receivers fitted with CVT regulators produce a good full size picture even when the input AC line voltage drops down to as little as half the normal line voltage.

11.10.4 Feedback Regulators

In this regulator, a sample of the output is fed back to a control stage which can then regulate the circuit for constant output. A block diagram. of the circuit is shown in Fig. 11.10.

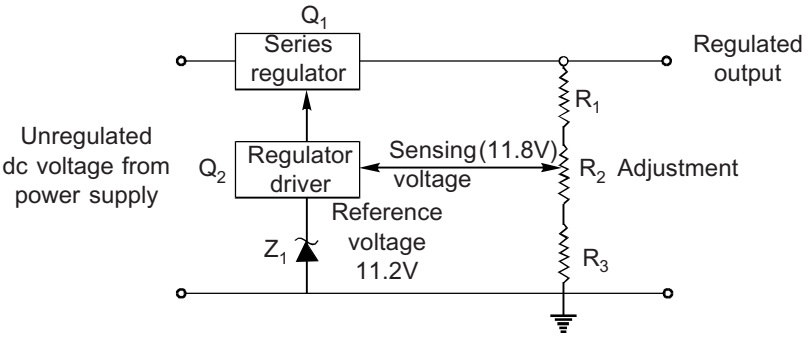


Fig. 11.10 Feedback Regulator with Series Regulator Transistor

Q_1 is the series regulator that provides output to the load. Q_2 is the driver to control the bias on Q_2. The driver has two input voltages, one is a reference voltage of 11.2 V stabilized by the zener diode in its emitter circuit and the other is the sensing voltage of approx. 11.8 V from a tap on the voltage divider R_1, R_2, R_3 across the output. The amount of conduction in the driver Q_2 determines the bias on the regulator Q_1. Moreover, the driver Q_2 compares the reference voltage of 11.2 V in the emitter to the sample voltage from R_2 which is approximately one-tenth the output applied to the base. Any difference is amplified to change the bias and conduction for the series regulator Q_1.

11.11 SWITCH MODE POWER SUPPLY (SMPS)

Operational voltages for solid-state circuits are more critical than those required for tube type circuits. In modern TV receivers and other similar solid-state

electronic circuits the use of ICs and modular construction is very common. To meet the stringent dc requirements of these circuits, a new power supply system has been introduced and is being used in all modern soild-state electronic circuits. This voltage regulating system develops a start up voltage for the horizontal oscillator. Once started the pulses from the horizontal oscillator are used to develop the required operational voltages for the rest of the receiver.

Basic Principle of SMPS A block diagram. of the SMPS circuit is given in Fig. 11.11.

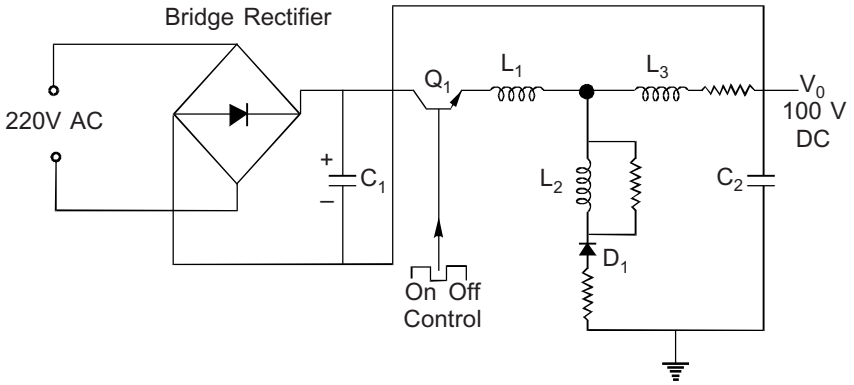


Fig. 11.11 Block Diagram of the SMPS System

The 220V AC input voltage is rectified by the bridge rectifier circuit and fed to the filter capacitor C_1 to charge to 300 V DC. This DC voltage (300 V) is connected through the switching transistor Q_1 to filter choke L_3. The switching transistor works as a switch which turns fully ON or OFF. During OFF period the energy stored across coil L_3 is delivered to condenser C_2 with rectifier diode D_1 in series. The diode allows the capacitor C_2 to charge due to induced voltage across coil L_3. The voltage across the capacitor C_2 is the output DC voltage connected to the load circuit. The unregulated dc supply is chopped by switching element at a rapid rate of more than 15 kHz (15625 Hz). The resultant high frequency pulse train is transformer coupled to an output network which provides final rectification and smoothing of the dc output. Regulation is accomplished by control circuits which vary the duty cycle-on off period-of the switching transistor and thereby control the output.

By making the switching frequency very high (more than 15 kHz) the values of the filter elements (capacitors and coils etc) are reduced to low values.

The rectifier diode D_1 must be a fast switching or fast recovery diode as it has to remove the negative induced voltage from the emitter of the switching transistor. The switching transistor is a fast switching medium power transistor such as 21413.

Advantages of SMPS SMPS possesses the following advantage over a normal conventional regulated Power supply system.

1. Due to the high switching rate of about 20 kHz the components like transformer, inductors and filter capacitors are much smaller and lighter than those required for line frequencies (60 Hz). This reduces both the weight and cost of the equipment.
2. Switching transistors are only on-off devices and dissipate much less power and efficiencies of the order of 65 to 85% can be obtained compared to 30 to 40% for the linear supplies.
3. It can operate under a low ac input voltage. It has a relatively long hold up period and will sustain even if the input power fails momentarily.

However, SMPS is very expensive and has a complex circuitry. Electromagnetic interference from the switching circuits necessitates special shielding of the circuits and equipment and filter circuits to reduce noise and interference.

11.12 INVERTERS

Emergency Power Supply Systems An inverter is an emergency power supply system, that supplies power to small household loads consisting of lights, fans, TV sets and other small electrical appliances in case of power supply failure. The emergency power supply comes from a set of battery accumulators which forms a part of the inverter system. Figure 11.12 shows a block diagram of an inverter system.

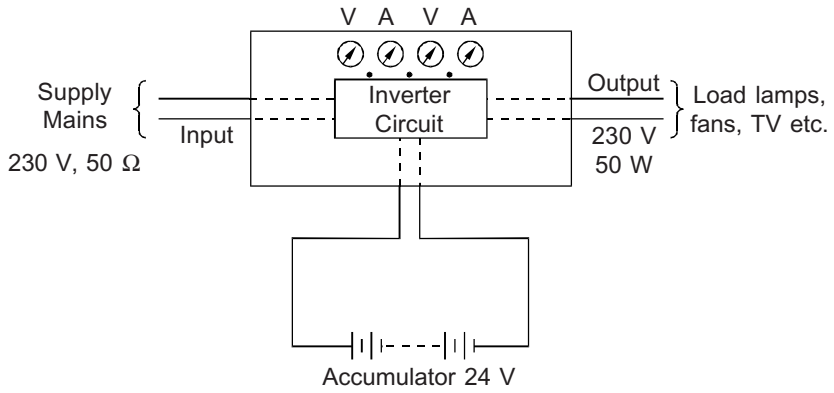


Fig. 11.12 Block Diagram of an Inverter System

Normally, the household load is fed from the mains supply while it is on. In this condition, the accumulators also get the charging current from a built in rectifier system. When the main supply fails, a relay system operates and the battery voltage gets connected to this inverter system which convertes the battery DC voltage into an AC supply of suitable voltage and frequency and this takes over the load for the duration of the power supply failure. As soon as the power supply is restored the relay operates again, the system falls back to its normal working. The charging of the batteries starts again and this continues till another power supply failure occurs.

Meters and indicators lamps fitted into the cabinet give the required information regarding the working of the inverter system.

Inverters of various capacities such as 200 VA, 500 VA and 1250 VA are available in the market.

11.13 POWER SUPPLY TROUBLES

Defects in most electronic equipments can be traced to power supply troubles. For troubleshooting, systematic measurement of AC and DC voltages at various points of the power supply circuit is necessary. The result of these measurements can give a clue to the actual trouble and help in isolating and replacing the defective component.

If no output DC voltage is available from the power supply, the trouble could be due to (i) open line cord or defective plug, (ii) blown-out fuse, (iii) defective power transformer, (iv) defective rectifier diode, (v) defective filter capacitors, or (vi) open filter choke or filter resistor.

If no DC output voltage is available across C_2 but DC voltage is available across C_1, the trouble is a shorted filter capacitor C_2, an open filter choke or resistor R_3 or a shorted bleeder resistor, if one is used. A resistance check of these components will indicate the trouble. If the voltage across C_1 is also zero, the trouble may be shorted C_1. Before isolating and checking C_1, an AC voltage test across the secondary of the power transformer is necessary to eliminate any possible defect in the line cord, fuse switch and power transformer. If there is no AC voltage across the power supply but AC voltage is available at the mains outlet socket, then a continuity check between the prongs of the power plug and transformer primary after closing the switch will show if the line cord, fuse, switch or the transformer primary is open and should be replaced. If all components from the line cord through the primary of the transformer are normal, the trouble is an open secondary winding in the transformer which must be confirmed by a continuity check. If the AC voltage is available across the secondary winding but no DC voltage is available either across C_1 or C_2, the trouble could be either a shorted filter capacitor C_1 or a defective rectifier diode which should be replaced. An open centre tap on the secondary of the transformer can also be the cause of the trouble.

If the DC output voltage is low, the trouble can be attributed to either an increase in the load current or leaky electrolytic capacitors. A weak or defective rectifier diode can also be the cause of this problem. The defective components including the filter capacitors should be replaced one at a time till the output DC voltage and hum level, if any, are restored to their normal value.

This procedure to detect faults applies equally to AC-DC or transformerless power supplies, except that the transformer troubles are eliminated in the latter.

SUMMARY

DC voltages required for the operation of electronic equipment using vacuum tubes and transistors are generally obtained by the conversion of AC mains

supply into DC by a process known as rectification. AC is converted into pulsating DC by means of a device called the rectifier. A filter circuit consisting of capacitors, chokes and resistors as filter elements changes the pulsating DC into smooth and steady DC.

A rectifier is a unidirectional device which conducts only during the positive half of the AC cycle and remains cut off during the negative alternation. Diodes posses unidirectional characteristics and are used as rectifiers. Diodes used as rectifiers may be either vacuum diodes or semiconductor diodes which require no filament heating.

Rectification may be half-wave rectification using a single diode, conducting only during positive alternation of the input cycle or full-wave rectification using two diodes so connected that the current flows through the load during the full cycle of the input AC voltage. A bridge rectifier using four diodes provides full-wave rectification without the use of a centre-tapped transformer. Voltage doublers are rectifier circuits capable of delivering a DC voltage that is twice the peak input AC voltage.

Filter circuits are used to convert pulsating DC into steady DC by smoothing out the pulsations. Practical filter circuits combine voltage stabilising action of shunt capacitors with the current smoothing action of a series choke coil. A filter circuit is a capacitor-input circuit or a choke input circuit depending upon whether the first component in the filter circuit is a shunt capacitor or a series choke coil. A bleeder resistor across the filter circuit helps maintain a constant output voltage for changing loads.

Voltage regulators of various types are used to prevent damage to transistors due to fluctuations in the line supply voltage. Switch Mode Power Supply System (SMPS) is now used in almost all modern solid state electronic circuits.

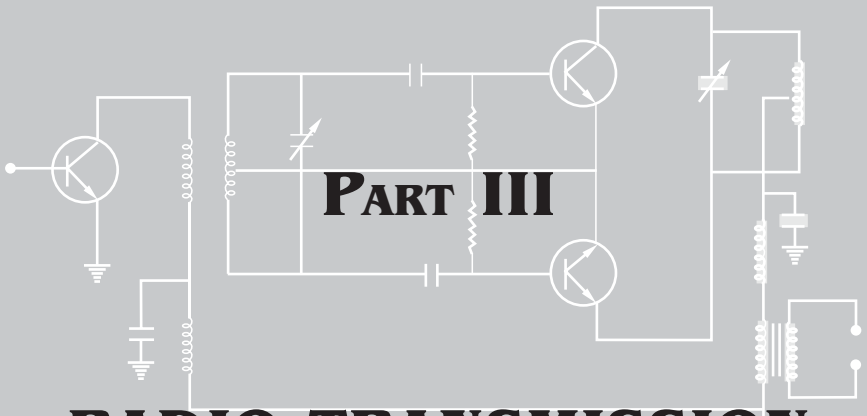
Inverters are emergency power supply system that supply power from accumulators in emergencies. Troubleshooting in power supply systems is carried out by systematic measurement of AC and DC voltages at various points and resistance checking of the suspected components.

Review Questions

1. Explain the function of power supplies in electronic equipment?
2. What is rectification? How the characteristics of a diode help in rectification?
3. Explain the operation of a half wave rectifier? How is the pulsating DC changed into steady DC?
4. What is full-wave rectification? Draw the circuit of a full-wave rectifier using semiconductor diodes.
5. Draw the circuit of a bridge rectifier showing the direction of current flow through various components.
6. Compare a conventional full-wave rectifier with a bridge rectifier.
7. What is a filter circuit and how does it function?
8. Explain the relative advantages and disadvantages of a choke-input filter circuit and a capacitor-input filter circuit.

9. What are the two common types of power supplies used in radio receivers? Compare and contrast the two systems of power supplies.
10. What is a voltage regulator? Name some of the important voltage regulators being used in modern electonic equipment.
11. What is SMPS? Explain with a block diagram the basic principle of SMPS.
12. Explain whether True or False.
    1. A filter circuit converts AC into DC.
    2. Full-wave rectification needs at least two diodes.
    3. A bleeder resistor is used to improve the regulation of power supply.
    4. The primary of a CVT uses a resonant circuit.
    5. A voltage doubler is not a rectifier circuit.

    (*Ans.* 1. False, 2. True, 3. True, 4. False, 5. False)
13. What is an inverter? Explain its working with a block diagram.



PART III

RADIO TRANSMISSION AND RECEPTION

chapter 12

Propagation and Transmission of Radio Waves

12.1 RADIO COMMUNICATION

Radio communication is the process of sending information from one place and receiving it in another place without using any connecting wires. It is also called "wireless" communication. Perhaps, the most important form of radio communication is radio broadcasting, including television broadcasting. Radio broadcasting not only provides home entertainment but also serves as a valuable educational aid. Other important applications of radio communication are radio telephone, radio telegraph, police wireless, radio aids to navigation (both air and sea), walkie-talkie and satellite communication. Radio communication is based on the properties of a special type of radiation called *radio waves*.

12.2 RADIO WAVES—CHARACTERISTICS

Radio waves are produced by rapidly changing currents flowing through a conductor. These radio waves spread out in space like ripples produced on the surface of a pond when a stone is dropped into the water. When these fast moving radio waves strike some other conductor placed in their path at a distant point, they produce in the second conductor weak currents of the same nature as the original current which produced these radio waves. Thus a communication called radio communication is established between two distant points.

Radio waves belong to a particular type of waves called *electromagnetic waves*, a form of energy resulting from a combination of electrical and magnetic effects of rapidly changing electric currents. Although not visible to the eye, radio waves travel with the velocity of light waves which is 186,000 miles per second or 300,000,000 metres per second. In fact, both light and radio waves

are electromagnetic waves. Other examples of electromagnetic waves are X-rays, cosmic rays and gamma rays produced by radioactive substances.

Sound also travels in the form of waves but sound waves are not electromagnetic waves. Compared to electromagnetic waves, sound waves travel at a much lower speed of 1,100 feet per second or 330 metres per second. This is the reason why a flash of lightning is seen first and the sound of thunder follows a little later.

12.2.1 Frequency and Wavelength

Figure 12.1(a) represents a complete cycle of a radio wave. The number of such complete cycles performed by the radio wave in one second is called the *frequency* of the radio wave. Figure 12.1(b) represents a wave which performs three complete cycles in the same interval of time as that of the wave in Fig. 12.1(a) to perform one complete cycle. The frequency of the wave represented by Fig. 12.1(b) is, therefore, three times the frequency of the wave represented by Fig. 12.1(a). The unit of frequency is hertz (Hz) which is one cycle per second. This unit is named after Henrich Hertz, who discovered radio waves.

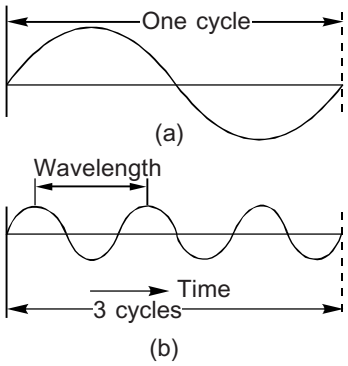


Fig. 12.1 Frequency and Wavelength of a Radio Wave
(a) One Cycle (b) Three Cycles

Radio waves generally possess a frequency of thousands and millions of hertz and are thus represented by larger units called kilohertz and megahertz.

1 kilohertz (kHz) = 1000 hertz (Hz) = 10^3 Hz

1 megahertz (MHz) = 1000,000 hertz (Hz) = 10^6 Hz

The *wavelength* of a radio wave is the distance travelled by the wave during one complete cycle. As shown in Fig. 12.1(b), it is the distance between two successive peaks in the same direction. Wavelength is usually expressed in metres. The frequency, wavelength and velocity of radio waves are connected by the following formula:

$$f\lambda = c$$

where f = frequency in hertz

λ = wavelength in centimetres

c = velocity of light in centimetres per second
$= 3 \times 10^{10}$ cm/s

The above formula helps to convert frequency into wavelength and vice versa.

Every radio broadcasting station or television station is allocated a fixed frequency for operation which is required to be maintained constant within prescribed limits to avoid interference with neighbouring stations. Thus a broadcasting station operating at 1000 kHz has a wavelength of

$$\frac{3 \times 10^8}{1000 \times 10^3} = 300\,\text{m}$$

EXAMPLE: A broadcasting station is operating at 30 MHz. What is the wavelength in metres?

$$\text{Wavelength } (\lambda) \text{ in metres} = \frac{\text{velocity of light in metres per second}}{\text{frequency in hertz}}$$

$$= \frac{300{,}000{,}000}{f\,(\text{Hz})} = \frac{300}{f\,(\text{MHz})}$$

In this case $f = 30$ MHz

so
$$\lambda = \frac{300}{30\ \text{MHz}} = 10\ \text{m}$$

Radio waves of different frequencies are used for different purposes. Radio broadcast stations normally operate at frequencies from a few hundred kHz to 30 MHz but television stations use frequencies above 40 MHz. A complete range of frequencies over which radio signals can be transmitted for various purposes and the classification of these frequencies is given later in this chapter.

12.3 MODULATION

Every transmitting station is assigned a radio frequency (RF) called the *carrier* which can travel over long distances in free space with the speed of light. However, the human ear cannot respond to these high frequencies. If the radio waves are to carry a message or information, some feature of the radio wave must be varied in accordance with the information to be communicated. The process by which the information is superimposed on the carrier is called *modulation.* In the case of radio broadcasts the information or the message generally consists of low frequencies in the range of 20 to 20,000 Hz. These low frequencies are called *audio frequencies* because the human ear can respond to corresponding sound frequencies in the same range.

Audio frequencies by themselves cannot travel long distances but when superimposed on the carrier frequency, they can cover the same distance as the carrier wave itself. A modulated wave is like an aeroplane carrying passengers who could not have reached their destination without the help of the aeroplane. In the case of television broadcasts the modulation frequencies are called *video* frequencies which correspond to the visual information in the picture to be transmitted.

For modulating a radio wave, the two important characteristics of the radio wave that can be varied are the *amplitude* and the *frequency* of the carrier wave. When the amplitude of the carrier is varied in accordance with the variation in the amplitude of the modulating signal (audio frequency), the modulation is called *amplitude modulation* (AM). If, however, the frequency of the carrier is varied in accordance with the variation in the amplitude of the modulating signal, the modulation is called *frequency modulation* (FM).

Both Amplitude and Frequency Modulation are used in radio and television broadcasts. In television the picture transmission uses amplitude modulation but sound transmission employs frequency modulation.

12.4 PROPAGATION OF RADIO WAVES

Radio communication is established between two distant points when radio waves leaving a conductor of suitable length called a *transmitting antenna* are picked up at the receiving end by another conductor of suitable length called the *receiving antenna*. There are a number of modes or paths by which the radio waves travel from transmitting antenna to receiving antenna. The more important of these modes are:

1. Ground wave
2. Sky wave
3. Space wave

12.4.1 Ground Wave

A ground wave travels from transmitting antenna to the receiving antenna along the surface of the earth. It is also known as the surface wave. The earth being a good conductor of electricity absorbs electrical energy from the ground wave as it glides along the surface of the earth. Energy is also absorbed from the ground wave by intervening objects like trees, buildings and hills, etc. The ground wave, therefore, gets weaker and weaker as it travels along the surface of the earth till it becomes so weak that it is no longer useful. The distance over which the ground wave can provide good communication is, therefore, limited. The absorption of energy also depends on the frequency of the ground wave. The absorption by the earth's surface increases with the frequency of the radio waves.

The ground wave provides the best and the cleanest reception within its range. This reception is constant at all times of the day and is not affected by seasonal or other changes in weather conditions. Ground wave propagation is, therefore, used for local and regional broadcast service. The frequencies used are low or medium frequencies up to about 1,600 kHz.

12.4.2 Sky Wave

The transmitting antenna sends out radio waves in all directions. A part of these radio waves travels upward towards the sky and enters a region of the upper atmosphere known as the *ionosphere* which consists of electrically charged

particles called ions. The ionosphere acts like a huge mirror or reflector placed in the sky and sends back to the earth a high frequency radio wave like a ray of light reflected from a mirror. Such a radio wave that is reflected from the ionosphere is called a *sky wave*, as shown in Fig. 12.2. A sky wave can cover much larger distances than the ground wave and makes long distance radio communication possible. In order to understand the propagation of sky waves, we must first know something about the properties of the ionosphere.

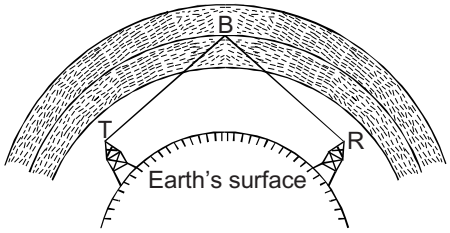


Fig. 12.2 Sky Wave Formed as a Result of Reflection from the Ionosphere

12.5 IONOSPHERE

The upper parts of the earth's atmosphere are constantly bombarded by ultra violet and many other types of radiations from the sun. These radiations not only heat up the atmosphere but also produce negatively and positively charged particles called ions. This highly ionised region of the earth's upper atmosphere is known as the *ionosphere*. This region is also known as the Kennelly-Heaviside layer because it was first discovered by two scientists, A.E. Kennelly and Oliver Heaviside of England.

The ionosphere, which extends from about 80 to 650 km above the earth, actually exists in the form of layers at different heights above the surface of the earth. These layers called D, E and F layers have different densities of ionisation.

The D layer which is actually a region extending between 50 and 90 km above the earth is responsible for much of the day time absorption of high frequency radio waves. At night this layer moves up and farther away from the earth, thereby providing better coverage for high frequency radio waves which cannot be received in the day because of attenuation by the D layer.

The E layer which is normally constant at a height of about 110 km above the earth surface, may at times show variable ionisation between 90 and 130 km.

The F layer which exists at a height of about 220 km is actually a combination of two layers called F_1 and F_2 of which the F_2 layer is more variable with typical heights lying between 250 to 350 km. At night the F_1 and F_2 layers merge into one another to form a single F_2 layer.

The ionisation properties of these layers and their height above the earth change from time to time. These changes take place not only between day and night but also with different seasons of the year. The variations mostly follow a regular pattern but sometimes sudden and irregular changes also take place which disrupt radio communication at certain frequencies.

12.6 PROPAGATION OF RADIO WAVES THROUGH THE IONOSPHERE

When a radio wave enters the ionosphere and travels upwards through layers of increasing ionisation density, it is reflected and gets bent earthwards like a light ray passing from a medium of lower refractive index to one of a higher refractive index. The extent to which a radio wave bends will depend upon the frequency of the radio wave and the ionisation properties of the layer at that time. At frequencies below 30 MHz, the radio wave gets bent as it moves up through the ionosphere and finally gets reflected back to the earth at a point *C* (Fig. 12.3). The radio wave so reflected back to earth is called a *sky wave* and it is this sky wave that makes long distance radio communication possible. Sometimes the sky wave striking the earth at point *C* again gets reflected from the earth and bounces back to the ionosphere from where it is again reflected back to earth at a point still further away from the transmitting antenna. Each of these *bounces* from the earth is called a *hop.* Very long distance communication extending to thousands of miles is sometimes accomplished in two or more hops, by sky waves.

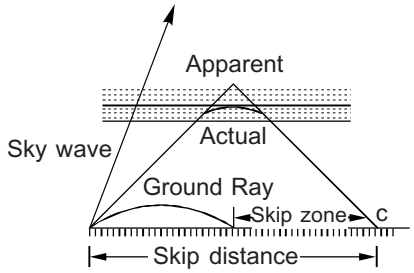


Fig. 12.3 Wave Propagation Showing Skip Distance, Skip Zone and Ground Ray

Radio waves of frequencies higher than 30 MHz are not normally reflected from the ionosphere. They move straight into the ionosphere layers and get lost due to absorption. As such, these waves are not useful for long distance radio communication.

12.6.1 Skip Distance

The distance between the transmitting antenna and the nearest point to which the sky wave returns after reflection from the ionosphere is called the *skip distance* (see Fig. 12.3). Within this distance the only signals that can be received are due to ground waves most of which are absorbed at high frequencies.

There is a zone around the receiving antenna which is covered by neither the ground wave nor the sky wave. Such a zone or belt is called the *skip zone* (Fig. 12.3). Skip distance depends on the frequency of the waves, the time of the day and the angle at which the wave enters the ionosphere. In order to keep

the skip distance same for a particular receiving point for uniform reception during the day and night, it becomes necessary to use different frequencies for transmission during the day and night.

Space Wave Radio waves above 30 MHz are not reflected by the ionosphere. Radio waves at these frequencies cannot travel more than a few hundred feet along the surface of the earth because of heavy absorption by the earth. These waves travel from the transmitting antenna to the receiving antenna in space at about 15 km above the surface of the earth. This space is known as the *troposphere* and the waves travelling through it are called *space waves*. The space wave commonly consists of at least two components as shown in Fig. 12.4. One of these, *TR*, travels directly from the transmitting to receiving antenna and is called the *direct wave*, whereas the other *TER*, reaches the receiving antenna as a result of reflection from the surface of the earth and is called the *reflected* wave. The reflected wave undergoes a phase change of 180° on reflection from the earth. The two components will, therefore, add or cancel each other out at the receiving point *R* depending on the distance *TER*, travelled by the reflected wave.

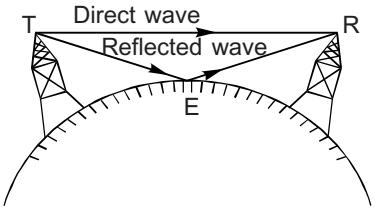


Fig. 12.4 Formation of a Space Wave

Space waves provide communication within the optical range or the *line of sight* distances only. This range does not normally extend beyond the horizon due to curvature of the earth. Hills, trees and tall buildings also obstruct these waves. For communication to be possible, the transmitting and receiving antennas should be able to see each other.

Frequencies above 40 MHz are used for television and FM broadcasts which consequently have a line-of-sight range only.

To extend the horizon for greater coverage at these high frequencies, the transmitting and the receiving antennas have to be raised to as great a height as possible.

Radio waves at these high frequencies are called *microwaves*. These microwaves are used to provide links between two points for telephone communication. With the help of the communication satellites, which can be stationed at suitable heights above the earth, these microwaves are also used for long distance television broadcasts and other radio communication channels extending over thousands of miles.

12.7 FADING

Fading is the name given to the phenomenon due to which the strength of the radio signal received at a point fluctuates with time. The period of these variations in signal strength may be short or long. Fading is specially noticeable at night for frequencies between 500 and 1,000 kHz and at distances of the order of 160 to 1,600 km from the transmitter. Fading may be due to various causes but one of the most common causes is that the direct ray and indirect ray may reach a particular receiving point in opposite phases due to different distances travelled by these rays. If the two rays are of comparable strength, they will tend to cancel each other out resulting in fading.

Selective fading results when the components of a modulated wave fade in a manner in which each is independent of the other. This results in distortion of a received signal as its waveform is not the same as that of the transmitted signal.

Special methods of transmission and reception have to be adopted to avoid fading.

12.8 RADIO WAVE TRANSMISSION

Figure 12.3 showed how radio communication was possible by radio waves which are electromagnetic waves like light waves. These radio waves spreading out in space from their point of origin travel at the speed of light (3×10^{10} cm/s in vacuum) and can induce weak currents of the same frequency in any metallic objects that fall in their way. When suitably modulated, these radio waves can be used to convey information or messages from their point of origin to the point of reception which may be thousands of miles away.

These radio waves get attenuated or weakened as they travel out in space, partly due to the absorption of their energy by reflection and refraction in the ionosphere, partly by the ground and partly by other objects which these radio waves strike. The behaviour of radio waves of different frequencies is different in so far as their reflection or absorption by other objects is concerned. The propagation characteristics of radio waves of different frequencies mainly decide the use to which these waves can be put to for communication purposes. Radio waves have, therefore, been divided into categories or classes with regard to their frequencies. Table 12.1 gives a summary of the classification of radio waves, their propagation characteristics and their typical uses.

To understand how radio waves are generated and radiated into space, consider alternating currents of suitable frequency fed into a conductor or wire of suitable length called the antenna. Fast moving alternating currents produce a moving electric field around the antenna. This field in turn produces a magnetic field at right angles to it. This combination of electric and magnetic fields constitutes the radio wave or electromagnetic wave which is a form of radiant energy.

A conductor carrying alternating currents will radiate a certain amount of electrical energy in the form of electromagnetic waves provided the length of the conductor carrying AC is comparable with the wavelength of the alternating

Table 12.1 Classification of Radio Waves and their Propagation Characteristics

Class	*Frequency range*	*Propagation characteristics and typical uses*
Very low frequency (VLF)	10 to 30 kHz	Low attenuation and propagation characteristics reliable all day—used for long distance communication
Low frequency (LF)	30 to 300 kHz	Day-time absorption more than VLF—used for marine communication and navigational aids
Medium frequency (MF)	300 to 3,000 kHz	High attenuation during day and less attenuation at night—suitable for broadcasting and marine communication
High frequency (HF)	3 to 30 MHz	Propagation characteristics vary with time of day, season and frequency used for long distance communication
Very high frequency (VHF)	30 to 300 MHz	Line of sight propagation—not affected by ionosphere used for television, FM transmission, radar, etc.
Ultra-high frequency (UHF)	300 to 3,000 MHz	Line of sight propagation—used for television and short distance communication
Super-high frequency (SHF)	3,000 to 30,000 MHz	The same as UHF

current flowing through it. Thus, a conductor carrying AC at 100 Hz will have to be 3×10^6 m or 30,000 km long for electromagnetic waves at this frequency to radiate appreciable amounts of energy. On the other hand, currents at 1,000 kHz will need a conductor of only 300 m length which can be constructed practically. It will be clear that HF waves can be conveniently radiated by a small radiator while LF waves require a large radiator or antenna system for effective radiation of electromagnetic energy in the form of radio waves. Waves having frequencies below a minimum limit are not quite suitable for radio communication. This is the reason AF produced by a microphone cannot be radiated directly even if they are made quite powerful by audio amplification. Different transmission characteristics possessed by radio waves of different frequencies are utilised in various ways for radio communication purposes.

Low frequencies from 30 to 300 kHz are not affected by the ionosphere and the ground losses are also very low at these frequencies. These frequencies can provide very stable and dependable transmission throughout the day and night and are quite suitable for radio broadcasting. However, longwave broadcasting is not in use in India at present.

Medium-wave frequencies from 535 to 1,605 kHz known as the *broadcast band*. Due to absorption by the ionosphere, these waves cannot travel more than 300 km during the day. However, the absorption is reduced at night and the sky waves at these frequencies can be transmitted up to about 5,000 km. Most regional broadcast stations, therefore, operate in the medium-wave band or the broadcast band.

For short-wave frequencies between 1,600 kHz to 30 MHz the absorption due to the ground is very good. The sky wave at these frequencies can, however, travel long distances up to about 20,000 km in one or more hops and these short waves are used for overseas transmission over long distances. However, the transmission provided by these frequencies is unstable and is affected by changes in ionospheric conditions over the long distance path.

At frequencies above 30 MHz, the ground wave transmission is practically impossible due to very high ground losses and the sky wave also penetrates the ionosphere without being reflected back to the earth. Transmission at these frequencies is possible only in a straight line connecting the transmitting and receiving antennas. The characteristics of waves at these frequencies resemble light waves and the curvature of the earth limits the transmission distance to the line-of-sight distance only. These frequencies which are used for TV and FM broadcasts provide very stable transmission signals which are not much affected by outside disturbances.

12.9 MODULATION OF RADIO WAVES

Radio waves are only silent carriers and convey no messages unless some of their characteristics are changed in accordance with the information to be transmitted. The method by which some feature of the radio wave, also known as the *carrier* wave, is varied in accordance with the information to be transmitted is called *modulation*.

There are various ways of modulating a radio wave. One method is to turn on and off a radio transmitter in accordance with a prearranged code like the dots and dashes of the Morse telegraph code. This system of radio telegraphy is the CW (continuous wave) system of radio transmission. However, for the transmission of sound or pictures, the features of the radio wave that are varied are the amplitude or the frequency of the carrier wave. Thus, the two most important methods of modulation are the *amplitude modulation* and the *frequency modulation.*

12.9.1 Amplitude Modulation

In amplitude modulation the amplitude of the radiated carrier wave is varied in accordance with the variations of amplitude of the modulating AF wave as shown in Fig. 12.5.

Amplitude modulation is used in radio broadcasting and radio-telephony, where the carrier wave amplitude is modified according to the strength of the audio signal produced by sound pressure variations on the microphone. In the

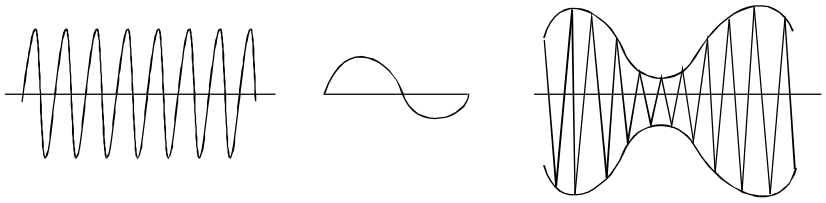


Fig. 12.5 Amplitude Modulation of an RF Carrier (a) RF Carrier Wave (b) AF Modulating Signal (c) RF Amplitude Modulated Carrier Wave

case of TV the modulation signal is the video signal produced by the TV camera from the variations of light intensity in the televised scene.

In Fig. 12.5, a single AF sine wave has been used to modulate a sine wave RF carrier. In actual practice, the audio signal produced by a microphone or the video signal from a TV camera may not be a simple sine wave but a complex wave of the form shown in Fig. 12.6 and the amplitude modulated RF wave will also be of a complex nature.

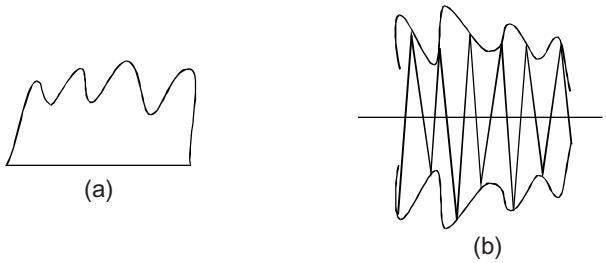


Fig. 12.6 (a) Modulating RF Carrier with Signal from a Microphone (b) Modulated RF Signal

For amplitude modulation, the two frequencies have to be mixed in a suitable manner. This can be done by impressing the carrier wave and audio signal on a device or circuit which has non-linear characteristics. In other words, the modulating device or modulator will possess the characteristics where the current is not directly proportional to voltage. Vacuum tubes (or transistors) have such non-linear characteristics and form very good modulators.

Sidebands When an RF carrier wave is modulated by an AF signal wave by mixing the two frequencies in a modulator tube operated on the non-linear portion of its characteristics, two new frequencies equal to the sum and difference of the combining frequencies are produced by the heterodyning process, similar to the one described in the case of a superheterodyne receiver. Thus, for each AF present in the audio signal, two new frequencies appear, one equal to the carrier frequency plus AF and the other equal to the carrier frequency minus AF. The two new frequencies which appear on either side of the carrier frequency are called *sidebands*. For example, if a carrier wave with a frequency

of 1,000 kHz is modulated with an AF signal of 1 kHz, the two sidebands produced will be: 1,000 kHz + 1 kHz and 1,000 kHz – 1 kHz. The sideband with the higher frequency (1,001 kHz) is called the *upper sideband* and the sideband with the lower frequency (999 kHz) is called the *lower sideband* as shown in Fig. 12.7.

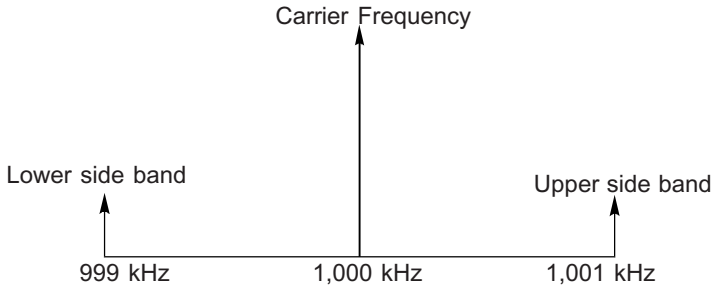


Fig. 12.7 Carrier and Sidebands in Amplitude Modulation

The *bandwidth* or *channel width* required for the transmission of a modulated wave is twice the modulating signal frequency which in the example given is 2 kHz.

In sound broadcasting the carrier wave is modulated not with just a single frequency, as in the previous case, but with a whole set of AFs constituting the voice or music to be broadcast. Each frequency produces two sidebands and the bandwidth required for the transmission of voice or music will be twice the highest frequency contained in the actual programme to be broadcast. In the radio broadcast of music, the highest radio frequency contained in the modulating signal may be as high as 10 kHz and the bandwidth of the transmitting channel required will be 20 kHz. In the case of TV also the bandwidth of the transmitting channel is twice the highest frequency contained in the video signal. It will thus be clear that the tuned circuits in radio transmitters and receivers must be so designed as to be able to pass a whole band of frequencies rather than only the carrier wave frequency.

12.9.2 Power Relations in an Amplitude Modulated Wave

According to Ohm's law, P, the power is given by the relation

$$P = \frac{V^2}{R}$$

where V is the voltage and R is the resistance in the circuit.

An amplitude modulated wave consists of the carrier and the two side bands.

Thus, the total power P_T is given by the relation

P_T = Carrier Power (P_{can}) + Power in lower sideband (P_{LSB}) + Power in upper sideband (P_{USB})

$$P_T = \frac{V_{can}^2}{R} + \frac{V_{LSB}^2}{R} + \frac{V_{USB}^2}{R}$$

where R is the resistance of the antenna in which the power is dissipated. All the voltages are r.m.s values

$$P_C = \text{unmodulated carrier power} = \frac{V_{can}^2}{R} = \frac{\left(\frac{Vc^2}{\sqrt{2}}\right)}{R}$$

$$= \frac{Vc^2}{2R}$$

$$\because \qquad V_{rms} = \frac{V_{peak}}{\sqrt{2}}$$

Since the amplitude of the side bands is proportional to the modulation index m

$$P_{LSB} = \frac{V_{LSB}^2}{R} = \left(\frac{mV_C/2}{\sqrt{2}}\right)^2 \Big/ R$$

$$= \frac{m^2 V_C^2}{4 \times 2 \times R} = \frac{m^2 V_C^2}{8R} = \frac{m^2}{4} \cdot \frac{V_C^2}{2R}$$

Similarly $\qquad P_{USB} = \frac{m^2}{4} \cdot \frac{V_C^2}{2R}$

$\therefore \qquad P_T = \frac{m^2}{4} \cdot \frac{V_C^2}{2R} + \frac{m^2}{4} \cdot \frac{V_C^2}{2R}$

$$P_T = P_C + \frac{m^2}{4} P_C + \frac{m^2}{4} \cdot P_C$$

$$P_T = P_C + P_C \cdot \frac{m^2}{2}$$

$$= P_C \left(1 + \frac{m^2}{2}\right)$$

If $m = 1$, $\qquad P_T = P_C \left(1 + \frac{1}{2}\right) = 1.5\ P_C \qquad (12.1)$

Thus a 100% modulated wave contains 50% more power than an unmodulated wave. All the extra power comes from the sidebands.

The transmitter should be capable of handling the extra power at 100% modulation. Equation 12.1 above establishes a relation between the carrier power and the sideband power at different values of modulation index.

EXAMPLE: An amplitude modulated transmitter radiates a total power of 400 Watts at 50% modulation. Calculate the power in each sideband.

$$P_T = P_C\left(1+\frac{m^2}{2}\right)$$

$$400 = P_C\left(1+\frac{m^2}{2}\right) = P_C\left(1+\frac{(0.5)^2}{2}\right)$$

$$= P_C\,(1 + 0.125)$$

$$= P_C \times 1.125$$

$$\therefore \quad P_C = \frac{400}{1.125} = 355.6$$

$$\text{Total sideband power} = P_C\,\frac{m^2}{2} = 355.6 \times 0.125$$

$$= 44.45 \text{ Watts}$$

$$\text{Power in each sideband} = \frac{44.45}{2} = 22.22 \text{ Watts}$$

Percentage Modulation In amplitude modulation the amplitude of the carrier wave is varied in accordance with the variations in the amplitude of the modulating signal. The extent of variations in the amplitude of the modulated wave can be expressed as the degree of modulation or depth of modulation, which depends on the relative amplitudes of the carrier and modulated wave at any instant during modulation. This degree of modulation is often converted into a percentage called the *percentage modulation.*

In Fig. 12.8, E_o represents the normal amplitude of the unmodulated carrier wave and $E_{\max}$ is the maximum value of the modulated wave at any particular instant, the percentage modulation is given by

$$\% \text{ Modulation} = \frac{E_{\max} - E_0}{E_0} \times 100\%$$

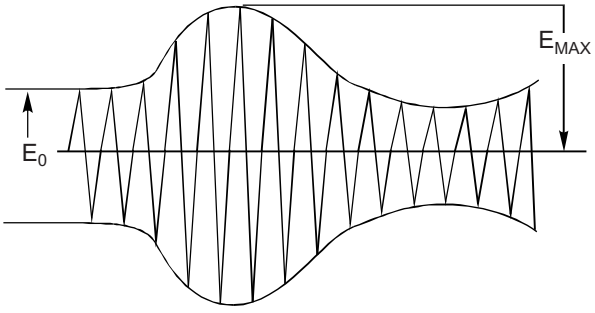


Fig. 12.8 Percentage Modulation

When $E_{\max} = 2E_0$, the modulation is 100 per cent as is clear from the above formula. The modulation is 50% when $E_{max} = 1\frac{1}{2}\ E_0$ and so on. Modulation above 100% is called overmodulation and results in a distorted output waveform.

The power required to modulate a carrier wave depends on the percentage modulation and the type of modulation. In amplitude modulation, the sidebands produced carry a part of the power contained in a modulated wave. It can be shown mathematically that the sidebands contain 50% as much power as the unmodulated carrier so that the two sidebands together make the power of a completely modulated wave 50% greater than the carrier power. A well-modulated carrier has a much greater coverage than an under-modulated carrier wave.

12.10 FREQUENCY MODULATION (FM)

Another method of transmitting information by modulation of the carrier wave is frequency modulation or FM. In FM the frequency of the carrier wave is varied in accordance with amplitude variations in the audio signal; the amplitude of the carrier wave remains constant throughout. Figure 12.9 shows the comparison between an amplitude modulated wave and a frequency modulated wave. It will be seen that in AM the amplitude varies in accordance with the audio modulation and the frequency remains constant, but in the case of FM the frequency varies in accordance with the audio modulation and the amplitude remains constant. This is the main difference between the characteristics of the two systems of modulation.

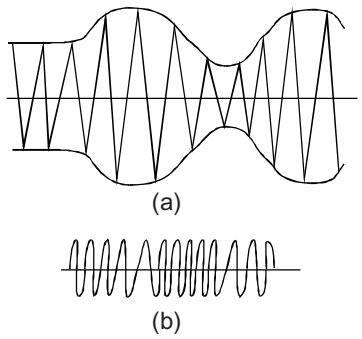


Fig. 12.9 Comparison between AM and FM Waves
(a) Amplitude Modulated Wave
(b) Frequency Modulated Wave

Frequency modulation possesses two main advantages over amplitude modulation. These advantages are:

1. The output from an FM signal is comparatively much less noisy than the output from an AM signal. This is because all natural and man-made noises like atmospheric, static, sparking from electrical machines, etc. can produce only amplitude modulation of the carrier and these cannot produce any effect on FM signals where the amplitude remains constant.
2. The quality of FM broadcasts is much superior to the quality of AM broadcasts. For high fidelity broadcasts of music programmes, the

modulating frequencies can extend up to 10 or 15 kHz. This requires a bandwidth of 20 to 30 kHz which cannot be easily accommodated in the broadcast band without making it overcrowded. To conserve the bandwidth and to provide more channels in the broadcast frequency spectrum, the maximum frequency range of music programmes has to be restricted to only about 7.5 kHz, thereby sacrificing fidelity. However, FM broadcasts are made in the VHF range of 88 to 108 MHz and a large number of FM stations capable of transmitting the full audio band from 20 Hz to 15 kHz can be operated in the same area without any difficulty.

The sound carrier in TV transmission is frequency modulated.

Modulation Index In FM the carrier frequency swings with the amplitude variations of the modulating signal. The maximum permissible frequency swing or frequency deviation has been fixed at ±75 kHz, thereby allowing a maximum bandwidth of 150 kHz. The actual deviation varies with the degree of modulation used by various services. In practice the degree of modulation or *modulation index* in FM is defined as the ratio of the frequency deviation to the modulating frequency, which is given by

$$\text{Modulation index} = \frac{\text{Carrier frequency deviation}}{\text{Modulating frequency}}$$

Thus for a carrier frequency deviation of ±75 kHz and a modulation frequency of 5 kHz, the modulation index will be 75 kHz/5 kHz = 15. The modulation index will vary with the audio-modulating frequency.

Method of modulation or detection of FM waves are different from those used for AM waves. In FM it is the frequency of the carrier that varies with modulation and the detector for FM is a circuit whose output varies with frequency. Such a detector circuit converts the variations of frequency into amplitude variations and is called the *discriminator*. Methods of detection of FM waves have been described in detail in Chapter 19 on television where FM is used for the modulation of the sound carrier.

FM broadcasts are normally made on frequencies in the VHF range. These waves resemble light waves and travel in straight lines. The distances covered by FM waves are limited to line-of-sight distances only because of the obstruction caused by the curvature of the earth and other obstacles in the path of these waves. However, with the increasing use of communication satellites, it is possible to cover much wider areas with FM broadcasts.

12.11 RADIO TRANSMITTERS

A radio transmitter is a device used for generating HF radio waves called carrier waves which can be modulated by the information to be transmitted. The modulated carrier wave is then fed into a suitable antenna system for radiation. The carrier frequency which is generally produced at a low level by an oscillator is then amplified by a number of stages of RF amplification till the required level of output power is obtained. This RF carrier is then suitably

modulated by any one method of modulation described earlier depending upon the type of information to be transmitted.

One method of modulation used in radio telegraph transmitters is to interrupt HF carrier at regular intervals in accordance with a telegraphic Morse code to produce the dots and dashes of the telegraphic system. A key is used for closing and opening the transmitter circuit. The output of the transmitter in such cases consists of continuous HF waves which can be made audible at the receiving end by converting these into LF audio notes by a heterodyning process. This type of transmitters are called *continuous-wave* (*CW*) *transmitters*.

For the transmission of speech or music as in broadcasting, the method of modulation used is either amplitude modulation (AM) or frequency modulation (FM). Transmitters using AM or FM system of modulation are also known as *radio-telephone transmitters*.

12.11.1 Amplitude-modulated Transmitters

Amplitude-modulated (AM) transmitters are mostly used for the broadcast of speech or music. These transmitters either operate on the broadcast band (535-1,605 kHz) or medium-wave band and provide steady service over a limited range. When used for long distance transmissions such as overseas broadcasts, these transmitters operate on higher frequencies between 3 and 30 MHz and are known as HF transmitters or short-wave (SW) transmitters. Figure 12.10 is a block diagram of an amplitude modulated transmitter.

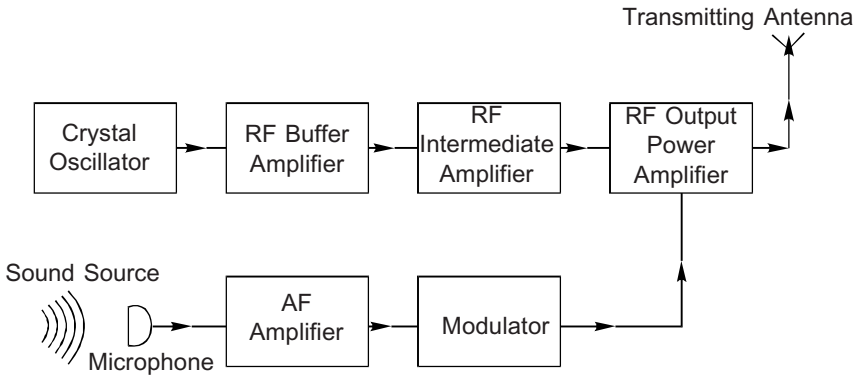


Fig. 12.10 Block Diagram of an AM Transmitter

An amplitude modulated transmitter can be divided into two main parts or chains.

Radio Frequency (RF) Chain The RF chain consists of circuits connected with the production of RF carrier power of the desired strength, i.e. 1 kW, 10 kW, 100 kW and so on. The required frequency is generated at a low power level by a stable oscillator, ordinarily a crystal oscillator. This is followed by a chain of Class C amplifiers that raises the RF power to a level sufficient to drive the final output power amplifier which is also a Class C amplifier. A

buffer amplifier is generally inserted between the oscillator stage and the first Class C intermediate amplifier to isolate the oscillator from the following stages, thereby avoiding any changes in the oscillator frequency due to variations in the loading and coupling circuits of the power amplifier stage.

When the frequency to be radiated is higher than that can be obtained directly from a crystal oscillator, harmonic generators or frequency multipliers are used in the Class C chain. An arrangement for producing stable and steady oscillator frequencies is called the *master oscillator*. The master oscillator is capable of producing frequencies that can cover the various frequency bands and conform to the internationally accepted standards of frequency deviation. Thus, the RF chain in an amplitude modulated transmitter normally consists of the master oscillator, buffer amplifier, multipliers and Class C amplifiers for producing RF voltages of sufficient amplitude to drive the final class C power amplifier.

12.11.2 Modulation Chain

In this chain AFs produced by the microphone in broadcasting studios are amplified in different stages of audio amplification and made sufficiently powerful to be able to modulate the Class C power amplifier to the required degree of modulation.

In transmitters using vacuum tubes the modulation voltage can either be injected into the final power amplifier directly or into any of the earlier Class C stages which operate at a lower level. For high level modulation of the final power amplifier, the modulating signal is inserted in series with the DC plate supply voltage of the transmitter and the system is called *plate modulation*. If the modulation voltage is injected into the control grid of the transmitter the modulation is known as *control grid modulation. Screen grid* or *suppressor grid* modulation results when the modulation is introduced into the screen grid or suppressor grid respectively of the transmitter stage. The plate modulation, being the most efficient and easily adjustable system is most commonly used in high power transmitters. The final modulator amplifier is normally operated as a Class B push-pull amplifier constituting a high efficiency Class B modulation system.

When the final RF power amplifier is modulated by the audio signal, the system of modulation is called *high level* modulation. Modulation in any other earlier stage is known as *low level* modulation. Modulating the oscillator stage itself is to be avoided as this affects the frequency stability of the transmitter.

12.11.3 Collector Modulation

Modern High power AM transmitters normally use Vacuum tubes for the output and penultimate stages but transistors are preferred for the earlier low power stages.

All-transistor transmitters are also used for low power transmitters radiating only a few kilowatts of RF power. Transistors in parallel are employed in such

cases. However, for maximum power output push-pull transistor amplifiers are used in almost all such cases.

Modulation methods employed in transistor transmitters are similar to those used in tube transmitters. Thus collector and base modulation of Class C amplifiers are used similar to Plate and grid modulation employed in Class C tube amplifiers, the properties and advantages being similar in both cases. Figure 12.11 shows a transistor transmitter output stage employing collector modulation. Simultaneous base and collector modulation is generally employed to avoid collector saturation which prevents 100% modulation if only collector modulation is used. Where FETs are used, drain and gate modulation is equally feasible.

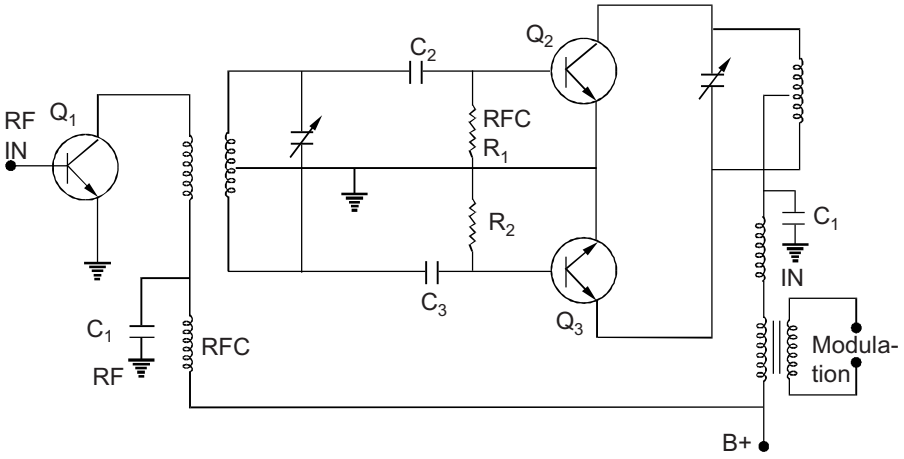


Fig. 12.11 Collector Modulation in All Transistor Transmitter

12.11.4 Some Transmitter Problems

(a) Neutralization Triode transmitting tubes used in High-power transmitters have an interelectrode capacitance (C_{ag}) between the anode and the grid of the tube. This capacitance tends to make the circuit unstable at high frequencies due to feedback from the anode to the grid. The tube has a tendency to oscillate. Circuits and methods employed to provide stability and freedom from feedback are called Neutralizing Circuits. A number of circuits are available for avoiding this feedback and one such circuit is shown in Fig. 12.12. This circuit is suitable for push-pull amplifier which is most commonly used in AM transmitters.

The circuit is self-explanatory. The principle of neutralization used in this case is that of connecting each grid through a neutralizing capacitor C_n equal to the anode grid (C_{ag}) capacity to a point which is at a potential with regard to ground 180° out of phase with its anode circuit. Such a point is very conveniently constituted by the anode of the valve forming the other half of the push-pull. The circuit gives perfect stability and complete freedom from feedback for all frequencies at which the reactance of the leads is negligible.

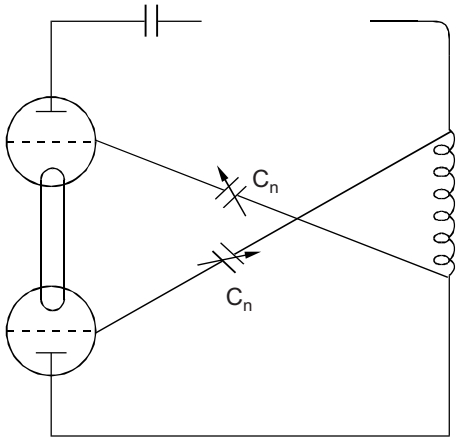


Fig. 12.12 Push-pull Neutralizing Circuit

12.11.5 Transmitter Cooling

In High power AM transmitters using big transmitting tubes, a part of the input power to the transmitting tube is converted into RF power and the balance of the power is dissipated as heat and has the effect of heating the anode and the copper glass seals of the glass envelope. Unless steps are taken to cool the anode and the glass-to-copper seals, the tube structure is likely to get damaged due to over heating. Various cooling methods used to cool the transmitting valves are described below.

Water Cooling In this method the anode is fitted with a jacket through which water circulates. As the anode is at high potential above ground, it is customary to feed the jacket through a long water path of high electrical resistance generally provided by means of a rubber tubing arranged in a coil. The use of distilled water and elaborate pumping arrangements for circulation of water makes it a cumbersome cooling method. This method is not used in modern transmitters.

The glass to copper seal has to be kept cool by air circulation during the operation of the valve.

Air Cooling Most modern valves use air cooling for the anodes of the valves. The anodes are provided with fins which are cooled by air. There are two methods of air-cooling.

(i) *Forced air cooling*. In this method air is blown round the fins on the anode and hot air is allowed to escape into the atmosphere.

(ii) *Suction air cooling.* In this method the air round the anode is sucked by means of suction pumps and the draught so created cools the anode and the surrounding seals.

The suction method has the advantage that it cools the enclosure of the transmitter and the components mounted in the enclosure compared to the blown method which throws hot air into the enclosure.

Vapour Cooling In this method the anode is surrounded by a jacket which contains a liquid with high specific heat. As the liquid evaporates it requires large quantities of heat which is secured from the anode kept at the specified temperature.

12.12 SINGLE SIDEBAND (SSB) TRANSMISSION

It has been shown earlier that when a carrier is modulated with a single sine wave, the resulting frequency spectrum consists of a carrier and two side bands. It has also been shown in this Section that total power contained in a modulated wave is given by

$$P_T = P_C \left(1 + \frac{m^2}{2}\right)$$

where P_T is total power, P_C carrier power and m is the modulation index. At 100% modulation the total power $P_T = P_C \left(1 + \frac{1}{2}\right) = 3/2 \; P_C$. The carrier contains 2/3 of the total power but it contains no information. The carrier can be suppressed without any detriment to information carrying capacity of the modulated wave thereby resulting in a saving of 66% of the total power. The carrier can, of course, be reconstructed at the receiver.

The power contained in the two side bands is $P_C \cdot \frac{m^2}{2}$ and both the side bands carry the same information and are mirror image of each other. If only one of the side bands is transmitted it will result in a further saving of 50% of the remaining power after the carrier is suppressed.

The system in which one of the sidebands is suppressed is called the Single-sideband or SSB. If the carrier and one of the sideband is suppressed the system is called Single Sideband Suppressed Carrier System (SSB-SC).

This system has many advantages and has been adopted by many communication systems. Besides saving considerable amount of power, it results in bandwidth saving and can transmit good quality communication signal at low power and narrow bandwidth.

The saving in power depends on the percentage of modulation. In broadcast AM transmitters, the modulation percentage varies from 0 to 100% modulation, with an average modulation of 30%. The saving effected at this average modulation can be calculated as belows

Power in each side band

$$= P_C \cdot \frac{m^2}{4}$$

If $m = 30\% = 0.3$

$$P_C \cdot \frac{m^2}{4} = \frac{(0.3)^2}{4} = \frac{.09}{4} P_C$$

$$\text{Total power at } 30\% = P_C \left(1 + \frac{(.3)^2}{2}\right)$$

$$= P_C \, 1.0425$$

$$\text{Saving in power} = \frac{1.0425 - .0225}{1.0425}$$

$$= \frac{1.0200}{1.0425} = .98 \text{ which is } 98\%.$$

In other words, a 100 kW (Unmodulated carrier) transmitter using SSB-SC transmission can save 98 kW of power at an average modulation of 30%. This is a huge saving in maintenance cost of the transmitter.

Suppression of Carrier In the production of a SSB-SC signal, the carrier is first suppressed by a circuit known as the balanced modulator and one of the remaining two sidebands is then filtered out.

Balanced Modulator Figure 12.13 gives the circuit of a balanced modulator.

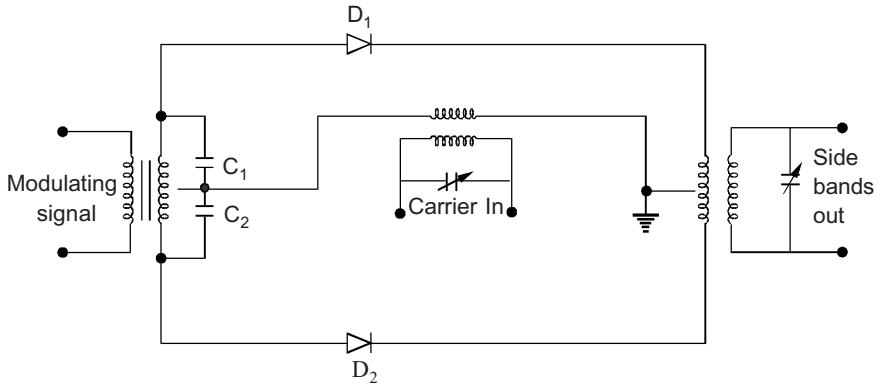


Fig. 12.13 Balanced Modulator

It makes use of two diodes D_1 and D_2 which have non-linear characteristics. The modulating signal is fed in push-pull and the carrier voltage in parallel. The modulated output currents of the two diodes are combined in the centre tapped primary of the push-pull output transformer and they therefore subtract in the output. If the system is made completely symmetrical, the carrier frequency will be completely cancelled or heavily suppressed if the system is not completely symmetrical. The output of the balanced modulator thus contains the two sidebands and some other miscellaneous components which can be taken care of by tuning of the transformers secondary winding. Thus the final output consists of only the two sidebands.

Removal of Unwanted Sideband Of the two sidebands present in the output of the balanced modulator, one can be suppressed for the generation of

the single-sideband suppressed carrier (SSB-SC) system. A number of methods are available for the suppression of the unwanted band but only two of these methods are discussed below.

(a) Filter Method of SSB-generation This method is simple but the design of the filter is not. It is necessary to provide a filter cut-off of one side which will separate signals differing by only twice the lowest modulating frequency. Circuits have been designed using powdered iron-core inductors, quartz crystals as resonators or mechanical resonant systems which perform satisfactorily in this regard.

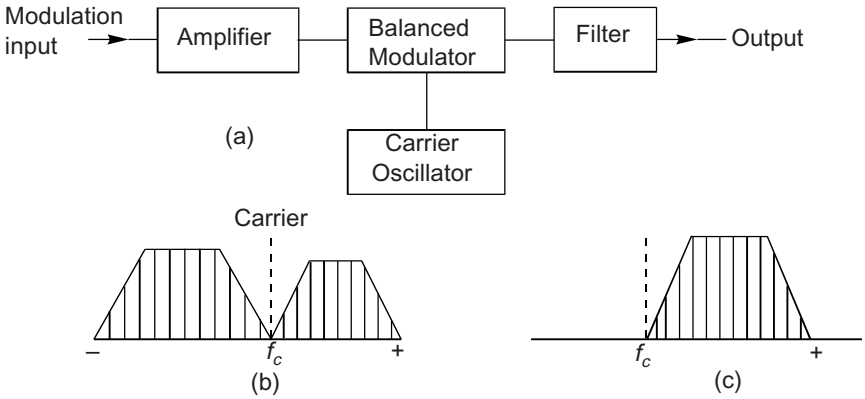


Fig. 12.14 (a) Filter Generation of SSB, (b) Output Balanced Modulator, (c) Output after Filter

The filter may be adjusted to select either the upper or the lower side band. After filtering the desired frequency is obtained by frequency multiplication.

(b) SSB by Phasing Method Generation of SSB signal by the phasing method requires two balanced modulators and two 90° phase shifter networks, one giving a constant phase shift over the modulating frequency range, the other providing a 90° phase shift over the range of carrier frequencies expected.

As in the Fig. 12.15 the modulation frequencies are fed to the two balanced modulators, one directly and other through 90° phase shift network. The carrier is likewise supplied to the two balanced modulators in quadrature. The outputs of the two modulators, consisting of sidebands only, are then added or subtracted to give either the upper or the lower sidebands as output*.

---

* The input to the two modulators are

	(1)	(2)
Modulators	$Em \sin \omega_m t$	$Em \cos \omega_m t$
Carrier	$Ec \sin \omega_c t$	$Ec \cos \omega_0 t$

It may again be assumed that the modulators have parabolic characteristics given by

$$Z = a_1 ed + a_2 ed^2 + \ldots$$

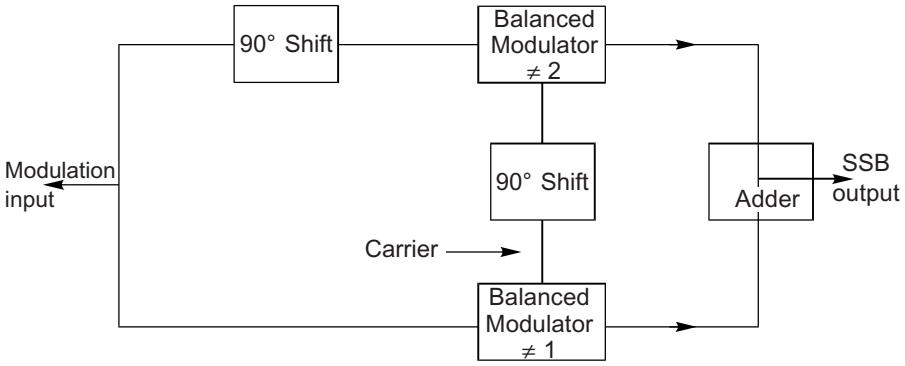


Fig. 12.15 Phase-shift Method of SSB Production

Summary

Radio or wireless communication is the process of sending information from one place and receiving it in another place by means of radio waves. Radio waves are disturbances produced in space by rapidly changing electrical currents flowing through a conductor. Radio waves are electromagnetic waves which travel with the velocity of light which is 18,6000 miles per second or 3×10^{10} cm/s. For communication of information, high frequency radio waves have to be modulated with low frequency waves corresponding to sound waves. These modulated carrier waves strike a receiving antenna and produce in it similar high frequency currents. These high frequency currents are selected, detected, amplified and again converted into sound waves by a radio receiver. Depending on frequency, radio waves travel from transmitting to receiving antenna either along the surface of the earth as *surface waves*, or by reflection from the ionised layers of

---

For modulator (1) the two inputs are

for T_1 $\quad ed_1 = E_m \sin \omega_{mt} + E_c \sin \omega_{ct}$

for Tt_2 $\quad ed_2 = -E_m \sin \omega_{mt} + E_c \sin \omega_{ct}$

Using these inputs and assuming an output circuit resonant to ω_c, the output of the balanced modulator is given as,

$$Eo_1 = 2ka_2 \sin E_c [\cos (\omega_c + \omega_m)t - \cos (\omega_c - \omega_m)t]$$

for modulator (2) the two inputs are

$$\left.\begin{aligned} e'd_1 &= E_m \cos \omega_m mt + E_C \cos \omega_c et \\ e'd_2 &= -E_m \cos \omega_m mt + E_C \cos \omega_c et \end{aligned}\right] \text{Both shifted } 90°$$

and again using the same process, the input of the second modulator may be written as

$$Eo_2 = 2ka_2 E_m E_c [\cos (\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$$

If the outputs of the two modulators are added

$Eo_1 + Eo_2 = Ka_2 E_m E_c \cos (\omega_m + \omega_c)t$ and the upper side band is obtained.

The subtraction of the two outputs gives the lower side band

$Eo_2 - Eo_1 = Ka_2 Em Ec \cos (\omega_c - \omega_m)t$.

K and k are different constants.

the ionosphere as *sky waves* or direct from transmitter to receiver as *space waves.* Low and medium frequency waves travel as ground waves and have a limited range due to absorption by the earth's surface. High frequency waves travel as sky waves and can cover very long distances. Very high frequency waves, above 40 MHz used for television, travel straight from the transmitter to receiver and have only a line-of-sight range. Communication satellites are used to extend the range of these very high frequency waves called *microwaves.* Fading is the phenomenon due to which the strength of the radio waves received at a point fluctuates. Special methods of transmission and reception have to be adopted to avoid fading. Radio waves are electromagnetic waves which travel in space with the speed of light which is 3×10^{10} cm/s. The propagation characteristics of radio waves of different frequencies mainly decide the use to which these waves can be put for communication purposes. Radio waves have accordingly been divided into various categories or classes with regard to their frequencies and propagation characteristics.

Modulation is the process by which certain characteristics of the radio waves are varied in accordance with the information to be transmitted. The two important methods of modulation are amplitude modulation (AM) and frequency modulation (FM). In AM the amplitude of the RF carrier wave is varied in accordance with the amplitude of the modulating audio frequency. AM is widely used in radio broadcasts and for picture transmission in TV broadcasts. In FM the frequency of the carrier wave is varied in accordance with amplitude variations of the modulating signal. FM is used for radio broadcasts and for the sound channel in TV broadcasts. The quality of FM broadcast is superior but the transmission range is limited to line-of-sight distances only.

A radio transmitter is a device for generating high frequency radio waves which can be suitably modulated by the information to be transmitted. Depending on the type of modulation used, a transmitter is known as a continuous wave transmitter, an amplitude modulated transmitter or a frequency modulated transmitter. An AM transmitter normally consists of an RF chain where the carrier waves are produced and amplified to the desired output power. In the modulation chain the audio frequencies from the microphone are amplified and made sufficiently powerful to be able to modulate the Class C amplifier.

Neutralization of triode transmitting tubes is necessary in high power transmitters to avoid instability due to inter-electrode capacitance of the tube. Transmitter cooling is another problem that has to be overcome by various methods of transmitter cooling.

Single sideband system with suppressed carrier (SSB-SC) is commercially used to save power as well as bandwidth. Balanced modulator is a device commonly used for the suppression of carrier. A number of methods are available for removal of one of the sidebands.

Review Questions

1. What is radio communication? Why is it called wireless communication?
2. What are radio waves?
3. What is the velocity of radio waves? How much time will a radio wave take to travel around the earth if the circumference of the earth is 25,000 miles?
4. What is meant by frequency and wavelength of a radio wave? What is the relationship between the two?
5. A radio station is broadcasting at 1,370 kHz. What is the corresponding wavelength of the station?
6. What is ionosphere? What are its different layers and how do they affect radio communication?
7. What is fading? How can fading be overcome?
8. Define skip distance. How can skip distance be kept constant?
9. What are electromagnetic waves? How are these waves used for radio communication?
10. What is the frequency range of the radio waves used for the following purposes:
    (i) Broadcasting, (ii) long distance communication and (iii) television.
11. What are two important types of modulation used for radio broadcasting? Give their relative advantages and disadvantages.
12. What are sidebands in Amplitude Modulation? Calculate the frequency of the sidebands when a 100 kHz carrier is modulated with a 500 Hz signal. (Ans. 100.5 kHz 99.5 kHz)
13. Give the block diagram of an amplitude modulated transmitter and explain the working of each unit.
14. Define modulation index both in the case of amplitude modulation and frequency modulation. How is the modulation index calculated in each case.
15. Derive a relation between the total modulated power and the unmodulated carrier power in an amplitude modulated wave.
    Calculate the carrier power of an AM radio transmitter which is radiating 10 kW of power at a modulation percentage of 60. (*Ans.* 8.47 kW)
16. What is single sideband system of radio transmission. What are its advantages.
    Calculate the percentage saving in power in a single sideband suppressed carrier (SSB-SC) AM system when the modulation percentage is 50. (*Ans.* 94.4%)
17. What is a balanced modulator? Explain the use of a balanced modulator in the production of a SSB-SC signal.

chapter 13

Transmission Lines and Antennas

INTRODUCTION

The RF energy generated in the tank circuit of a transmitter has to be fed to a transmitting antenna which is generally a wire or a mast situated at a considerable distance from the transmitter. The wires connecting the output of the transmitter to the radiating antenna are called transmission lines. At higher frequencies the length of the transmission lines becomes an appreciable fraction of the wavelength being transmitted and these transmission lines are not just connecting wires between the transmitter and the antenna but play a significant role in the efficient propagation and radiation of the RF energy. The size, separation and general lay out of the system of wires determine their characteristics. These characteristics will now be studied in detail.

13.1 TRANSMISSION LINES

A transmitting antenna is normally constructed in an open space and in most cases is located several hundreds of feet away from the output stage of the transmitter. The RF energy from the output stage of the transmitter has to be transferred to the antenna by means of connecting wires known as *transmission lines* or *feeder lines*. A transmission line may be a pair of parallel wires separated by a certain distance (Fig. 13.1) or it may be a coaxial cable consisting of a grounded outer conductor and an insulated inner conductor. A good transmission line should be able to convey RF currents from the transmitter to the antenna without excessive losses and without itself radiating any energy. Transmission lines are required even at the receiving end where the receiving antenna and the receiving equipment are generally located at separate places. The RF energy at the receiving end being very small, prevention of RF losses on the transmission lines is even more important.

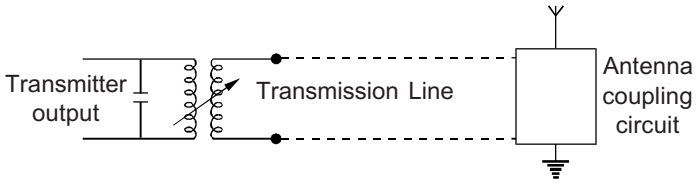


Fig. 13.1 A Parallel Wire Transmission Line

A parallel wire transmission line will also have distributed capacitance and inductance as in the case of an antenna. When of a suitable length, the transmission line will also form a resonant circuit at the frequency of the RF generator. Standing or stationary waves will be formed on the transmission line which will start radiating electromagnetic energy in the same way as a resonant antenna does. This is not desirable. To prevent undesirable radiation by a transmission line, the two parallel wires constituting the transmission line are brought close together to a separation of about 5 to 15 cm. Since the flow of currents in the two wires is in opposite directions, the standing waves formed on the two wires cancel each other out. In fact, the half-wave resonant antenna can be considered as an extension of a resonant transmission line whose two limbs have been spread apart to allow formation of standing waves which result in radiation of radio waves by the antenna portion only. Figure 13.2 shows the distribution of current and voltage on an half-wave antenna connected to a resonant transmission line. The radiation from the vertical leads of the transmission line is mutually cancelled.

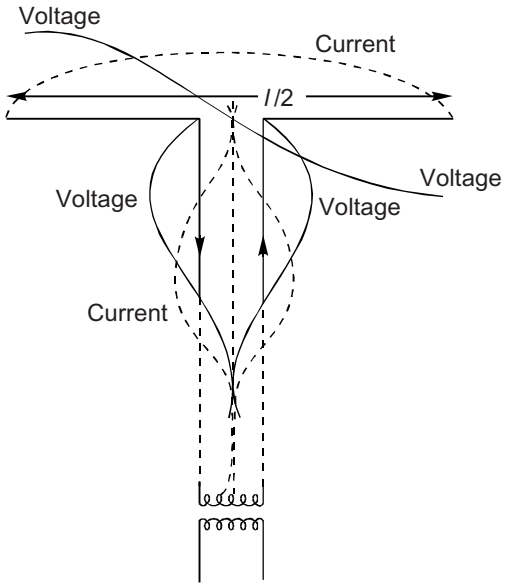


Fig. 13.2 Current and Voltage Distribution on a Resonant Line Connected to a Halfwave Antenna

13.2 CHARACTERISTIC IMPEDANCE

A transmission line has distributed capacitance and inductance all along its length and presents a certain amount of impedance to the generator feeding RF energy into it. This impedance is called the *characteristic impedance* or *surge impedance* of the transmission line and is represented by Z_0. It can be calculated from the formula

$$Z_0 = \sqrt{\frac{L}{C}}$$

where Z_0 = characteristic impedance

L = inductance per unit length of the line

and C = capacitance per unit length of the line.

If a transmission line is terminated in an impedance equal to its characteristic impedance, all the energy travelling down the transmission line will be absorbed and there will be no standing waves formed. Such a line behaves like a line of infinite length which will allow no reflection of waves from the far end and thus prevents loss of energy from the transmission line by radiation due to standing waves. In actual practice the transmission line is connected to the antenna in such a manner that the antenna itself presents an impedance to the transmission line equal to the characteristic impedance of the same.

13.3 TYPES OF TRANSMISSION LINES

The type of transmission line most commonly used for feeding half-wave antennas is the two-wire transmission line which has a characteristic impedance of 500 to 600 Ω depending on the size of conductors and the distance between the two parallel conductors. In the case of a coaxial transmission line consisting of a copper wire running along the centre of an outer and larger tube and insulated from it by spacers, the characteristic impedance is much lower. A coaxial transmission line having an outer tube diameter of 0.95 cm will have a characteristic impedance of about 75 Ω. The outer tube which is generally earthed acts as a shield to prevent radiation. A twisted pair transmission line consisting of two insulated wires twisted around each other has a characteristic impedance of about 73 Ω. A twisted pair transmission line has greater flexibility than a parallel wire line.

The impedance of the half-wave Hertzian antenna varies from 73 Ω at the centre to very high values at the end. Any of the transmission lines described so far can be connected to suitable points on the half-wave antenna which corresponds to the characteristic impedance of the transmission line. A properly terminated line becomes a non-resonant transmission line and may be of any length.

For efficient transfer of RF energy from the transmitter to the antenna, the transmission line input impedance must be properly matched to the output impedance of the transmitter at one end and the output impedance of the transmission line must be matched to the input impedance of the antenna at the

other end. For this purpose, suitable coupling networks or impedance matching devices have to be used at both ends of the transmission line. At the transmitter end the coupling network generally consists of an air-cored transformer with variable coupling between the primary and secondary windings to adjust the loading of the line. At the antenna end, the coupling network may consist of a combination of inductors and capacitors unless the transmission line can be connected directly to the antenna as mentioned earlier. The arrangement normally adopted for matching in broadcast stations is shown in Fig. 13.3.

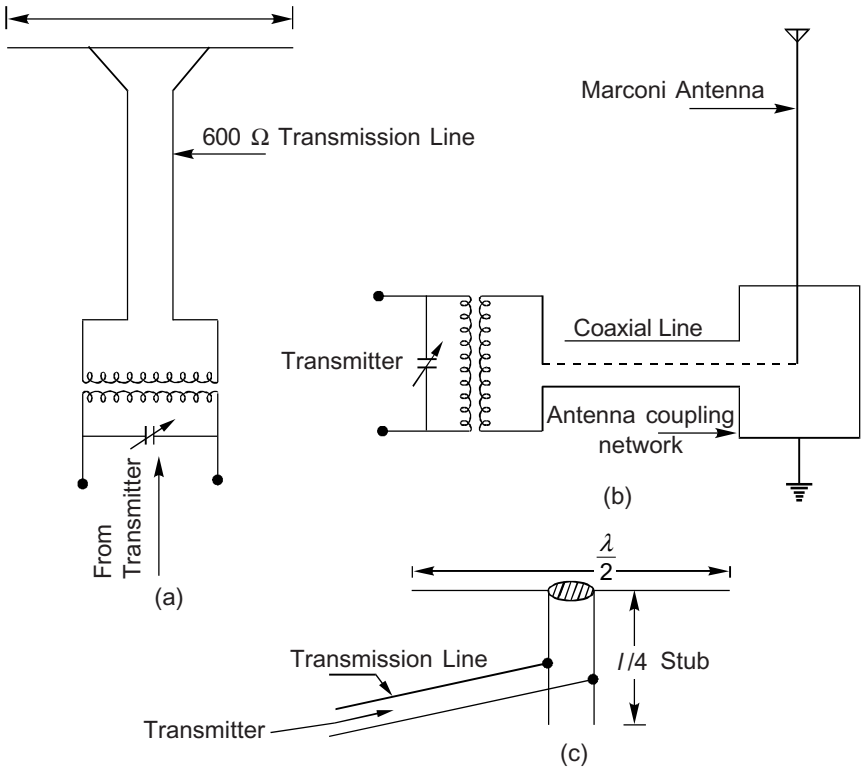


Fig. 13.3 Matching Transmission Line to Antenna (a) Two-wire Transmission Line Matching (b) Coaxial Line Matching (c) Stub Matching

In the case of HF dipole antennas a quarter-wavelength of the transmission line called the *stub* is attached to the non-resonant transmission line at a point where the characteristic impedance of the line equals the characteristic impedance of the stub, which acts as a coupling transformer between the transmission line and the antenna.

13.4 ANTENNAS—DEFINITION

In order that the high frequency energy produced in the tank circuit of the power amplifier of the transmitter can be radiated into free space as radio waves, the modulated RF carrier currents must be transferred to a wire

structure called the *antenna* or *aerial*. The function of the antenna is to radiate into space the strongest possible radio waves produced by the power amplifier of the transmitter. For this the RF energy from the transmitter is transferred to the antenna in a most efficient manner by connecting wires called *transmission lines*. For reasons of efficient radiation of radio waves, in most cases the antenna is located at a remote place from the transmitter. In such cases, the transmission lines connecting the output stage of the transmitter to the antenna carry the RF energy currents in the same way as the power transmission lines connecting the generating station to the substation or distribution station carry the 50 Hz low frequency power supply currents.

Whereas the function of a transmitting antenna is to convert the RF energy into radio waves which travel in space with the velocity of light, the function of a receiving antenna is to pick up these radio waves and convert them into HF currents which are transferred to the radio receiver by means of transmission lines. But for the different functions performed by the transmitting and receiving antennas, the two types of antennas have identical properties.

13.5 PRODUCTION OF E.M. WAVES

A wire in open space carrying HF currents can convert a part of the RF energy into radio waves or electromagnetic waves under suitable conditions. The actual mechanism of the conversion of RF energy into radio waves is rather complicated and involves complex mathematical calculations. However, by making certain reasonable assumptions, it can be shown that energy gets detached from a conductor or a circuit whenever the current in the circuit changes. For this, a wire or a conductor in open space can be considered to consist of a large number of capacitors formed between the conductor and the earth and a large number of small inductors or coils spread over the entire length of the conductor. When these capacitances and inductances are lumped together, the antenna can be represented by a circuit consisting of a capacitor whose plates are connected by a vertical wire having a certain amount of self-inductance as shown in Fig. 13.4(a). The generator in the circuit represents the source of RF energy which feeds the antenna and whose frequency is generally the resonant frequency of the oscillatory circuit.

In order to study the phenomenon of detachment of energy from the antenna in the form of radio waves, it will be necessary to consider the distribution of electric and magnetic fields round the antenna wire during one complete cycle of the applied RF voltage. Considering the distribution of the electric field first, we may start the cycle when the capacitor is charged to its maximum potential difference, making the top plate positive with regard to the bottom plate, and the current in the wire is zero. At this moment the field around the antenna is entirely electric with the lines of electric force connecting the positive charge or the upper plate to the opposite negative charge on the lower plate as in Fig. 13.4(b).

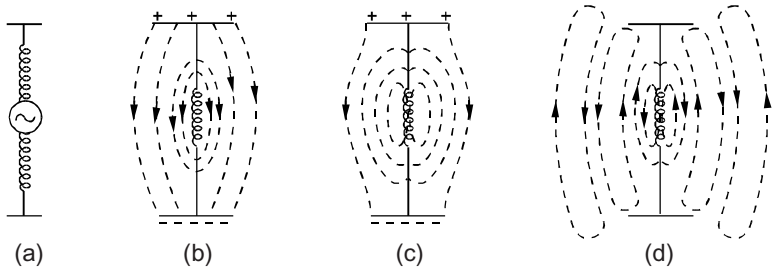


Fig. 13.4 Production of Electromagnetic (radio) Waves (a) Antenna can be Represented by a Capacitor Whose Plates are Connected by a Wire with Self-inductance (b) Capacitor Charged to Maximum Potential and Field Round the Antenna is Entirely Electric (c) Current Flows and Electric Field Starts Collapsing (d) New Electric Field Starts Building up before the First One has Actually Disappeared

After the moment of maximum pottential difference in the cycle, the electrons start moving upward, thereby constituting a flow of current. The electric field starts collapsing and the ends of the lines of force tend to come together along the wire as indicated in Fig. 13.4(c). As per Lenz's law, the current continues to flow even after the electric field has collapsed completely and the capacitor starts getting charged in the opposite direction producing new lines of electric force in the reverse direction to the earlier field. As a result of some time lag between the collapse of the initial field and the changes in potential responsible for the collapse, the new electric field starts building up before the first one has actually disappeared. The first disturbance is driven outward in the form of closed loops by the new electric field due to mutual repulsion as shown in Fig. 13.4(d) where the direction of the lines of force in the inner surface of the first and the outer surface of the second is the same.

In addition to the electric field, the circuit is also surrounded by rings of the magnetic fields which are produced as a result of varying currents through the wire. The intensity of the magnetic field depends on the current strength and the direction also alternates with the current. However, in the immediate vicinity of the current carrying wire, the magnetic lines of force are at right angles in space to the electric lines of force. The two fields are also at right angles to the direction of propagation of the waves. The electric and magnetic fields which exist together are 90° out of phase in time and at right angles to each other in space. The two fields moving together constitute what is known as an electromagnetic wave or a radio wave which travels in space with the velocity of light of 3×10^{10} cm/s. Relative directions of the electric field (Z), the magnetic field (X) and the direction of propagation of the wavefront (Y) are indicated in Fig. 13.5.

13.6 ANTENNA AS A RESONANT CIRCUIT

An antenna circuit's distributed inductance L, capacitance C and resistance R can be represented by a tuned circuit as in Fig. 13.6. When coupled to a

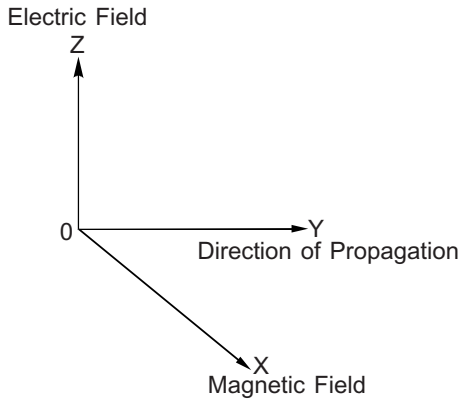


Fig. 13.5 Relative Directions of Electric Field, Magnetic Field and Direction of Propagation of Radio Waves

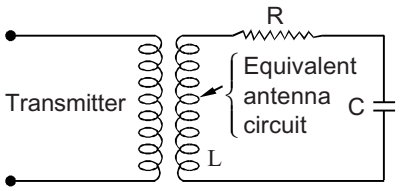


Fig. 13.6 Antenna as a Resonant Circuit

transmitter it will absorb the maximum energy for radiation when the equivalent tuned circuit is in resonance with the transmitter frequency. It is thus clear that for most efficient radiation of energy, the antenna wire length must bear a definite relation to the wavelength of the radiated radio waves.

In a resonant circuit the capacitor charges and discharges every quarter cycle of the resonant frequency. A centre-fed antenna of the type shown in Fig. 13.7 will resonate with the tank circuit frequency of the transmitter if the time taken by the electrons to reach the ends of the antenna rod and return to the centre is the same as the time taken for the charging and discharging of the tank circuit capacitor, which is the time for half a cycle of the tank frequency. In terms of wavelength this is equal to half a wavelength, which means that the length of each rod of the antenna should be a quarter wavelength or the length of the entire antenna should be half a wavelength. Such an antenna is called the *dipole* or *half-wave* antenna. Figure 13.7 shows the representation of a dipole antenna with the current and voltage distribution along the length of the half-wave antenna indicated on it.

The electron flow stops at the ends of the wire where the electrons pile up and there is no current flow. Such a point is called the current *node*. The electrical potential is the maximum (antinode) at these points. At the centre of the antenna, the electrons have the greatest motion and there is no accumulation of electrons giving rise to a point of minimum electrical potential or a voltage node. The current is maximum at this point and a current loop is

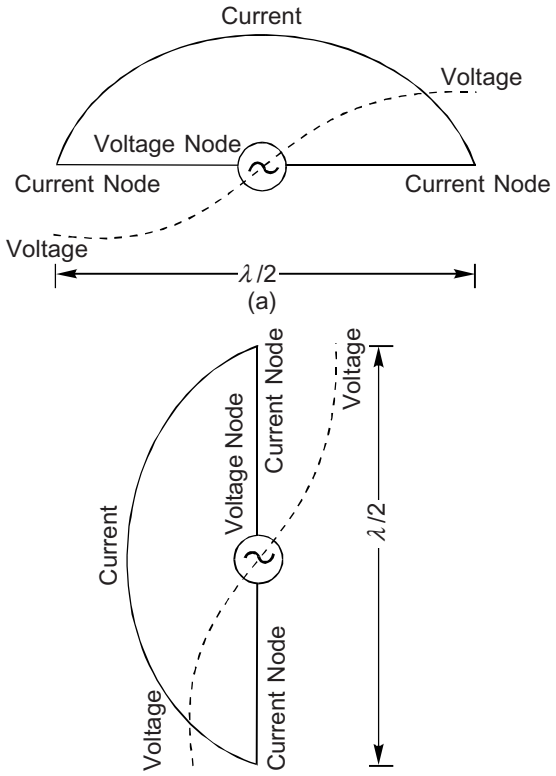


Fig. 13.7 Current and Voltage Distribution in a Half-wave Antenna (a) Horizontal Dipole (b) Vertical Dipole

formed. Thus, we find that a current node appears at a point where there is a voltage antinode and vice versa. This formation of nodes and antinodes of current and voltage along the length of the antenna gives rise to what is known as standing waves or stationary waves which are responsible for the electromagnetic waves leaving the antenna in the form of electric and magnetic lines of force.

For calculating the length in metres of a dipole or half wave antenna for operation at a particular frequency, we can make use of the formula connecting the frequency, wavelength and the velocity of electromagnetic waves (light waves) as already explained in Fig. 12.2. The following example will illustrate the method of calculation to be adopted.

EXAMPLE Calculate the length of each rod of a dipole operating on a frequency of (i) 1000 kHz and (ii) 7.5 MHz.

The formula applicable in this case is

$$f \cdot \lambda = c \qquad (13.1)$$

where f = frequency of the wave

λ = wavelength of the wave

c = velocity of electromagnetic waves which is 3×10^{10}cm/s

(i) f = 100 kHz

$c = 3 \times 10^{10}$ cm/s

λ = ?

$$\lambda = \frac{c}{f} = \frac{3 \times 10^{10}}{1000 \times 10^{3}} = 3 \times 10^{4} \text{ cm}$$

or 3×10^{2}m = 300 m

Total length of the dipole $\left(\frac{\lambda}{2}\right)$ = 150 m

Length of each rod $\left(\frac{\lambda}{4}\right)$ = 75 m

(ii) f = 7.5 MHz

$$\lambda = \frac{3 \times 10^{10}}{7.5 \times 10^{6}} = \frac{3 \times 10^{4}}{7.5} = \frac{30}{7.5} \times 10^{3} \text{ cm}$$

= 4000 cm = 40 m

Total length of the dipole $\left(\frac{\lambda}{2}\right)$ = 20 m

Length of each rod $\left(\frac{\lambda}{4}\right)$ = 10 m

The above calculations indicate that the length of a dipole antenna depends on the frequency of operation. A dipole antenna of the type discussed above is also known as a Hertz antenna after the name of its inventor, Heinrich Hertz.

13.7 EARTH AS REFLECTOR OF RADIO WAVES

From the above example, it will be clear that the length of a dipole or half wave antenna becomes unmanageably large at lower frequencies in the broadcast band. This creates practical difficulties in the construction of transmitting antennas, particularly of the vertical type. In such a case use is made of the reflecting properties of the ground which acts like a reflector or electrical mirror for radio waves. The length of the vertical dipole is made only a quarter of a wavelength $\lambda/4$ and one end of the antenna is grounded as shown in Fig. 13.8. The reflected image of the vertical $\lambda/4$ antenna forms the other limit of the vertical dipole and the two together behave like a normal half wave antenna. Such an antenna is called a Marconi antenna after the name of its inventor, Guglielmo Marconi.

In practice the actual length of the antenna is kept 5% less than the calculated length. This is due to the fact that the velocity of the electric current in the antenna wire is about 5% less than the velocity of the radio waves.

The effective length of an antenna can, in certain cases, be increased or decreased by connecting an inductor or capacitor in series with the antenna wire as shown in Fig. 13.9. By connecting an inductance in series with the antenna, the total inductance of the antenna increases, its natural frequency of

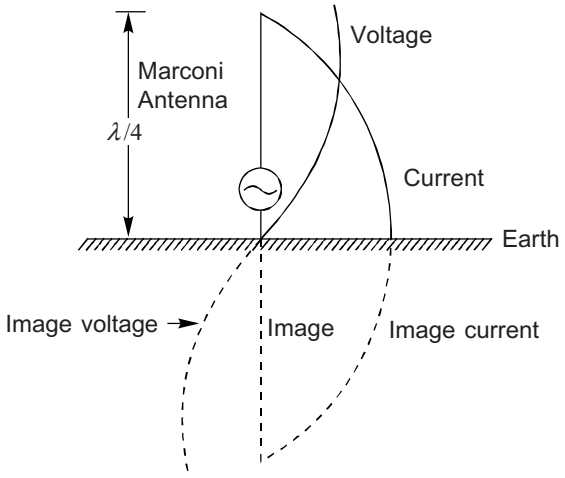


Fig. 13.8 Current and Voltage Distribution in a Marconi Antenna

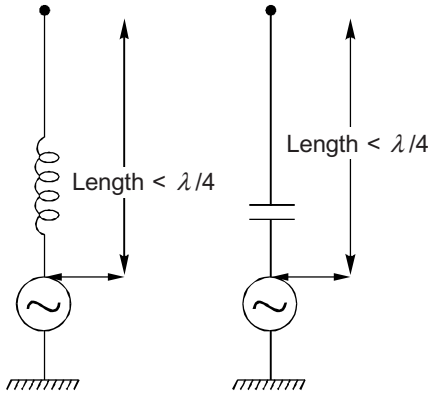


Fig. 13.9 Increasing or Decreasing the Effective Length of a Marconi Antenna

resonance decreases and the effective length of the required wire becomes more than the actual length. However, when a capacitor is connected in series, the total effective capacitance of the antenna decreases and the natural resonant frequency increases, thereby making the effective length less than the actual length of the antenna.

13.8 ANTENNA ARRAYS

A horizontal dipole of the type described earlier will radiate energy uniformly in all directions in a vertical plane and its radiation pattern can be represented by a circle drawn with the dipole at its centre as in Fig. 13.10(a). In the horizontal plane it will radiate maximum energy in a direction at right angles to its own line but no energy will be radiated in the direction of its own line. The actual radiation pattern will be like a figure of 8 as in Fig. 13.10(b).

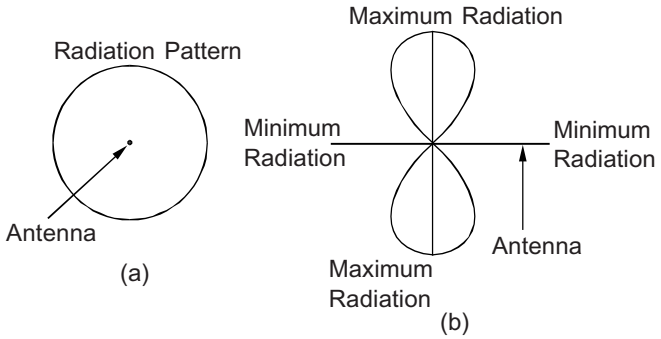


Fig. 13.10 Radiation Pattern of a Horizontal Half Wave Antenna

The radiation patterns of antennas when drawn as a graph are also known as *polar diagrams.* If two half-wave antennas are placed near each other, the radiation from the two antennas may cancel out in certain directions and add up in other directions thus forming a directional radiation pattern depending on the distance between the two antennas and the relative phase of the currents flowing in each antenna.

Two or more half-wave antennas stacked horizontally or vertically and suitably spaced form an *antenna array.* Such antenna arrays are often used in broadcasting to form radio beams which are directed towards the sky at a particular angle to be suitably reflected from the ionosphere to cover certain target areas.

Two popular types of antenna array are (a) Broadside array and (b) End-free array.

(a) Broadside Array As indicated in Fig. 13.11 a broadside array consists of a number of identical radiators (dipoles) equally spaced along a line and carrying the same amount of current in phase from the same source.

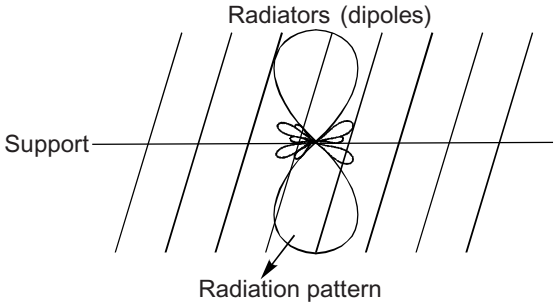


Fig. 13.11 Broadside Array with Radiation Pattern

As indicated by the directional pattern on the diagram this array is strongly directional at right angles to the plane of the array.

(b) End-Fire Array The physical arrangement of the end-fire array is shown in Fig. 13.12. The arrangement of the radiating elements is the same as in the

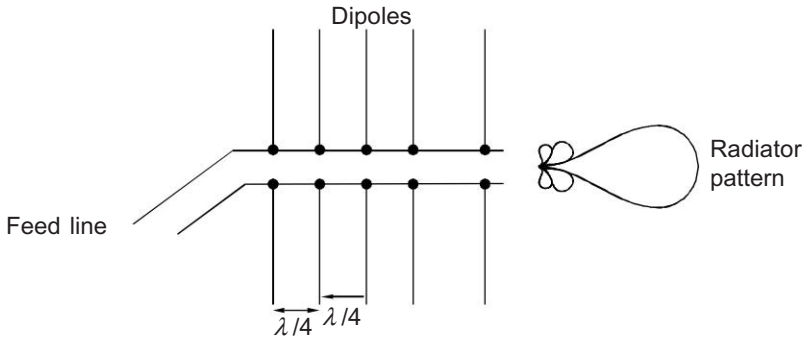


Fig. 13.12 End-free Array will Rareless Pattern

broadcast array. Each element contain the same magnitude of current but with a progressive phase difference between these currents. The phase difference is progressive from left to right, with each succeeding element lagging in phase on the preceding element.

The directional pattern is in the plane of the array and not at right angles to it. Moreover, the directional pattern is unidirectional and not bidirectional.

It is possible to combine several different arrays to obtain higher directional pattern. Such combinations of arrays are often used in point to point relay particularly for overseas broadcasting.

13.8.1 The Yagi-Uda Antenna Array

A Yagi-Uda antenna array consists of a driven element and one or more parasitic elements. Figure 13.13(a) shows the relative position and lengths of the various elements and Fig. 13.13(b) indicates the radiation pattern produced by the array.

As shown by the radiation pattern it is a relatively unidirectional array with a moderate gain of about 7 dB. It is used as an HF transmitting antenna and

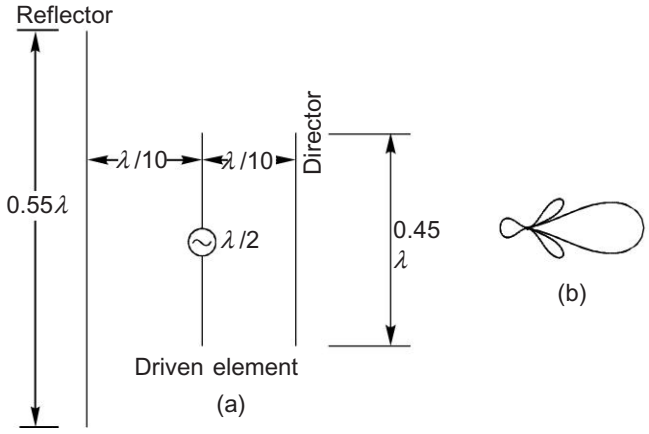


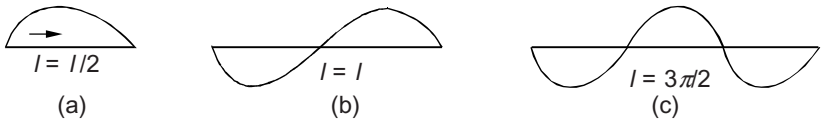
Fig. 13.13 (a) Yagi-Uda Array, (b) Radiation Pattern

also as a receiving antenna for VHF Television frequencies. The back lobe of the directional pattern can be reduced and front-to-back ratio of the antenna can be improved by bringing the radiation closer but this will lower the impedance of the array.

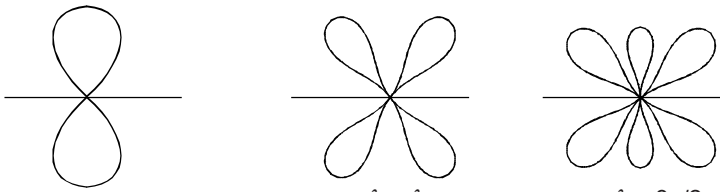
When used as a TV receiving antenna, the Yagi-Uda array makes use of folded dipole. As used in practice, it has one reflector and several directors which are either of equal length or decreasing slightly away from the driven element. It is a compact and relatively broadband antenna most suited for TV reception in the VHF range.

13.9 RESONANT AND NON-RESONANT ANTENNA

A resonant antenna is a transmission line whose length is an exact multiple of wavelength and is open at both ends. A dipole antenna mentioned earlier is a good example of a resonant antenna. Figure 13.14 shows three lengths for a resonant antenna which are integral multiples of half-wave lengths. Current distributions along the length of the radiator are also shown. The radiation patterns of the resonant antenna are also shown below.



(a) Current distribution on resonant dipoles



(b) Random pattern of resonant dipoles

Fig. 13.14

The radiation pattern depends on the length of the antenna. In the case of $\frac{\lambda}{2}$ dipole, the radiation is maximum at right angles to it and eventually falls to zero in line with the antenna. The radiation pattern is a figure of eight with its axes at right angles to the dipole.

When the length of the dipole increases, to λ, the polarity of current in one half of the antenna is opposite to that on the other half. As a result, the radiation at right angles to the antenna will be zero because the field due to one half fully cancels the field due to the other half. The radiation pattern develops four lobes, the max. of the lobe being at 54° to the antenna. As the length of

the antenna increases to $\frac{3\lambda}{2}$, the current distribution changes and the radiation pattern takes the shape shown in Fig. 13.14c. The number of the lobes in the radiation pattern goes on increasing with the length of the antenna but the angle of the largest lobe with the directions of antenna decreases. This property is made use of in forming antenna arrays as will be explained later.

13.9.1 Non-Resonant Antennas

A non-resonant antenna corresponds to a non-resonant transmission line which is correctly determined and on which only forward travelling wave exists and there are no standing waves. The radiation pattern though similar to the resonant antenna, is unidirectional. In fact there are half the number of lobes compared to the resonant antenna. This is due to the absence of the reflected wave, which otherwise combines vertically with the forward wave to create the radiation pattern.

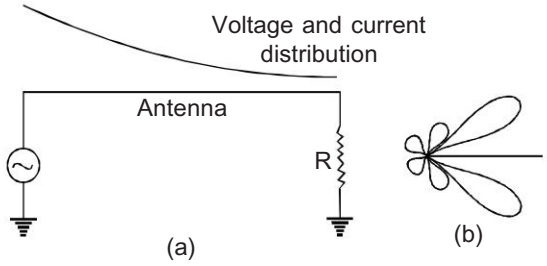


Fig. 13.15 Non-resonant Antenna (a) Layout and Current Distribution, (b) Typical Radiation Pattern

13.10 THE RHOMBIC ANTENNA

A very widely used non-resonant antenna array is the rhombic antenna shown in Fig. 13.16.

This consists of non-resonant elements as arranged differently from other antenna array. It is a planar rhombic which may be thought of as a piece of parallel wire transmission line pinched-out in the middle. The four legs are

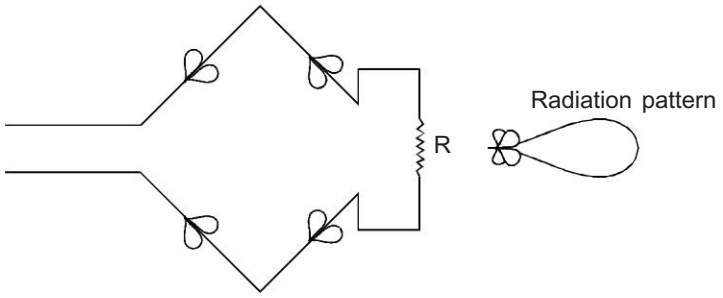


Fig. 13.16 Rhombic Antenna with its Radiation Pattern

considered as non-resonant antennas. This is achieved by treating the two sets as a transmission line correctly terminated in its characteristic impedance at the far end so that only forward waves are present. The radiation pattern is unidirectional as shown in Fig. 13.16.

Since the termination absorbs some power, the rhombic antenna must be terminated by a resistor which for transmission, is capable of absorbing about 'one-third' of the power fed into the antenna. The terminating resistance is about 900 Ω and the input impedance varies from 650 to 700 Ω.

A rhombic antenna is a multiband antenna capable of operating satisfactorily over most of 3-30 MHz range either for reception or for transmission. It is a widely used antenna array especially for point to point working in the HF range.

13.11 TELEVISION TRANSMITTING ANTENNAS

TV signals are transmitted by space wave propagation and so the height of the antenna must be as high as possible in order to increase the line-of-sight distance. Further, a TV transmitting antenna is omni-directional (radiating equally in all directions) and to obtain an omnidirectional pattern in the horizontal plane an arrangement known as turnstiles array is often used.

13.11.1 Turnstile Array

In this type of antenna two crossed dipoles are used in a turnstile arrangement as shown in Fig. 13.17(a). These dipoles are fed in quadrature i.e. the currents fed to them are 90° out of phase by using an extra $\frac{\lambda}{4}$ length in the feeder line of one. Each dipole has a figure of eight pattern in the horizontal plane but crossed with each other as shown in Fig. 13.17(b). The resultant energy of the two dipoles is the vector sum of the two fields from each dipole 90° apart and gives a constant vector sum in all directions. The turnstiles are stacked one above the other for vertical directivity as shown in Fig. 13.17(c).

13.11.2 Dipole Panel Antenna System

Another antenna system that is often used for band I and band III transmission consist of dipole panels mounted on the four sides at the top of the antenna tower as shown in Fig. 13.18.

Each panel consists of an array of full wave dipoles mounted in front of reflectors. For obtaining unidirectional pattern, the four panels mounted on the four sides of the tower are so fed that the current in each lags behind the previous by 90°. This is achieved by varying the field cable length by $\frac{\lambda}{4}$ to the two alternative panels and by reversing of the polarity of the current.

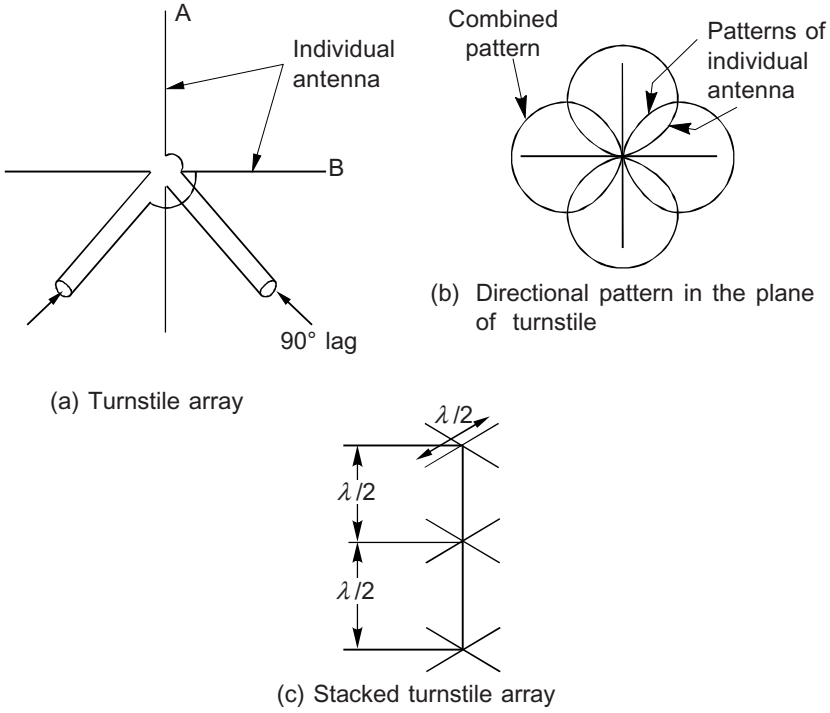


Fig. 13.17

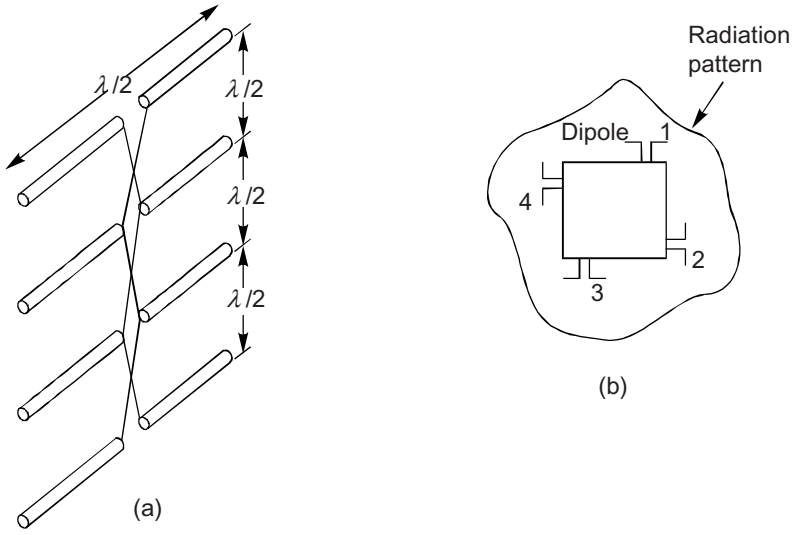


Fig. 13.18 (a) Panel of Dipoles, (b) Radiation Pattern of Four Tower Mounted Dipole Antenna Panels

13.12 FOLDED DIPOLE

An ordinary $\frac{\lambda}{2}$ dipole antenna has an impedance of about 72 Ω at the feed point. This impedance increases to four times its value (72 × 4 = 288 Ω) by converting it into a folded antenna rod construction in which the half wave dipole is joined at the ends by a continuous rod of the same length near and parallel to it as shown in Fig. 13.19(b). If the diameter of the two rods is equal, the current flowing in the folded dipole becomes one half and the impedance increases to four times as shown below.

$$P = Z_0 \times I^2 \tag{13.2}$$

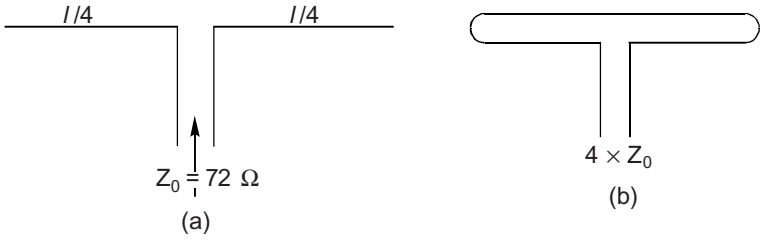


Fig. 13.19 (a) Dipole Antenna, (b) Folded Dipole

where P = Power fed into the dipole, Z_o the impedance at the feed point and I the current flowing into the antenna. If the current is halved on the length being doubled, and power input remaining the same, the new input impedance Z_1, is given by the relation

$$P = Z_1 \times \left(\frac{I}{2}\right)^2 \tag{13.3}$$

$$\therefore \quad Z_1 \cdot \left(\frac{I}{2}\right)^2 = Z_0 \cdot I^2$$

$$Z_1 \cdot \frac{I^2}{4} = Z_0 \cdot I^2$$

or $Z_1 = 4\ Z_0$

If $Z_0 = 72\ \Omega$ as stated earlier

$Z_1 = 4 \times 72 = 288 \approx 300\ \Omega$

If the feeder line used also has an impedance of 300 Ω as in the case of a twin parallel wire feeder, the impedance matching becomes easier with a folded dipole.

Moreover, the reaction of the folded dipole is lowered due to the paralleling action of the fold and the Q of the combined length is lowered and this gives it a larger bandwidth.

As a wideband antenna it is quite useful as a receiving antenna for TV signals consisting of both the picture signal and the sound signal.

Parasitic Elements The directivity of a dipole antenna can be increased by placing additional rods or elements on either side of the dipole and parallel to it. These rods or elements that have no electrical contact with the dipole are called parasitic elements. Depending on their length and distance from the dipole, these can act as reflector or director for the signal being received.

Reflector This is a parasitic element which is about 5% longer than the dipole and placed at a distance of 0.25 λ behind it. Being longer than the driver and close to it, reduces signal strength in its own direction and increases it in the opposite direction. This in effect amounts to reflection of energy towards the driver element and is thus called a reflector.

Director This is a parasitic element shorter than the dipole by about 4 to 5% and placed in front of the dipole at a distance of less than 0.25 λ. This element receives energy by induction from the driver dipole and tends to increase radiation in its own direction and is, therefore, called a director. The number of directors and its length can be varied to obtain increased directivity and broadband response.

Receiving Antennas A receiving antenna intercepts the radio waves radiated by the transmitting antenna and absorbs energy from the electromagnetic waves travelling past the former. The currents induced in the receiving antenna have the same waveshape as the radio wave striking it and these currents are transferred to the radio receiver by suitable transmission lines for detection and conversion into sound waves. As a general rule a good radiator is a good absorber also. The receiving antennas have the same properties and follow, more or less, the same design considerations as the transmitting antennas except for the magnitudes of the currents and voltages involved in the two cases.

A resonant half-wave antenna seems to be the obvious choice for a receiving antenna whose purpose is to extract the maximum power from the radio waves. Such half-wave antennas are actually used at higher frequencies where length of the antenna is within practical limits. However, for reception in the broadcast band 535-1605 kHz, where the length of a resonant antenna is very large, resonant antennas are not used in receivers but a built-in loop of wire is used because the high-powered transmitters used in the medium-wave band provide a sufficiently strong signal within the service area. It is only in the case of areas far removed from the transmitter and known as *fringe* areas that a long length of wire suitably strung between two points on the top of the building is used for reception as shown in Fig. 13.20.

13.13 TELEVISION RECEIVING ANTENNAS

An antenna most suitable for the reception of TV signals is a resonant horizontal dipole which can be conveniently constructed because of the VHF and UHF band of frequencies used for TV broadcasts. A TV receiving antenna is a directional antenna which is not designed for any particular frequency but is meant to cover the entire band of TV frequencies containing both the picture

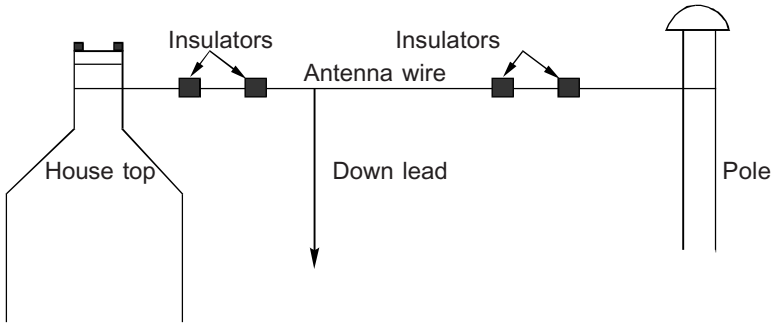


Fig. 13.20 Receiving Antenna—inverted *L* Type

and sound carrier frequencies. The length of the half-wave antenna is a geometric mean between the length of the highest frequency and the lowest frequency in the regular TV band.

Impedance matching in the case of TV antennas is very important for maximum transfer of energy from the antenna to the TV receiver. A resonant dipole has an impedance of about 73 Ω at the centre and can be directly matched to a coaxial cable transmission line which has practically the same impedance. In practice, however, the dipole used in a TV antenna is a folded dipole which has an impedance equal to four times the impedance of an ordinary dipole and can be conveniently matched to a 300 Ω parallel-wire transmission line.

Besides the folded dipole which is the *driven element* a TV antenna makes use of two *parasitic* conductors to increase its directivity. These are the reflector and the director. The reflector is a rod about 5% longer than the antenna and is mounted behind the folded dipole at a distance of about $\lambda/4$ from it. The signals coming from the front side of the dipole strike the antenna first and then the reflector. Currents are induced in both. The reflector reradiates these currents so as to reinforce the currents induced in the folded dipole. The received signal in the antenna is stronger than if there had been no reflector. The signals coming from the back of the antenna strike the reflector first and then the antenna itself. The signals reradiated by the parasitic reflector in this case are out of phase with the signals striking the folded dipole from behind and the two mutually cancel each other. The effect of placing the reflector behind the TV antenna is to give it a clear directivity pattern.

The director consists of a straight rod mounted in front of the folded dipole. The director is about 4% shorter than the dipole and is mounted about 0.1 wavelength in front of the dipole. The action of the director is similar to that of the reflector and the entire antenna array gets a highly directive reception pattern toward the direction of the transmitting station. More than one director are sometimes used to increase the directivity of the TV antenna. The use of reflectors and directors in a TV antenna not only increases its directivity but also its gain and sensitivity. Figure 13.21 shows a modern TV antenna which is a slightly modified form of Yagi-Uda antenna described in Fig. 13.21.

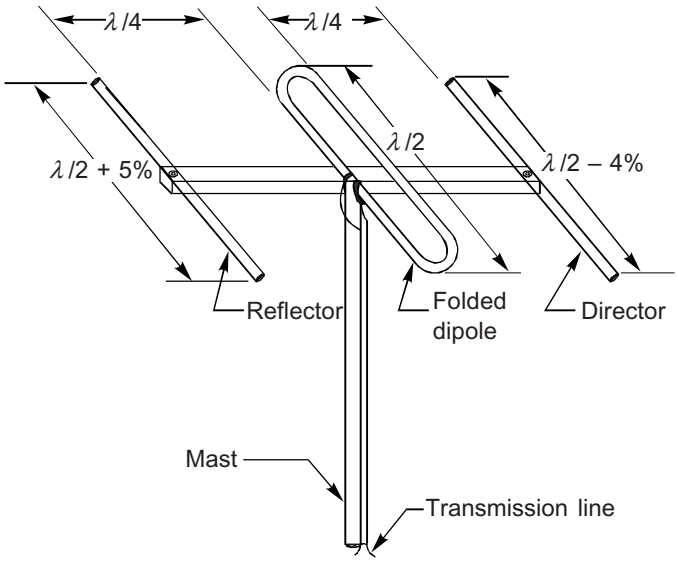


Fig. 13.21 A Television Antenna

SUMMARY

RF energy from the tank circuit of a transmitter is fed to the antenna by means of connecting wires called transmission lines. A transmission line has to be terminated by its characteristic impedance which depends on its distributed capacitance and inductance and is given by the formula $Z_0 = \sqrt{\frac{L}{C}}$. Two wire transmission lines have a characteristic impedance equal to 500 to 600 Ω whereas the coaxial transmission lines have a transmission impedance of about 75 ohms.

The function of a transmitting antenna is to convert the RF energy into radiowaves which travel in space with the velocity of light. An antenna forms a resonant circuit at the frequency of transmission for maximum conversion of RF energy into radiowaves. Two types of commonly used antenna are the Hertz antenna or the dipole antenna and the vertical quarter-wave antenna called the Marconi antenna with one end earthed.

Two or more halfwave antenna stacked horizontally or vertically and suitably spaced form an antenna array which is used in broadcasting to form radio beams which can be reflected from the ionosphere to cover certain target areas.

A non-resonant antenna is a correctly terminated transmission line which is without standing waves. Rhombic antenna is a good example of a non-resonant antenna.

A television transmitting antenna is an omnidirectional antenna which is formed by dipole elements stacked suitably to give a omnidirectional radiation pattern.

A folded dipole is formed by joining the ends of a half-wave dipole with a continuous rod of the same length. It has an impedance of $75 \times 4 = 300$ ohms. It forms a useful part of a TV receiving antenna.

A receiving antenna intercepts radio waves and converts them into currents at the same frequency. A half-wave dipole is mostly used for short-wave reception. For television receiving antenna a Yagi-Uda antenna array is most commonly used.

Review Questions

1. What is a transmission line? Describe two types of transmission lines commonly used with radio transmitters.
2. What is meant by the characteristic impedance of a transmission line? Why is it necessary to terminate a transmission line with an impedance equal to its characteristic impedance?
3. What is the function of an antenna? How does an antenna behave like a tuned circuit?
4. Calculate the length of each limb of a half-wave antenna suitable for radiating at a frequency of 1.5 MHz.
5. What is the difference between a Marconi antenna and a Hertz antenna? Which of them will be suitable for broadcast of medium wave frequencies?
6. What is an antenna array? Explain the difference between a broadside array and an end-fire array. Show their directional patterns.
7. Describe a Yagi-Uda antenna array. Why is this antenna suitable for the reception of TV signals?
8. Explain the difference between a resonant and a non-resonant antenna. Give examples.
9. What is the main difference between a TV receiving antenna and a TV transmitting antenna? Describe a turnstile antenna array and give its directional radiation pattern.
10. Describe a folded dipole and give its practical applications.
11. Describe the action of a director and reflector in a TV receiving antenna.
12. Fill up the blanks with words selected from the list given at the end of the question:
    (1) A good radiator is also a good _____ of electromagnetic waves.
    (2) A picture carrier in a TV signal is _____ modulated and the sound carrier is _____ modulated.
    (3) A non-resonant transmission line is one that is terminated with its _____ impedance.
    (4) In a TV receiving antenna the _____ is placed in front and the _____ is placed behind the folded dipole.
    (5) The process of _____ is the reverse of modulation.
    (6) Frequencies in the _____ range are suitable for only line-of-sight communication.

    List of words: Director, reflector, absorber, amplitude, frequency, detection, VHF, characteristic.

chapter 14

Radio Receivers

14.1 FUNCTIONS OF A RADIO RECEIVER

A radio receiver is required to perform three main functions. These are:

1. Selection of the desired frequency from a large number of modulated carrier frequencies that strike the receiving antenna at one and the same time.
2. Separation of the modulating audio frequencies from the modulated carrier frequency by the process of *detection* or *demodulation*.
3. Conversion of AF currents into sound waves that can be easily heard by the human ear.

In addition to the fundamental requirements mentioned above, a radio receiver also contains circuits for RF amplification, frequency conversion and AF amplification. These circuits are meant to improve the performance of a radio receiver and to get sufficient output power for driving a loudspeaker. Based on these considerations, the main functions of a radio receiver can be indicated by means of a block diagram of the type shown in Fig. 14.1. The block diagram

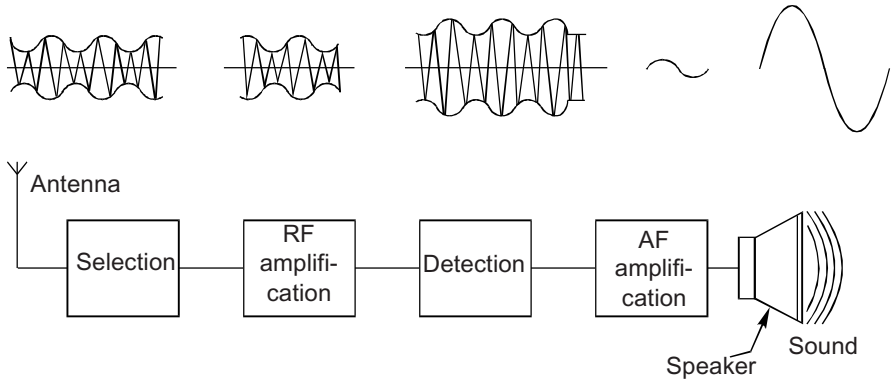


Fig. 14.1 Block Diagram of a Simple Receiver

also indicates the waveform of the radio waves as they pass through different stages of the radio receivers.

14.2 AM AND FM RECEIVERS

Two methods of modulation used in radio broadcasts are amplitude modulation (AM) and frequency modulation (FM). A receiver meant to receive amplitude modulated waves is known as an AM receiver and a receiver designed for the reception of frequency modulated waves is called an FM receiver. The circuits used in these two types of receivers are not similar. In India, the radio stations use mainly amplitude modulation for their radio broadcasts. As such, all domestic broadcast receivers in use are of the AM type. Only AM receivers will be described in this chapter.

Both AM and FM are used by TV stations. Whereas the picture signals from a TV transmitter are amplitude modulated, the sound signals are frequency modulated. FM receivers, therefore, form a part of TV receivers and will be described in Chapter 19 on TV receivers.

14.3 CHARACTERISTICS OF A RECEIVER

There are three main characteristics by which the quality of a receiver can be judged. These are *selectivity, sensitivity* and *fidelity*. These are also known as the performance characteristics or the specifications of a receiver.

Selectivity Selectivity of a receiver is its ability to select a desired signal frequency without any objectionable interference from other neighbouring stations. It is a measure of the extent to which the receiver can reject all other neighbouring stations and accept only the desired station. A good, selective receiver will select the desired station and reject all other unwanted stations. The selectivity is generally expressed in the form of a curve shown in Fig. 14.2. In this curve, the strength of the input signal at the resonant frequency required to produce a given output is taken as the reference and the strength of the modulated carrier at neighbouring frequencies required to produce the same output is plotted on the vertical axis. The sharper the selectivity curve, the more selective a receiver is.

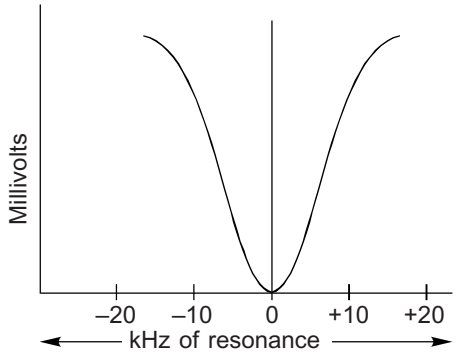


Fig. 14.2 Selectivity Curve

The selectivity of a receiver increases with the number of tuned RF stages in the circuit. A very selective receiver allows only a limited band of frequencies to pass through. As such, many of the side bands do not appear in the output and the quality of reproduction of the receiver suffers.

Sensitivity Sensitivity of a receiver is its ability to respond to weak signals. This is expressed as the minimum voltage or power that must be applied to the input of the receiver for getting a standard output of 0.5 W in the loudspeaker or in a resistive load substituted for the loudspeaker. For measuring sensitivity, the input RF signal must be modulated 30% at an AF of 400 Hz. Moreover, the RF signal voltage is applied through an artificial antenna called a dummy antenna. This dummy antenna creates the same input conditions for a receiver as a real antenna installed on the roof top. It can be made by connecting an inductance (coil) of 20 μH in series with a capacitance of 200 μμf.

The sensitivity of a receiver is expressed in microvolts. The smaller the input in microvolts the greater is the sensitivity of a receiver. A high grade broadcast receiver will have a sensitivity of less than 10 μV. The sensitivity curve for a standard broadcast receiver is shown in Fig. 14.3.

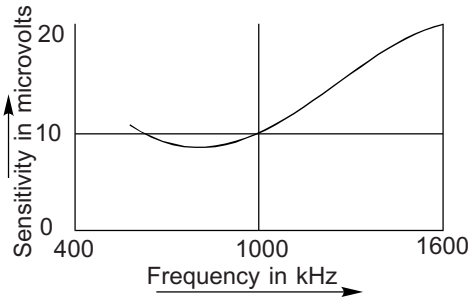


Fig. 14.3 Sensitivity Curve for a Standard Broadcast Receiver

Fidelity Fidelity is the ability of a receiver to reproduce faithfully all the audio frequencies with which the carrier is modulated. This is generally expressed as a frequency response curve shown in Fig. 14.4.

The fidelity of a broadcast receiver mainly depends on the response of the audio stage but the tuned RF stages also limit the response of the receiver by

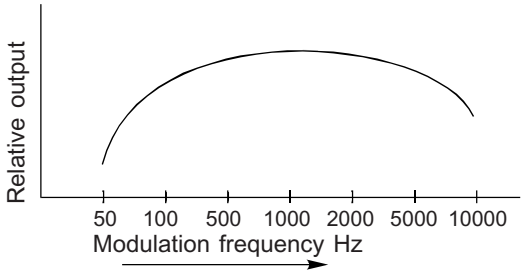


Fig. 14.4 Fidelity Curve

not allowing the higher sidebands to be reproduced properly. Selectivity and fidelity in a receiver are opposed to each other. Modern broadcast receivers are capable of reproducing properly all modulation frequencies from 30 Hz to about 8 kHz. However, the international regulations do not allow a bandwidth of more than 5 or 6 kHz for broadcast purposes.

The noise level of a receiver is another important characteristic particularly at shortwave bands. Noise produced in the receiver by its own circuits should be sufficiently low so that even the weakest signals received can suppress this noise.

14.4 TYPES OF RECEIVERS

The basic requirements of selection, detection and sound reproduction for a receiver can be met by a tuned *LC* circuit for selection, a diode (or a crystal) for detection, and a pair of headphones for sound reproduction. This was the earliest and the simplest form of a receiver and was called a *straight receiver*. It did not require a power supply and provided no amplification for the radio frequencies or the detected audio frequencies. It could be used for listening to only the local stations. The crystal set described earlier is a form of straight receiver.

TRF Receiver The invention of the vacuum tubes or radio valves made it possible to amplify the detected audio frequencies to drive a loudspeaker. With the development of RF amplification techniques, it was possible to add one or two stages of RF amplification before detection so that a sufficiently strong RF signal could be delivered for detection by the diode. Thus, with RF amplification before detection and AF amplification after detection, a new type of receiver known as the TRF or *tuned radio frequency* receiver came into existence. A TRF was the first practical type of receiver to be designed and constructed. A block diagram of a TRF receiver together with the waveform of the signal at different stages in the receiver is given in Fig. 14.5.

A TRF receiver tends to be selective and its fidelity is not so good. Also, the sensitivity of a TRF receiver varies with the received frequency. In spite of

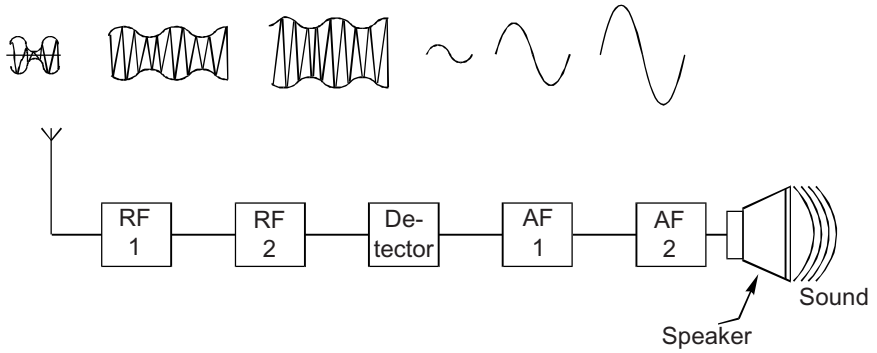


Fig. 14.5 Block Diagram of a TRF Receiver

this, a TRF receiver still finds good use in many special applications in the field of radio communication.

However, the TRF receiver has been largely replaced by the modern superheterodyne receiver which possesses many advantages over all other types of radio receivers.

14.5 SUPERHETERODYNE RECEIVER

A superheterodyne receiver is the most popular type of radio receiver. Almost all modern radio receivers are of the superheterodyne type popularly known as "superhet." In order to get a clear picture of the working of a superheterodyne receiver, it is best to refer to Fig. 14.6. Each block in this diagram represents a definite stage in the working of a superhet. The function of each stage will be described here briefly. However, the various stages will also be discussed in detail later in the chapter. A picture of the waveform of the signal as it enters and emerges from each stage is also shown in the block diagram. A brief description of the functions of the various stages is given below.

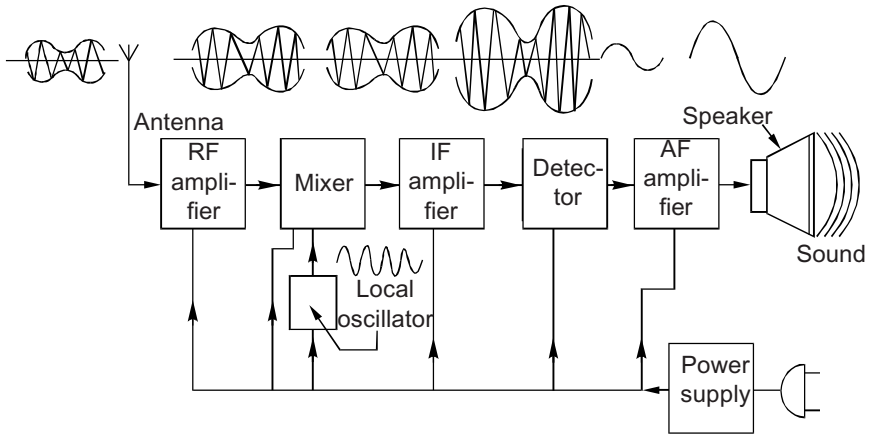


Fig. 14.6 Block Diagram of a Superheterodyne Receiver

The antenna picks up the modulated carrier signals of all the stations which are within the receiving range of the radio receiver. A tuned circuit in the RF stage selects only the desired modulated carrier signal and rejects all other signals.

The selected signal is amplified by the RF amplifier stage before it is applied to the mixer stage. Most broadcast receivers, however, do not have this RF amplifier stage.

In the mixer stage the incoming signal frequency is converted into a new fixed frequency by the mixing or beating of this frequency with an unmodulated frequency produced by a local oscillator. This new fixed frequency, which is the difference of the frequency produced by the local oscillator and the incoming modulated signal frequency, is called the *intermediate frequency* or IF. Its value is fixed at 455 kHz for almost all broadcast receivers. Thus the frequency generated by the local oscillator is always 455 kHz above the frequency of the

station selected by the RF stage. In fact, the main advantage of a superheterodyne receiver lies in having a fixed IF. It is much easier to design RF amplifiers for a fixed IF than to design RF amplifiers for amplification of a whole band of frequencies as in the case of a TRF receiver.

The IF of 455 kHz which contains all the modulation of the carrier, is suitably amplified by a fixed-tuned amplifier called the IF amplifier. It is then fed on to the detector stage.

The detector stage separates the audio component from the modulated IF which is bypassed to ground. The audio frequencies so recovered are applied to the audio stage for amplification. This detector is also known as the *second* detector, the first detector being the mixer stage.

The audio stage generally consists of two audio amplifiers. The first amplifier is a voltage amplifier and the second stage is a power amplifier, which develops sufficient power to drive a loudspeaker. The loudspeaker converts the electrical audio frequencies into sound waves of corresponding frequencies.

Voltages required for the operation of the various stages described above are obtained from a built-in unit called the power supply source. This is generally obtained by rectification and suitable filtering of the mains 230-V 50 Hz AC supply. In the case of battery-operated sets, the supply is obtained directly from the batteries for which no additional filtering is required.

14.5.1 Circuit Details

Before describing the circuit details of the various stages of a superheterodyne receiver, it is necessary to mention here that transistors are rapidly taking the place of valves or vacuum tubes in all types of electronic circuits. Therefore, present day radio receivers can be divided into two separate classes or categories—those using vacuum tubes or valves and known as valve radios and those using transistors in place of valves and popularly known as transistor radios or merely transistors. Circuit details are similar in both types of receivers but the main difference is in the size and specifications of the components used. Because of the much smaller magnitude of voltage and current involved in the case of transistor radios compared to valve type radios, the use of miniature components has been possible in the construction of transistor radios. This has resulted in considerable economy in the size and cost of transistor radios and reasonably priced pocket size transistor radios are becoming more popular each day.

14.5.2 Antenna Circuits

Modulated radio waves from different stations strike the wire of the receiving antenna and produce in this wire RF currents of the same frequency as the frequency of the arriving waves. These RF currents flowing through the primary L_1 of the transformer T_1 induce similar currents in the secondary L_2 of the transformer T_1. The secondary coil L_2 can be tuned to resonance at the desired frequency by the variable capacitor C_1 (Fig. 14.7(c)).

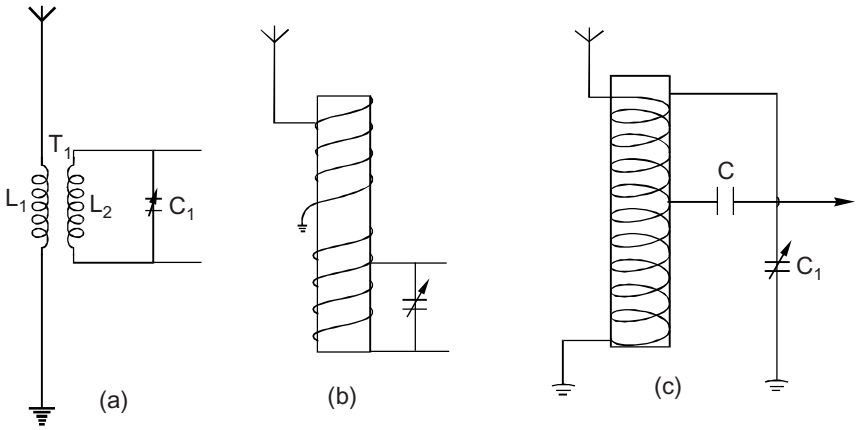


Fig. 14.7 Antenna Coils (a) Tuned Antenna Circuit (b) Primary and Secondary of the Transformer Wound on a Ferrite Rod (c) Antenna Coil as an Auto-transformer

The resonant voltages so developed are fed to the input of the RF amplifier.

To improve the efficiency of the transformer T_1, the primary and secondary of this transformer are wound on a ferrite rod made of a magnetic material of high permeability as shown in Fig. 14.7(b). The ferrite rod is also known as the *loopstick*. To further simplify matters, a single coil is wound on a form and tappings provided at different points for tapping the secondary voltage. A ferrite rod is then slid into the form and its position fixed to provide a good pick up. This arrangement works as an autotransformer and provides a satisfactory antenna coil for broadcast frequencies. The arrangement is shown in Fig. 14.7(c).

Actually the antenna itself possesses some stray inductance and capacitance and its dimensions can be so chosen as to make it resonate at a particular frequency. However, in present-day broadcast receivers the antenna is designed broadly to resonate at the mid-frequency of the broadcast band. In fact, with the strong signals received from most broadcast stations and the excellent sensitivity provided by the modern heterodyne receivers, an external antenna is not required particularly in the case of portable transistor receivers.

14.5.3 Radio Frequency Amplifier (RF Stage)

The desired RF signal selected by the tuned antenna circuits is applied to the input of a radio frequency amplifier which is also called an RF stage. After amplification by the RF stage, the modulated RF signal is passed on to the next stage which may be another RF stage or the converter (mixer) stage. The RF stage provides the following advantages:

1. It increases the sensitivity of the receiver because of additional amplification.
2. It improves the selectivity of the receiver due to additional tuned circuits.
3. It eliminates the image frequency interference.

4. It reduces the noise level because of a stronger signal fed to the convertor stage.
5. It improves the AVC action because another controlled stage is added to the RF chain. Broadcast receivers do not have an RF stage.

14.5.4 Special Features of a Superheterodyne Radio Receiver

Besides the three characteristics of a radio α receiver viz. selectivity, sensitivity and fidelity discussed earlier, the superheterodyne receiver possess certain features which are peculiar to this type of receiver only. These features will now be discussed.

(1) Image Frequency In a superheterodyne receiver, the intermediate frequency (*IF*) is the difference between the local oscillator frequency (f_0) and the frequency of the received signal (fs_1) or

$$f_0 - fs_1 = IF \tag{14.1}$$

If there is another signal fs_2, whose frequency is higher than the local oscillator frequency f_0, by amount equal to the intermittent frequency *IF*, and if this frequency fs_2 also manages to reach the converter stage, it will produce the same *IF* as fs_1 and will be equally well received by the *IF* stage or

$$fs_2 - f_0 = IF \tag{14.2}$$

adding (14.1) and (14.2)

$$fs_2 - fs_1 = 2IF$$

or

$$fs_2 = fs_1 + 2IF$$

This signal fs_2 where frequency is higher than the frequency of the signal fs_1, by an amount equal to twice the intermediate frequency *IF* is called the *image frequency*. Thus in an AM receiver with an *IF* of 455 kHz, a signal at 1000 kHz will be simultaneously received with its image frequency of 1000 kHz + 2 × 455 kHz = 1910 kHz. This is not desirable and steps must be taken to reject this image frequency by suitably designing the RF stages of the receiver.

Image Frequency Rejection The most effective method of image frequency rejection is to introduce a tuned RF stage between the mixer and the antenna stage of the receiver and this stage will favour the desired signal and discriminate against the undesired image signal. Both these factors will improve image rejection ratio ρ as is clear from the following expression

$$\rho = \frac{\text{gain at the desired signal } (fs)}{\text{gain at the image frequency } (fs_2)}$$

A single tuned circuit will increase the value of the numerator and decrease the value of the denominator thereby effectively increasing the value of ρ. In fact, the suppression or rejection of the unwanted image signal will be more effective, the greater the number of tuned circuits involved before the mixer stage.

Another method of improving the image rejection ratio is to increase the ratio of the intermediate frequency and the signal frequency by selecting a higher value of *IF* so that the ratio $\frac{fs_2}{fs}\left(\frac{fs + 2f_1}{fs}\right)$ becomes large. In the case of broadcast receivers which operate in the frequency range 535 to 1605 kHz range, the *IF* value is generally 455 kHz and value of fs_2/fs is large and the use of tuned RF stages is not necessary in the case of AM broadcast receiver for image frequency rejection. However the use of higher *IF* value becomes necessary for frequencies above 3 MHz for improved image frequency rejection. FM receivers which normally operate in the 88 to 108 MHz range, *IF* value employed is 10.7 MHz. In the case of TV receivers operating in the VHF and UHF range the *IF* used is 38.9 MHz.

Image rejection depends on the front-end selectivity of the receiver and must be applied before the *IF* stage. Once the spurious frequency reaches the *IF* stage, it becomes difficult to remove it from the wanted signal.

(2) Double Spotting It is a phenomenon that occurs in the HF or short-wave ranges of a superhet radio receiver. In double spotting, the same station is heard on two different points on the tuning dial of the receiver. This phenomenon is similar to image reception, the difference being that in double spotting the same signal f_s is heard at two settings of the tuning dial but in image reception two different signals f_s and f_{si} are heard on one and the same setting of the tuning dial. The phenomenon can be explained as follows.

When a signal f_s is received, the local oscillator is tuned to a frequency $f_s + f_i$ or when local oscillator frequency is reduced, the same signal will be received at a frequency given by $f_s - f_2$. The frequency difference between the two tuning position on the dial will be $(f_s + f_i) - (f_s - f_2) = 2f_i$ indicating that the second position of the same signal is that of the image frequency. The signal in the second portion will be reduced in strength depending upon the rejection ratio of the receiver p.

As a numerical example we may consider a strong signal received at 14.5 MHz, the local oscillator will be tuned to 14.5 + 455 = 15.145 MHz. The same signal will again be received at 13.590 MHz when the local oscillator frequency will be 14.045 MHz. This is exactly 455 kHz below the frequency of the original signal i.e. 14.50 kHz. The two signals will beat to produce an IF of 455 kHz and neither will be rejected by the IF amplifier. RF stage tuned to 13.590 MHz will reject the strong signal at 14.5 MHz and double spotting will be avoided.

Since double spotting occurs at higher frequencies in receivers having poor image rejection ratio, one solution to both the problems i.e. image rejection and double spotting is to improve the sharpness or the selectivity of the RF amplifier or to increase the value of IF or a combination of both.

Double spotting is not desirable in a receiver because a weak station may be masked by the reception of a nearby strong station at the spurious point on the dial.

It is interesting to note that the spurious point being exactly $2f$ below the correct frequency, the phenomenon of double spotting can be used to measure the intermediate frequency (IF) of an unknown receiver.

14.5.5 Tracking and Alignment in a Superheterodyne Receiver

A superheterodyne receiver has a number of tuned circuits which must be correctly tuned to their respective tuning frequencies if a station is to be received properly. For ease of operation, the various tuned circuits are mechanically coupled so that only one tuning control and dial are required for the tuning of the receiver. This is called single dial tuning.

Satisfactory signal dial tuning of a receiver requires that all resonant circuits in the radio frequency section of the receiver be tuned together and that their resonant frequencies correspond to a predetermined standard printed scale. In turn this means that no matter what the received frequency, the RF and mixer input circuit must be tuned to it. Thus when capacitive tuning is employed, exact alignment of the radio frequency circuit is obtained at the high frequency end of the tuning range by the use of an adjustable tuned capacitor which is in parallel with the coil as shown in Fig 14.8(a).

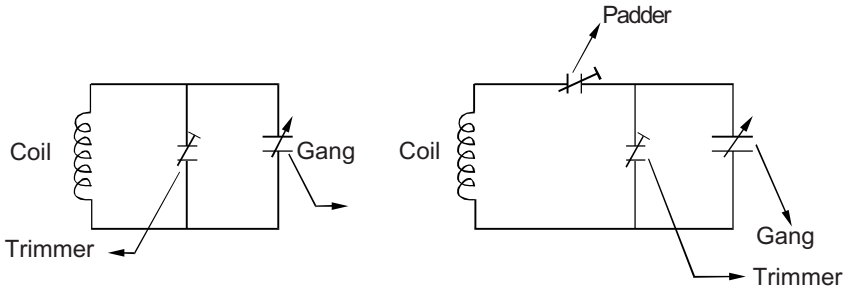


Fig. 14.8 (a) R.F. Circuit (b) Oscillator Circuit

The local oscillator must simultaneously be tuned to a precisely higher frequency than the frequency of RF section by an amount equal to the intermediate frequency. In capacitor tuning, this is generally achieved by using a gang capacitor in which different sections are made as identical as possible. Tracking is thus obtained by using a coil of somewhat less inductance for the oscillator than for the RF side and also using with the oscillator coil series capacitors called padders and parallel capacitors known as trimmers shown is Fig. 14.8(b). By proper adjustment it is possible to obtain exactly correct tracking at three frequencies in any tuning range. Any errors that exist are called *tracking errors* and can result in an incorrect frequency being fed to the IF stage and the tuned station will not correspond to the correct position on the dial. The three frequencies of correct tracking called crossover frequencies are chosen in the design of the receiver so that one frequency is just above the bottom end of the

band as 600 kHz, the other a little below the top end at 1500 kHz and the third at the geometric mean of the two i.e. $\sqrt{600 \times 900}$ = 950 kHz. The results are shown in Fig. 14.9.

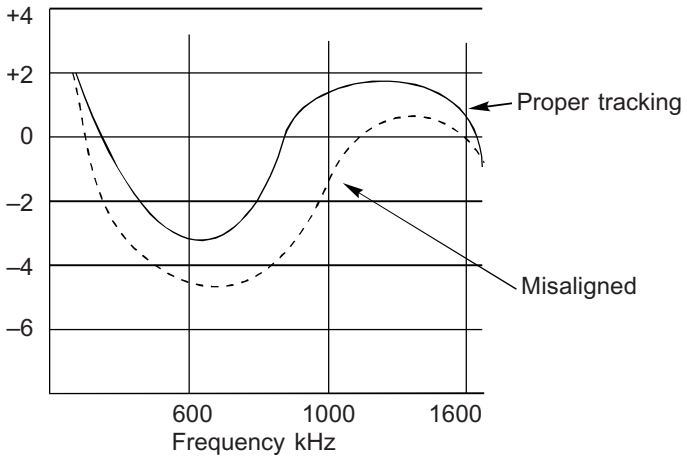


Fig. 14.9

If the maximum tracking error does not exceed 3 kHz with a range, the error at other frequencies will be satisfactory.

14.5.6 Tracking and Local Oscillator Frequency

Broadcast AM receivers both in the broadcast range (540 to 1650 kHz) and the short wave range up to 36 MHz make use of the Hartley and Colpitt type of oscillators which generally operate at a frequency above the signal frequency by an amount equal to the intermediate frequency which is normally 455 kHz in domestic broadcast receivers.

It is possible to design a circuit where the local oscillator frequency is below the signal frequency but this is likely to create practical difficulties in tracking and alignment as explained below.

(a) Local Oscillator Frequency Above the Signal Frequency For a Broadcast receiver operating in the 540 to 1650 kHz range the oscillator frequency at the lower end will be 540 + 455 = 955 kHz and the local oscillator frequency at the higher end of the range will be 1650 + 455 = 2105 kHz. This gives a ratio of max. to min. frequency of the oscillator as $\frac{2105}{995}$ = 2.2.

Normal tunable gang capacitors have a capacitance ratio of 10 : 1 for the max. and min capacitance. The equivalent frequency ratio will be $\sqrt{\frac{10}{1}}$ = 3.2 : 1. Thus the required frequency ratio of 2.2. : 1 is well within the tuning capabilities of the variable capacitor. The entire frequency range can be tuned in one sweep of the gang capacitor.

(b) Local Oscillator Frequency Below the Signal Frequency In this case the lowest oscillator frequency will be 540 – 455 = 85 kHz and the higher frequency equal to 1650 – 455 = 1195 kHz giving a frequency ratio of 14 : 1 and the ratio is much beyond the tuning capabilities of the gang capacitor. It will not be possible to cover the entire frequency range in one sweep.

This is the main reason for keeping the local oscillation frequency higher than the signal frequency in all receivers with variable frequency oscillators.

Another reason for keeping the local oscillation frequency higher than the signal frequency is that the tracking difficulties get minimized if the frequency ratio rather than the frequency difference between the two frequencies is kept constant.

In case (a) above the ratio of local oscillator frequency to signal frequency is $\frac{995}{540} = 1.84$ at the lower end of the band and it is $\frac{2105}{1650} = 1.28$ at the upper end.

In the case of (b) above these ratios would be $\frac{540}{85} = 6.35$ and $\frac{1650}{1195} = 1.38$. The variation is much greater than in case (a) above and will result in much greater problems in tracking.

14.6 MIXER STAGE

The mixer or convertor stage performs the following functions:

1. It tunes and amplifies the desired signal.
2. It generates or produces an unmodulated RF signal of its own at a frequency higher than the frequency of the received signal.
3. It mixes the locally generated oscillator frequency with the received signal to produce a new fixed frequency called the intermediate frequency.
4. It maintains the same constant frequency difference between the local oscillator frequency and the signal frequency to which the receiver is tuned. This difference is the IF.

14.7 IF AMPLIFIER

The IF amplifier is an RF amplifier whose function is to tune and amplify the voltages at IF. The input voltage for the IF amplifier is received from the IFT_1 which is a transformer fixed-tuned to IF and is placed in the base circuit of the IF amplifier transistor. After amplification by the IF amplifier, the IF voltage is again selected by another transformer IFT_2 which is also fixed-tuned to IF and is placed in the collector circuit of the IF amplifier transistor, IFT_2 will supply the input for the next succeeding stage of the superheterodyne receiver. IFT_1 which supplies the input for the IF amplifier is called the input IFT and IFT_2 which selects the output from the IF amplifier stage is called the output IFT. A well-designed IF stage with high gain IFT can provide an overall gain of 100 or so. Two or three IF stages are used when higher gains are required. However,

most modern superheterodyne receivers using transistors use two IF stages to get enough gains.

An IF of 455 kHz has now been standardised for all broadcast type receivers although some older models still use an IF of 175 or 260 kHz. An IF of 455 kHz provides better protection against image frequency interference than an IF of a lower value. 455 kHz is a compromise value between high IF and low IF.

The circuit of a typical IF stage with input and output IFTs is shown in Fig. 14.10.

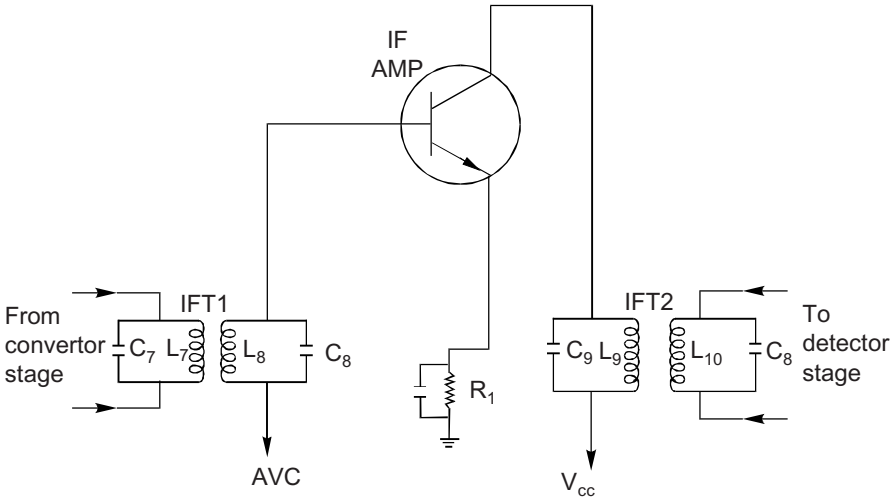


Fig. 14.10 IF Amplifier Stage

14.7.1 IF Transformers

Each IF amplifier makes use of two RF transformers known as IFTs. The input IFT (IFT_1) couples the output circuit of the convertor stage to the input or the base circuit of the IF amplifier and the output IFT (IFT_2) couples the collector circuit of the IF amplifier to the detector stage or to the input circuit of the second IF amplifier if there are more than one IF stages.

An IFT actually consists of two inductively coupled windings which can be independently tuned to the IF of the receiver, which is 455 kHz in most broadcast receivers. The tuning of the coils or windings can be done by means of small variable mica capacitors or air-trimmer capacitors. In this case the IFT is known as a trimmer-tuned IF transformer. In another type, the capacitors are fixed but the tuning is done by means of movable dust iron core slugs. Such IFTs are called permeability tuned or slug tuned IFTs.

14.7.2 IF Amplifiers—Choice of IF Frequency

The choice of IF in a receiver is very important because the IF stage is a vital stage in a superhet receiver. Both the high and low value of IF have their advantages and disadvantages. As such, the IF value is a compromise between a high value and a low value.

High value of IF results in poor selectivity and poor adjacent channel rejection necessitating the use of special sharp cut off filters in the IF stage. It also increases tracking difficulties.

On the other hand a low value of IF results in following difficulties.

1. A poor image rejection ratio as is clear from Equation 14.2.
2. Cutting of the sideband due to sharp selectivity response. This affects the quality of broadcasts.
3. The stability of the local oscillator needs improvement as any drift in local oscillator frequency forms a high percentage of the low value of IF.
4. IF must not fall within the tuning range of the receiver to avoid undesirable heterodyne whistles.

The use of IF for various types of receivers has almost been standardised with slight variations allowed in exceptional cases.

1. All AM receivers operating in the medium, high or even long range (150 to 350 kHz) make use of the standard IF of 455 kHz. Variations are, however, allowed within the range 438 to 465 kHz.
2. Certain special types of AM receivers like single sideband receivers and communication receivers make use of two or more values of IF. The ranges often used are 1.6 to 2.3 MHz or else above 30 MHz.
3. FM receivers operating in the frequency band 88-108 MHz invariably use 10.7 MHz as the IF.
4. Television receivers operating in the VHF (54 to 223 MHz) and UHF (470 to 940 MHz) use IF values varying between 26 to 46 MHz, the exact value depending on the system used. PAL system uses an IF of 38.9 MHz for the picture signal and 33.4 MHz for the sound signal.
5. Microwave and radar receivers operating in the Gigahertz range (1 to 10 GHz) use 30, 60 or 70 MHz as IF depending on the application of the receiver.

14.7.3 Adjacent Channel Selectivity

In Section 14.3, selectivity has been defined as the measure of the extent to which the receiver can reject all other neighbouring stations and accept only the desired stations. The selectivity of the RF stage of a receiver is a function of the frequency and is best at low frequencies and gets worse as the frequency increases. This is shown in Fig. 14.11.

Providing several sharply tuned circuits in the RF stage creates tracking difficulties for these tuned circuits. Therefore, the selectivity of the RF stage is left much wider than necessary for single channel operation and final selectivity is obtained in the design of the IF amplifier stage.

Adjacent channel selectivity is the ability of the receiver to avoid interference from the sideband of the adjoining channels when the frequency spectrum is overcrowded with a large number of channels. For getting the maximum number of channels in the assigned spectrum, the channel distribution is made

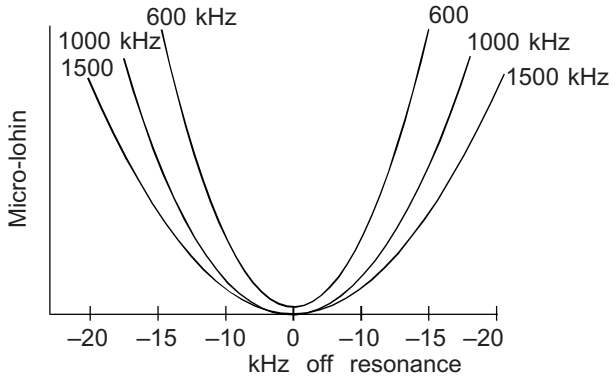


Fig. 14.11 Selectivity

as close together as possible. A channel separation of 10 kHz is typical of AM broadcast stations in the MF and HF band. FM and television stations operating in the higher frequency bands need much wide-spacings between adjacent channels.

It should be possible for two stations to occupy adjacent channels with minimum spacing between them and the receiver should be able to separate them. For this the ideal IF channel band pass characteristics will be a flat topped response within the pass band and sharp vertical cut off outside the pass band providing infinite attenuation or rejection outside the pass band. However, practical filters with the ideal pass band characteristics are not possible and adjacent channel interference to a certain minimum extent is unavoidable. However, good quality broadcast receiver should be able to provide 60 to 80 dB of adjacent channel rejection.

Some of the methods adopted for providing adjacent channel selectivity by the design of the IF stage are given below.

1. The simplest method is to use an IF amplifier consisting of two or more stages with several under coupled IF transformers, all tuned to IF. Each stage provides its own rejection and the total response is the product of the rejections provided by each stage. With suitable selection of Q values, a 3 dB band pass response is possible.

2. Stagger Tuning A much better pass band characteristics than those obtained by a single-tuned circuit can be obtained without undue sacrifice of selectivity by a process known as *stagger* tuning. In this method an odd number of tuned circuits are used and these are so tuned that one circuit is tuned to the circuit frequency and the other two circuits are tuned successively to frequencies

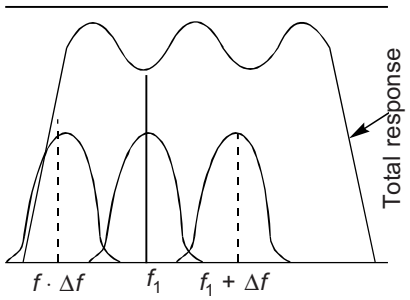
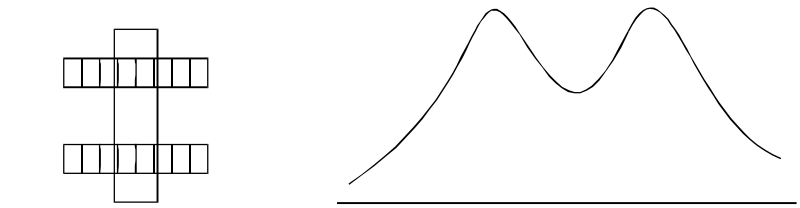


Fig. 14.12

which are above and below the centre frequency by a small amount Δf. Thus if one circuit is tuned to a frequency $f_1 + \Delta f$, the other circuit is tuned to a frequency $f_1 - \Delta f$, f_1 being the centre frequency as shown in Fig. 14.12.

The overall response is the product of the individual response but it has several peaks as shown on the top but this can be smoothened by adding more tuned circuits tuned closer together.

3. Use of Double Tuned Transformers When two circuits are tuned to the same frequency and then tightly coupled together as in IF transformer shown earlier (Fig. 14.13) the overall response has a double hump as shown below.



(a) Overconfined turned IF transformed

(b) Double tuned response

Fig. 14.13

The system has the advantage over the stagger tuned system that isolating amplifier need only be placed between pairs of tuned circuits instead of between individual circuits thus reducing the number of amplifiers required.

IF band pass is more expensive communication receivers and other special types of receivers makes use of IC (Integrated circuit) amplifiers and special type of filters like criptal lattice filter and mechanical resonant filters. However, HF operational amplifiers are now available in IC form, allowing large reduction in the size of the receiver. The entire receiver including the IF amplifiers can now be built with the help of one or two chips easily available in the market.

Permeability tuned IFTs are used in almost all modern broadcast receivers. The entire IFT assembly including the coils and capacitors is enclosed in a metal can which provides both support and shielding for the IFT assembly. Holes are provided in the metal can at suitable places so that the trimmer capacitors or the dust iron core slugs can be adjusted for tuning by passing a screw driver or some other alignment tool through the holes without removing the metal can. A view of the IFT assembly together with a schematic symbol for each of the two types of IFT is shown in Figs. 14.14(a) and (b).

With each IF stage, two IF transformers containing four tuned circuits all tuned to the IF of 455 kHz are added to the receiver chain. This makes the receiver highly selective so that no other frequency except the IF can get through. Since the IF of 455 kHz also contains the sidebands of the audio frequencies with which the original carrier was modulated, for satisfactory

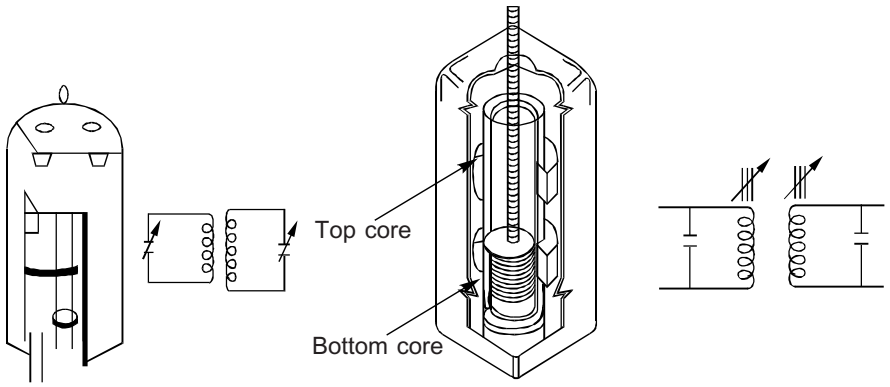


Fig 14.14 IF Transformers (a) IF Transformer–capacitor Tuned (b) IF Transformer–permeability Tuned

reproduction, it is necessary that sideband cutting, which results from a very selective circuit, should be avoided. This is done by broadening the response of the IF amplifier. There are several methods of doing this but the most common method is to adjust the coupling of two windings of the IFT by varying the distance between the windings. The response becomes broader as the distance between the two windings is reduced as shown in Fig. 14.15. The degree of overcoupling is adjusted till the sidebands from 7 to 10 kHz above and below the IF are able to pass through the circuit along with the IF. Such an adjustment is made in the factory and cannot be easily changed afterwards.

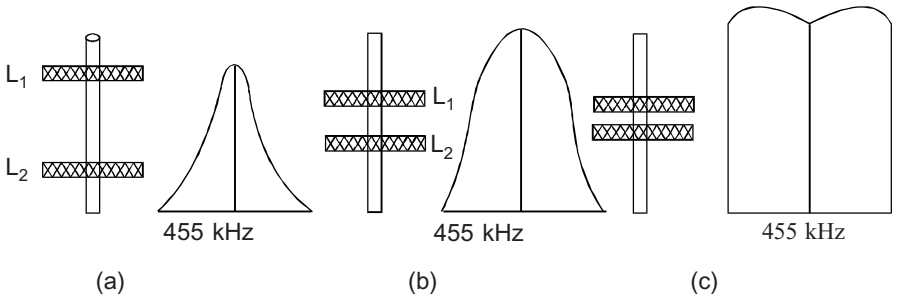


Fig. 14.15 Coupling in an IFT

14.8 DETECTOR STAGE

The output from the IF amplifier consists of the modulated IF at 455 kHz. This IF is modulated with the same audio frequencies with which the original RF signal picked up by the antenna was modulated. The amplified and modulated IF is applied to the input of the *detector stage.* It is the function of the detector stage to separate the audio frequencies from the IF (455 kHz). The audio frequencies so separated are then applied to the audio stage which drives the loudspeaker of the receiver to reproduce the same sound that was converted

into electrical audio frequencies by the microphone at the studios of the broadcasting station. Because of the nature of its function the detector stage is also known as the demodulator. It is also sometimes called the second detector to distinguish it from the mixer stage which was originally given the name of first detector.

Another important function performed by the detector stage is to provide automatic volume control or AVC for the superheterodyne receiver. AVC is the process which helps maintain practically the same volume of sound from the speaker whether the station received by the antenna is a strong local station or a distant weak station. This will be described in detail later.

A detector stage in a modern receiver employs a diode in an half wave rectifying circuit similar to that employed in a power supply source.

To draw an exact parallel between the half wave rectifier and the detector stage of a receiver, the circuit of a simple detector stage with associated filter circuits is shown in Fig. 14.16.

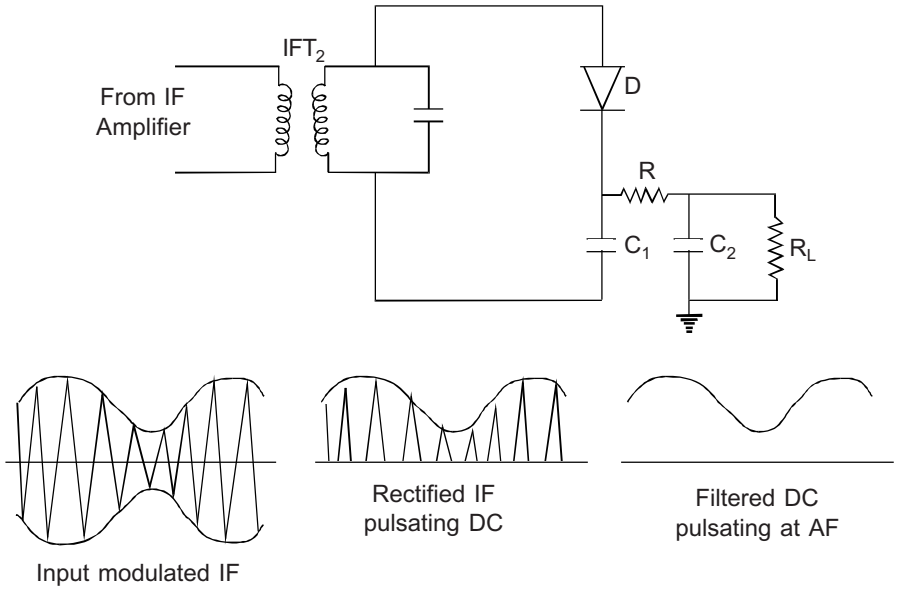


Fig. 14.16 Detector Stage

In this case the input for the diode is the modulated IF at 455 kHz in place of the 50 Hz mains supply. The main difference, however, lies in the filter section. In the filter section the resistance R replaces the choke and the capacitors C_1 and C_2, have a much lower value (0.001 μF) compared to the value of filter capacitors (20 μF) in the case of the rectifier for the mains supply. This difference in the value of the filter capacitors in the two cases arises due to the difference in the two frequencies (50 Hz and 455 kHz) to be rectified and filtered. The higher this frequency the lower the value of the required filter capacitor. Based on the principles described the circuit of a standard detector

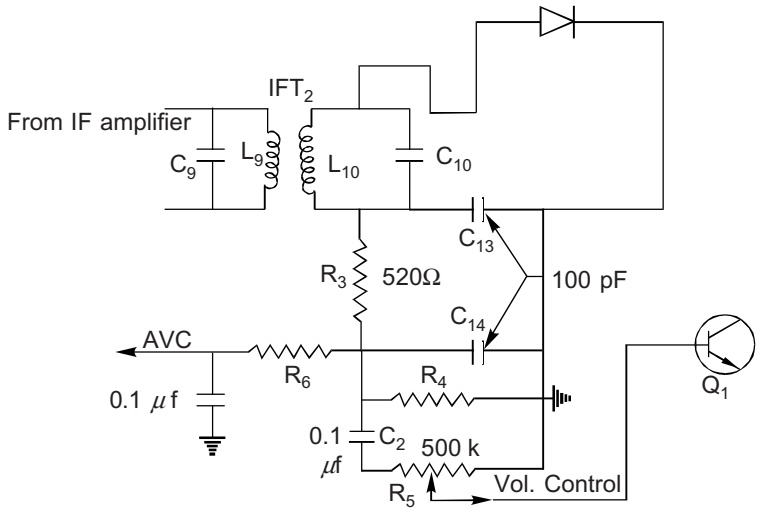


Fig. 14.17 Detector-amplifier Stage

stage in a superheterodyne receiver is given in Fig. 14.17. This circuit performs a number of functions besides detection.

The output from the IF amplifier is applied to the Anode of the diode from the secondary of IFT_2. The diode rectifies only on positive half-cycles of RF and the rectified current flowing through R_3 and R_4 produces a voltage drop on these two resistors. The RF portion of the rectified current is filtered by the filter consisting of R_3 (47 kΩ) and capacitors C_{13} and C_{14} each of which is 0.001 μF (100 pF). Thus the voltage appearing across R_5 is only the audio component of the modulated IF. R_5 is a variable resistor (500 kΩ) or a potentiometer from which a suitable value of audio voltage can be selected and applied to base of transistor Q_1 which forms the first stage of audio amplification. R_5 serves as the manual volume control for the receiver. C_2 (0.15 μF) is coupling capacitor for the first audio stage.

Manual Volume Control As mentioned above, the potentiometer R_4 helps to supply variable audio input for the first audio amplifier stage and the output from the audio stage can be controlled by rotating the shaft of the potentiometer with a knob fitted at the end. The audio voltage actually applied to the base of the transister Q_1 is the voltage between ground and the variable or sliding point of the potentiometer. This potentiometer which is generally a carbon potentiometer of 0.5 mΩ value forms the manual volume control of the receiver. The manual volume control in modern broadcast receivers has also the ON-OFF switch fitted in the same unit and is operated from the same common shaft. A typical volume control with the switch connection is shown in Fig. 14.18.

14.8.1 AGC (Automatic Gain Control)

One of the functions of the detector stage is to produce AGC or automatic gain control for the 'rf' stage of the receiver.

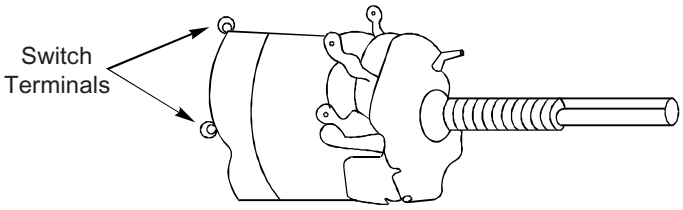


Fig. 14.18 Volume Control with On-off Switch

Almost all modern receivers are furnished with AGC (or AVC) which permits tuning to stations of varying signal strengths without appreciable change in the strength of the output signal. AGC thus *irons* out input signal amplitude variations and the gain control does not have to be re-adjusted every time the receiver is tuned from one station to another except when the disparity in signal strength or the receiver sensitivity is very great. In addition, AGC helps to smoothen out the rapid fading which may occur with long distance shortwave reception. It also prevents the overloading of the last IF amplifier which might otherwise have occurred.

How AGC is produced and applied can be explained with the help of Fig. 14.19.

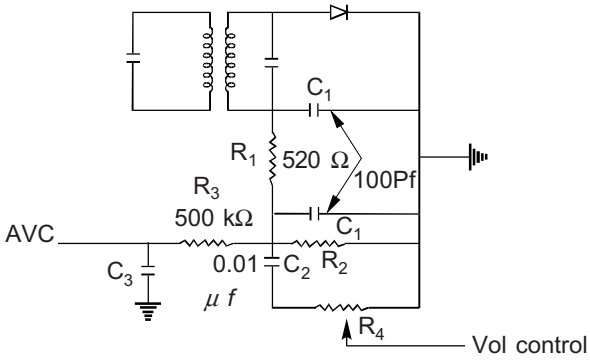


Fig. 14.19 AGC Circuit of a Transistor Receiver

R_1C_1 functions as a resistance capacitance filter that prevents the radio frequency ripple voltage developed across R_1 from reaching the output-terminal. Capacitor C_2 is for blocking the dc component of the rectified output from the volume control potentiometer R_4 and must have a small reactance at modulating frequencies compared to the value of R_4. The combination R_3C_3 is a resistance capacitance filter proportioned to remove modulation components. It provides a DC voltage propotional to the rectified carrier and free from modulation components and suitable for AGC purposes. This dc voltage controls the gain of the RF stages to keep it steady with the variations in the strength of the RF signal.

14.8.2 Simple AGC

In this system of simple AGC the overall gain of the receiver is varied automatically with the changing strength of the signal to keep the output substantially constant. A dc bias voltage derived from the detector is applied to a selected number of the IF, RF and mixer stages in the case of vacuum tube receivers which are voltage controlled. However, in the case of transistor receivers which are current controlled devices, a bias current is fed back so that some power is required for bias control. In this case the gain of the relevant amplifier is controlled by variation of the base current, provided sufficient AGC power is available.

14.8.3 Delayed AGC

Simple AGC discussed above does not provide ideal AGC characteristic for a receiver because the gain is reduced by AGC action even when the output is less than desired value. As a result, circuit modification involving a second diode are sometimes used to make the AGC system inoperative until the signal at the second detector reaches the desired level. A double-diode is used in the case of tube type receivers but in the case of transistor receiver two separate diodes are employed, one being the detecter diode and the other the AGC diode as shown in Fig. 14.20.

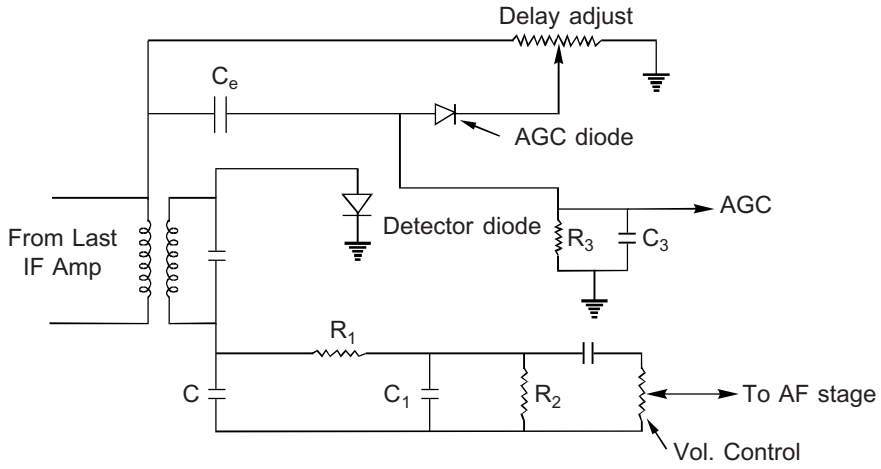


Fig. 14.20 Delayed AGC

These diodes can be connected either to separate windings of the IF transformer as shown in Fig 14.20 or both to the secondary without too much interference. As indicated a positive bias is applied to the cathode of the AGC diode to prevent conduction until a predetermined level has been reached. A control is often provided as shown, to allow manual adjustment of the bias on the AGC diode and hence of the signal level at which AGC is applied. If mostly weak signals are likely to be received, the delay setting may be quite

high which means no AGC until signal level is fairly high. It is, however, preferable to make it as low as possible to prevent over loading of the last IF amplifier by unexpected stronger signals.

Various types of AGC characteristics are shown in Fig. 14.21.

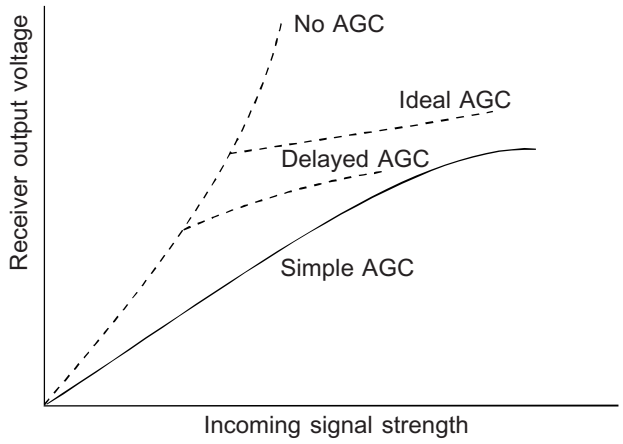


Fig. 14.21 AGC Characteristics

14.8.4 Distortion In AM Detection

In diode detection described earlier, the capacitor *C* charges to the peak value of the carrier voltage during the positive half cycle and then slowly discharges through the load resistance *R* during the negative half cycle of the modulating voltage. For the detected voltage to be an exact replica of the modulating signal waveform, the change in capacitor voltage should be sufficiently rapid to follow the modulation envelope. If it does not, amplitude distortion will result. There are two types of distortions that are observed in practical diode detection. These are Diagonal clipping and Peak negative clipping.

Diagonal Clipping This can be explained with the help of Fig. 14.22. In this, Fig. 14.22 (a) represents the normal diode rectification circuit and Fig. (b) is its equivalent circuit.

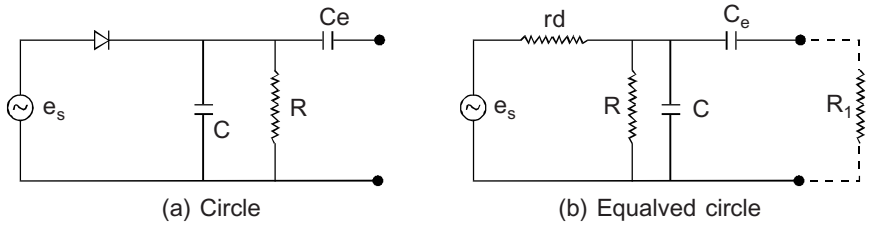


Fig. 14.22 Diode Rectifier

In the above circuit, the capacitor *C* charges in series with the diode resistor *rd*. Whenever the applied voltage exceeds the voltage across the capacitor *C*,

the conduction starts and the capacitor charges through resistance *rd* which is very small compared to *R*. The conduction starts at some point A in Fig. 14.23 and continues till point *B* in a manner indicated by line *AB*. At point *B* the source voltage at carrier frequency is falling faster than *C* can change and the voltage e_c across the capacitor becomes larger than e_s and diode conduction stops due to reverse biasing. The capacitor *C* then discharges through the resistance *R* which is large compared to *rd* and time constant *RC* is large and the voltage falls to *C* when the conduction starts again and the same process is repeated. Thus, between *B* and *C* the capacitor discharges exponentially rather than follow the modulation curve. This results in diagonal clipping. This type of distortation does not occur below 60% modulation at the highest frequency but the size of the filter capacitor becomes a limiting factor for a distortionless detector circuit.

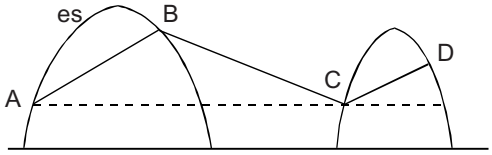


Fig. 14.23 Diagonal Clipping

It can be proved mathematically that distortion is not excessive if the time constant *RC* does not exceed $\frac{1}{w_m \cdot ma}$.

Where w_m is the modulation frequency ($2\pi f_m$) and m_a is the modulation index as defined in Chapter 12.

Negative Peak Clipping This is a distortion that results from the loading effect of C_e and R_1 [Fig. 14.22] C_e is the dc blocking capacitor and R_1 represent the input resistance of the following AF stage. Capacitor C_e is large enough to allow the modulation component of the voltage to pass unattenuated to R_1 and as a result C_e maintains a constant voltage at the mean load voltage of $V = 1$ volt and V_1 follows the modulation envelope maintaining the relationship $V - V_1 = V_{ce}$ = constant

However *V* cannot drop below the min. level set by

$$V_{\min} = V_{ce} \cdot \frac{R}{R + R_1}$$

$$= \frac{R}{R + R_1} \text{ when } V_{ce} = 1V$$

below this minimum level the capacitor *C*, no longer follows the modulation envelope resulting in negative peak clipping as shown in Fig. 14.24.

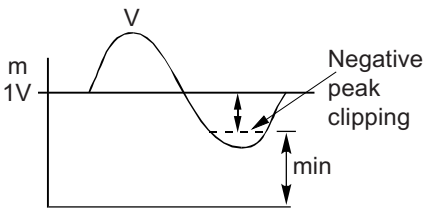


Fig. 14.24 Negative Peak Clipping

It can further be shown that the max., modulation m_{max} should not

exceed the value max $\leq \frac{R_1}{R_1 + R}$ to avoid negative peak clipping. A maximum modulation of about 70%, without negative peak clipping can be achieved in well designed practical diode detector circuits in AM Broadcasting system.

14.9 TRANSISTOR RADIO RECEIVERS

A transistor can perform practically all the functions of a triode vacuum tube but it has many advantages over a triode valve. It is small in size, light in weight, requires no heating power and works on low voltage. Because of these advantages, transistors are rapidly replacing vacuum tubes in all electronic equipment including radio receivers.

A transistor receiver is like any other normal receiver in which valves have been replaced by transistors. With the use of transistors, the size and weight of all other components required for the circuit is also considerably reduced. Thus, reasonably priced pocket transistor receivers are now available in all convenient shapes and sizes.

Except for some receivers meant to receive only local stations, all other modern transistor receivers are of the superheterodyne type. A transistor superheterodyne receiver makes use of the stages which function in the same way as the stages of a valve type receiver. However, there are certain differences in the circuits employed in the two cases because of some basic differences between the properties of transistors and vacuum tubes. In order to clearly bring out the points of difference and similarity between the circuits used in the two types of receivers, the circuit diagram of a standard transistor receiver is given in Fig. 14.25. A brief description of the various stages and their functions is given.

14.9.1 Convertor Stage

A convertor circuit is shown in Fig. 14.26. In this, only one transistor serves as a convertor but sometimes two separate transistors, a mixer and a local oscillator, perform the function of a convertor. The signal from the ferrite antenna coil L_1 is fed to the base of the transistor Q_1. Resistors R_1 (6.8 K) and R_2 (39 K) form a voltages divider across the battery for the emitter-base bias. Resistor R_3 (3.3 K) in the emitter lead is a DC stabilizer. The oscillator coil is L_2 and it feeds its local oscillations to the emitter of Q_1 through capacitor C_3. The tuning and oscillator capacitors are ganged as in a tube receiver. The station signal to the base and the local oscillator signal to the emitter mix to produce the modulated IF signal at the collector which applies it to the IF transformer. Capacitor C_2 (0.02 μF) is an IF decoupling capacitor to prevent IF signals from being fed back to the base. Taps are used at the oscillator coil and the IF transformer winding for impedance matching. Q_1 is a PNP transistor in which the collector is at a negative potential with respect to the base.

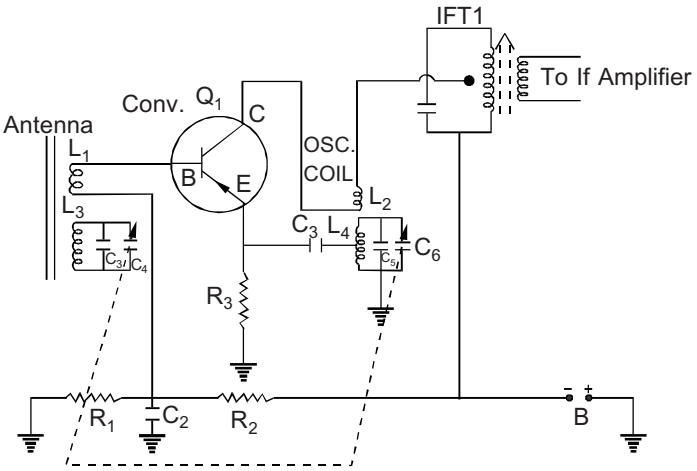


Fig. 14.26 Convertor Stage

14.9.2 IF Amplifier Stage

The IF amplifier stage is shown in Fig. 14.27. A transistor receiver normally has two IF amplification stages compared to only one IF stage in a tube receiver. This is due to the fact that a transistor amplifier has less gain than a similar tube amplifier.

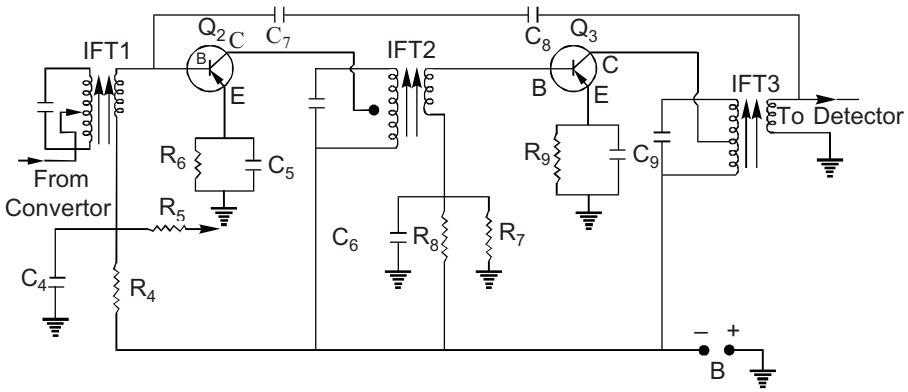


Fig. 14.27 IF Amplifier Stage

The IF signal (455 kHz) from the convertor is fed to the IF input transformer IFT1. Transformers IFT1, IFT2 and IFT3 are permeability tuned transformers with slug adjustments. Q_2 and Q_3 are PNP transistors and the bias voltages for various transistor elements are arranged accordingly. Emitter components R_6 and R_9 are DC stabilisers and are bypassed by capacitors C_5 and C_9 respectively to avoid negative feedback. AVC voltage is fed to the base of the first IF transistor Q_2. Resistors R_4, R_5, R_7 and R_8 are voltage dividers to provide the emitter base bias. Capacitors C_4 and C_6 are decoupling capacitors. To prevent oscillations of the IF amplifiers, capacitors C_7 and C_8 are used as neutralising

capacitors which couple the signal from the output back to the input in the proper phase.

14.9.3 Detector Stage

A transistor receiver makes use of a semiconductor diode in its detector stage. This diode has similar properties of detection and rectification as a vacuum tube diode. Besides the detection of the IF signal, the diode also develops the AVC voltage for the first IF amplifier. The diode detector (Fig. 14.28) receives the modulated signal from the last IF stage. The diode load is made up of resistor R_{10} and volume control R_{11}. The tap between the diode and R_{10} is the source of the AVC voltage. Capacitor C_{10} (0.01 μF) filters out any IF signal in the detector output. Audio voltage in the AVC line is filtered out by R_5 (5.6 K) and C_4 (8 μF).

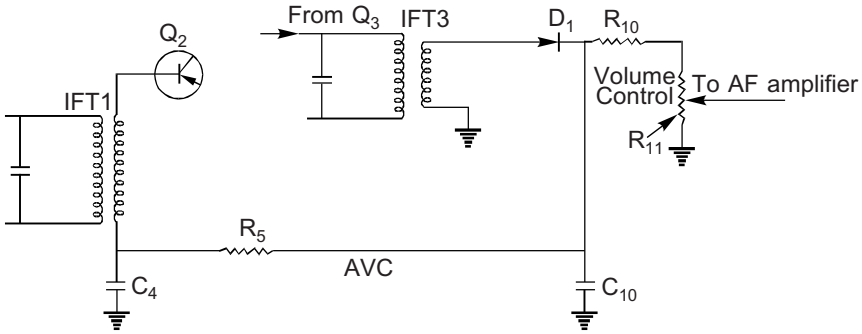


Fig. 14.28 Detector Stage

14.9.4 Audio Amplifier Stage

The AF stage normally consists of two audio amplifier stages—the audio driver or voltage amplifier stage and the power output stage.

Audio Driver Since the detector output is insufficient to drive the audio power output stage, an audio amplifier called the audio driver is placed between the two stages. This is generally a Class A voltage amplifier. A circuit diagram of the audio driver is shown in Fig. 14.29.

The audio signal from the detector is fed to the base of the driver transistor Q_4 through the coupling capacitor C_{11} (8 pF). Resistors R_{12} and R_{13} are the voltage dividers for the base emitter bias. Resistor R_{14} is the DC stabiliser bypassed by capacitor C_{13} (20 μF), C_{14} (0.005 μF) is the RF bypass filter. C_{12} (30 μF) is an electrolytic capacitor and is shunted across the battery supply for decoupling signals from all stages.

Power Output Stage The power output stage in a transistor radio receiver is generally a Class B pushpull amplifier particularly in the case of transistor receivers operated from batteries where power consumption is an important consideration. It is known that a properly designed Class B amplifier using a

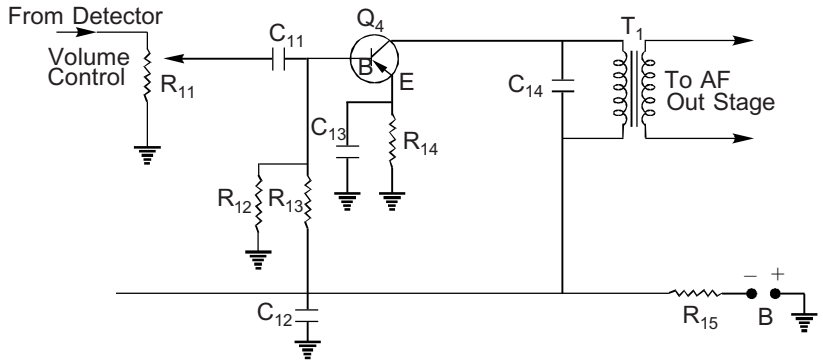


Fig. 14.29 Driver Stage

matched pair of transistors draws no current from the batteries when no input audio signal is being received. There are a number of circuits in use for the output stage but only two popular circuits will be described here. Figure 14.30 shows the circuit of a conventional Class B pushpull amplifier for the output stage.

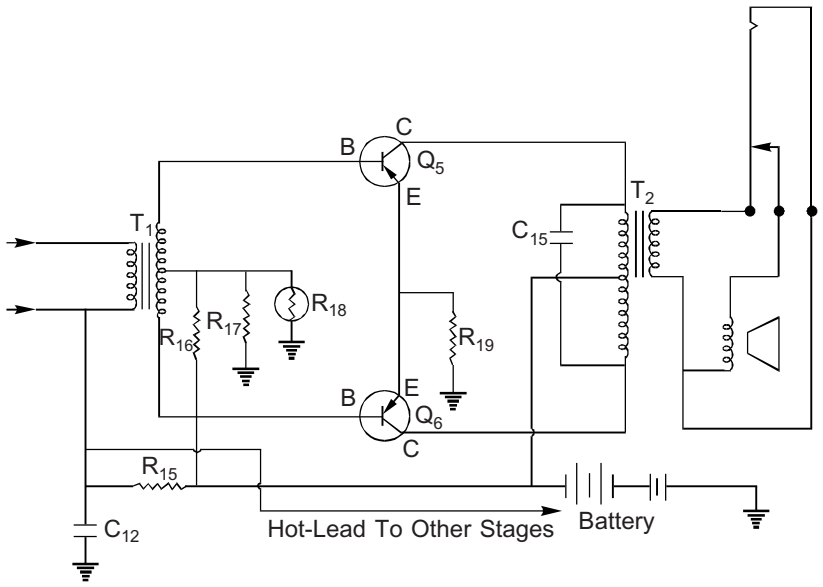


Fig. 14.30 Pushpull Power Output Stage

The audio signal is fed to the bases of the output transistors Q_5 and Q_6 from the driver stage through audio transformer T_1. Since the transformer T_1 has a centre tap, the audio voltage applied to the bases of Q_5 and Q_6 are 180° out of phase, which fulfills the requirements for a pushpull amplifier.

Resistor R_{19} (15 Ω) is the DC stabilising resistor which provides negative feedback, R_{16} (7.5 K) and R_{18} (330 Ω) provide the voltage divider for the

emitter-base bias. Thermistor R_{18} is a stabiliser against thermal changes. R_{15} (220 Ω) and C_{12} (50 μF) form a decoupling filter. Capacitor C_{15} (0.02 μF) is a bypass capacitor across the primary of the output transformer T_2 for bypassing the higher audio frequencies. The output transformer feeds the output signal to the loudspeaker or the earphone through the earphone jack. The output transformer also provides the impedance matching between the high output impedance of the pushpull amplifier and the low impedance of the voice coil of the loudspeaker.

Transformerless Output Stage A circuit without the output transformer is being commercially employed for the output stage by some manufacturers of transistor radio receivers. This results in reduction both in the cost and weight of the receiver. A circuit diagram for such a transformerless output stage is given in Fig. 14.31.

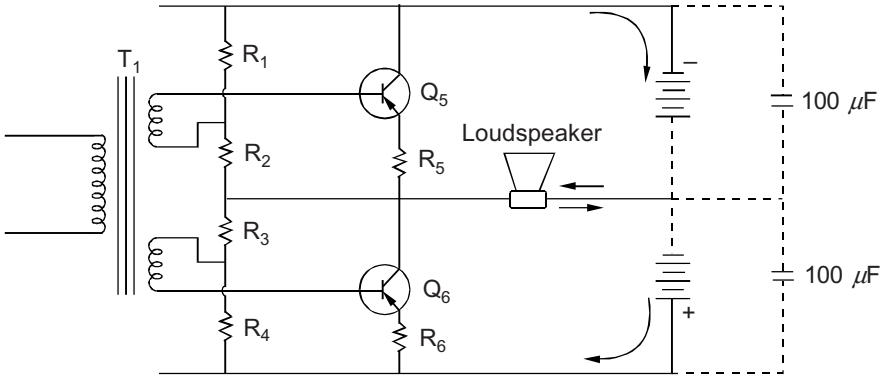


Fig. 14.31 Pushpull Output Stage without Output Transformer

The input transformer T_1 is of a special design and has two separate and independent windings. The base voltage of the two transistors Q_5 and Q_6 can be adjusted by the values of R_1, R_2, R_3 and R_4 so as to allow a known value of current to flow through the collectors. Resistors R_5 and R_6 are for bias stabilisation.

The signal voltages applied to the bases of Q_5 and Q_6 by the two secondaries of the transformer T_1 are opposite in phase. The collector currents will also vary according to the signal voltage applied to the two bases. These collector currents will flow in opposite directions through the common load which in this case is the loudspeaker. The voltage developed across the voice coil of the loudspeaker is due to the difference between the two collector currents thereby creating the same conditions as exist in a normal or conventional pushpull amplifier.

It is interesting to note how impedance matching between the output impedance of the pushpull amplifier and the loudspeaker impedance is achieved without the use of a matching transformer. The two collector impedances are in parallel and hence the output impedance of the combination is one half the

output impedance of each collector circuit (assuming a matched pair) and one fourth the output impedance of a normal pushpull circuit where the collector impedances are in series. A reduction in the output impedance of the collector can also be effected by suitable selection of the battery voltage (reverse bias). Further, by selecting transistors with low output impedance and the loudspeaker of proper impedance value, it is possible to obtain good impedance matching conditions. It will be seen from Fig. 14.31 that one side of the loudspeaker is to be connected to the middle point of the battery which is not easily available. In such a case, two electrolytic capacitors (100 to 400 μF) can be connected as shown dotted in the diagram to obtain the mid-point of the battery.

The circuit diagram of a transistor receiver in which the power output stage is without the output transformer is given in Fig. 14.32.

14.10 MULTIBAND RECEIVERS

The receivers described so far are single band receivers which are capable of receiving stations operating on the broadcast band or medium wave (MW) band which extends from 535 to 1605 kHz. The frequency band or the wave band that can be covered is decided by the inductance of the coil and the capacitance of the variable capacitor that form the tuned circuits in the antenna section and the oscillator section of the convertor stage. In modern superheterodyne receivers it becomes necessary to have a number of short wave (SW) bands in addition to the (MW) band, so that stations broadcasting on SW frequencies can also be received. A receiver which includes one or two SW bands in addition to a MW band is called a *multiband receiver*.

One coil and a variable capacitor can be tuned over one band of frequencies. For a second band of frequencies, a different coil with different number of turns is required both for the antenna section and the oscillator section of the convertor stage. The variable capacitors, which are parts of a two gang capacitor will remain the same. Thus for changing from one wave band to another, one set of coils and trimmers should be disconnected and another set of coils and trimmers should be connected to the circuit, simultaneously in the antenna section and the oscillator section of the convertor stage of the receiver. This involves a number of switching operations that must be performed before a waveband can be changed. This is done with the help of a special multicontact switch known as the band change switch or a wave change switch or merely as the band switch. The most common type of band switch used in multiband receivers is the rotary band switch which will be described here.

14.10.1 Rotary Band Switch

A rotary band switch consists of movable and fixed contacts. The movable contacts are known as poles and the fixed contacts are called positions. The number of positions per pole is equal to the number of wave-bands in the multiband receiver. The number of poles is determined by the number of terminals in each set of a coil and a trimmer that has to be changed both in the

antenna section and the oscillator section of the convertor stage. Thus, in a 5-valve 3-band superheterodyne receiver without any RF stage, the number of contacts that require changing for every wave change is two in the antenna section and two in the oscillator section thereby requiring 4-poles with 3-positions for each pole. The band switch required is a 4-pole three positions switch or a 4-pole three way switch.

Each rotary band switch is made up of separate sections called *wafers*. A wafer with fixed and moveable contacts is shown in Fig. 14.33.

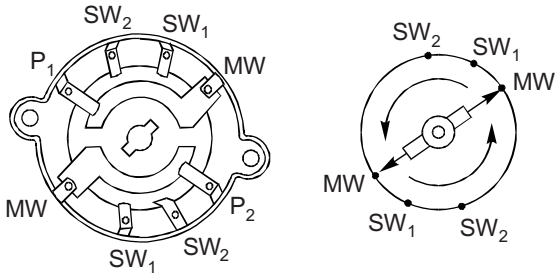


Fig. 14.33 Wafer of a Band Switch

Fixed contacts MW, SW_1 and SW_2 are punched on a circular plate of fibre or some other suitable insulating material. A smaller circular plate of the same insulating material carries two semicircular metallic strips with projections on one end, which will make contact with the positions MW, SW_1 and SW_2, one by one as the inner fibre plate is rotated with a central shaft. The fixed poles P_1 and P_2 punched on the outer fibre plate remain permanently in contact with their respective rotating metallic strips.

Two such wafers will be required for a normal three band radio receiver with one medium wave band and two short-wave bands SW_1 and SW_2. These wafers are arranged one above the other in the form of separate decks supported by means of nuts and bolts passing through metallic sleeves. For a wave change, the inner fibre plate can be rotated clockwise or counter clockwise by means of a common central shaft with a knob fixed at one end. A metallic plate with grooves is fixed on the top and this metal plate carries a leaf-spring and a ball to keep the switch held firmly in a particular position of the switch.

Receivers having RF stages between the antenna and the convertor stage and receivers having a larger number of wave bands require complicated rotary band switches with larger number of poles and positions and consequently use rotary switches for performing many auxiliary functions such as shorting the coils not in use, lighting pilot lamps to illuminate selected wavebands and connecting the pickup terminals to the audio section etc.

In a multiband transistor radio receiver, the switching arrangement for the convertor stage is very much similar to the arrangement described. A three band transistor receiver will require a 6-pole 3-way band change switch as shown in Fig. 14.34.

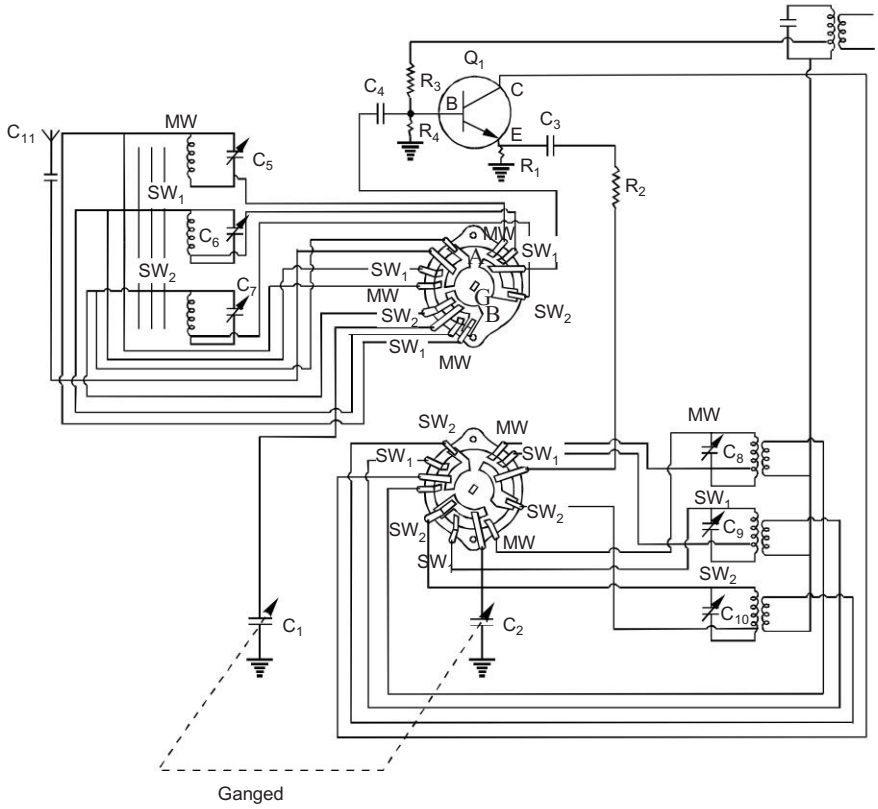


Fig. 14.34 Convertor Stage of a 3-band Transistor Receiver (Wiring Schematic)

In this diagram, two separate wafers have been shown, one for the antenna section and the other for the oscillator section of the convertor stage. Each wafer has three poles and three positions MW, SW_1 and SW_2 for each pole. The poles for the antenna section wafer are A (antenna), B (base) and G (gang) whereas the three poles for the oscillator section have been shown as C (collector), E (emitter) and G (gang).

A circuit diagram of a simple 3-band transistor receiver using six transistors and one diode is shown in Fig. 14.35. Different poles of the band change switch and the positions controlled by each pole are indicated in a schematic way in the circuit diagram.

Some of the rotary band switches used in transistor radio receivers have separate wafers for the poles and the positions. This however, does not change the wiring arrangement for the band change switch.

Other band change switches in common use are the piano type band change switch and the sliding type band change switch. Push button type of tuning switches can be used to listen to certain selected stations without turning the manual tuning control. Push button type switches are more useful than manual tuning control. Push button type switches are mostly useful in automobile receivers.

14.11 FM RECEIVERS

Frequency modulation is extensively used in receivers operating in frequency ranges much above that used for AM receivers. FM receivers are mainly used for broadcasting, Television (sound only), Police radio and military systems. VHF range from 108 to 118 MHz is specially reserved for FM Radio Broadcasting. A block diagram of an FM receiver is given in Fig. 14.36.

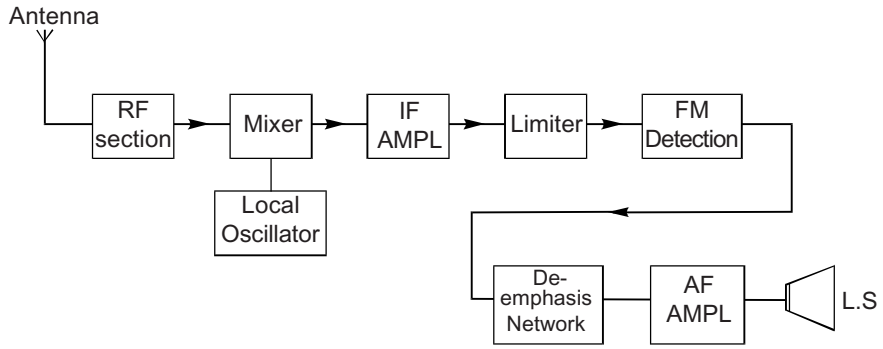


Fig. 14.36 Block Diagram of FM Receiver

FM receiver is a superheterodyne type receiver and all stages upto and including IF amplifier correspond to similar stages in AM receivers. The main difference lies in the FM detection stage which is preceded by a limiter to avoid excessive distortion due to unavoidable amplitude variations.

A stage wise description of the block diagram is given below:–

1. RF Section It consists of at least one RF amplifier to improve the signal-to-noise ratio which will be low otherwise due to large bandwidth requirements. Tuned RF stages also help ensure good image rejection ratio.

2. Mixer Stage The Mixer stage makes use of FET and the local oscillator uses bipolar transistors with the Colpitt and Clapp circuits being quite popular. These oscillators operate in the VHF range and tracking is not a problem at these high frequencies because the tuning frequency range is only 1.25:1 which is much less than in AM broadcast receivers.

3. IF Amplifier The IF amplifier typically operates at a relatively high frequency of 10.7 MHz and a bandwidth of about 200 kHz. This is the bandwidth for broadcast receivers operating in the 88-108 MHz frequency range.

The receiver bandwidth varies with the type of service for which the receiver is intended. In frequency modulation broadcasts the standard provide for a max. frequency deviation of 75 kHz corresponding to a total bandwidth of 150 kHz. However, in types of service like the Police Radio, where the main objective is to obtain intelligible communication, frequency deviation as low as 15 kHz is often used.

4. Limiter Stage In order to make full use of the advantages offered by FM, it is necessary that any amplitude variation of the incoming signal must be minimized by either introducing an amplitude limiting stage before the FM detector or by employing a ratio detector or by a combination of both. An amplitude limiter has already been described and a ratio detector which has inbuilt propular amplitude limiting device is described below. An amplitude limiter is only a clipping device which will keep the output of IF amplifier constant despite changes in the input signal.

14.11.1 FM Detection

The main difference between an AM and FM receiver lies in the FM detection process.

In FM detection the frequency variations of an FM signal are first converted into AF amplitude variations similar to those that were originally responsible for the frequency deviation as per definition of frequency modulation. The conversion process should be efficient, linear and insensitive to amplitude variations. The simplest method is slope detection.

Slope Detection In this method the FM wave is applied to a tuned circuit which is tuned slightly away from the center frequency of the FM signal f_0. The selectivity curve of the tuned circuit and the variations of the input signal are shown in Fig. 14.37.

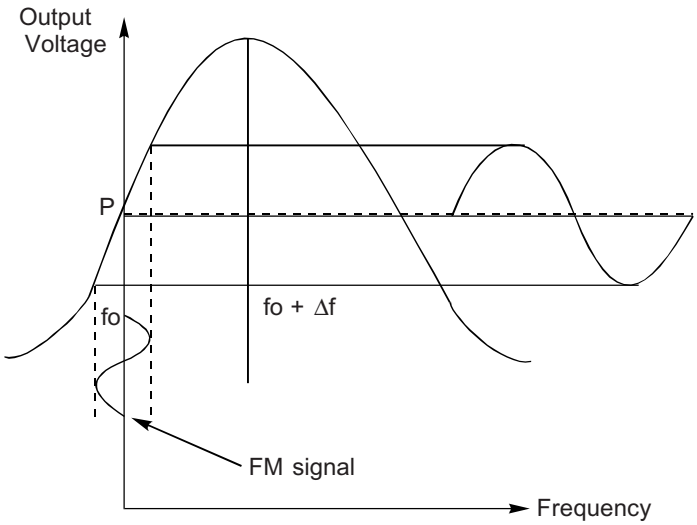


Fig. 14.37 Slope Detection

The output voltage developed in the tuned circuit will vary as the frequency of the applied FM signal. The output voltage of the tuned circuit will vary in amplitude as the frequency deviates away from the center frequency on both sides of point *P* on the selectivity curve . The AM signal developed as a result

of FM detection is then applied to a diode detector for AM detection to make it suitable for use as an input signal for the AF stage. The method of slope detection is neither efficient nor linear and reacts to amplitude variations.

A large number of other circuits are available for FM detection but the two methods most commonly used are (a) the phase discriminator and (b) the ratio detector.

The Phase Shift Discriminator (Foster-Seeley Discriminator) The circuit of a phase shift discriminator is shown in Fig. 14.38.

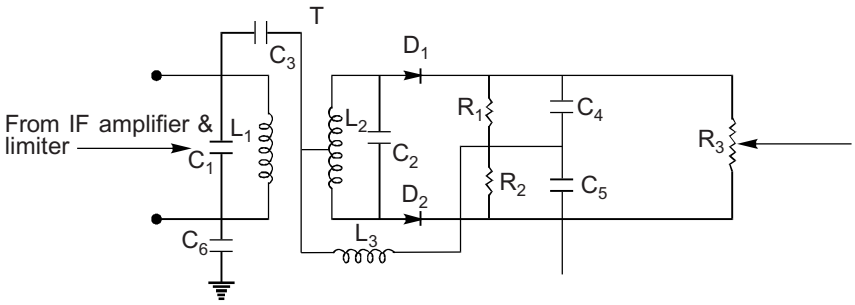


Fig. 14.38 Phase Shift Discriminator

The output from the IF amplifier and limiter is applied to the diode D_1 and D_2 through the secondary winding L_2 of the transformer, T. Both the primary L_1 and the secondary L_2 of the transformer are tuned to the same IF frequency of 5.5 MHz as used in TV receiver. The IF signal from the primary L_1 is also applied to the center of the secondary winding L_2 through a capacitor C_3. With this arrangement the voltage applied to each diode is the vector sum of the primary voltage and half the secondary voltage which are 90° out of phase when the input frequency is 5.5 MHz. In this position equal voltages are applied to the two diodes and the output voltages across R_1 and R_2 are equal and opposite resulting in zero output voltage. However, when the IF frequency deviates from the center frequency the phase difference of 90° between the primary and half secondary voltages no longer holds, the voltages applied to the two diodes are not equal and a difference voltage appears at the output. The amount of output voltage depends on the frequency deviation and the sign of the voltage depends on whether the deviation is above or below the center frequency of 5.5 MHz. The frequency deviations of the original IF have thus been converted into corresponding amplitude variations of the output voltage which are amplified by the audio amplifier.

The phase-shift discriminator has the disadvantage that its output is sensitive to amplitude variations and so needs a limiter stage to precede the detector stage.

Ratio Detector A detector circuit which is not sensitive to amplitude variations and so does not require a limiter to be included in the circuit is shown in Fig. 14.39. This circuit which is known as a *ratio detector* differs from the

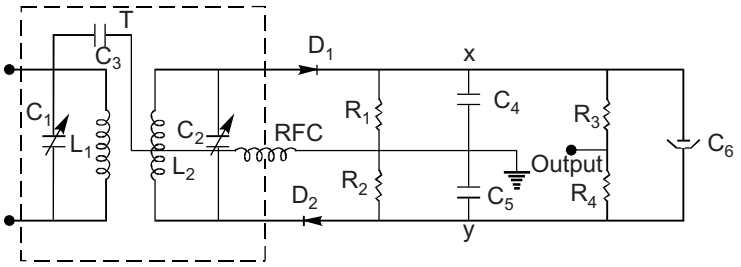


Fig. 14.39 Ratio Detector

circuit of the phase-shift discriminator in two respects (a) the two diodes D_1 and D_2 are connected in series across the secondary of the transformer T_1 and (b) the output is taken between the ground and the center tap of resistance R_3 and R_4.

The primary as well as secondary of the transformer are again tuned to the same center frequency (5.5 MHz) and the voltages applied to D_1 and D_2 have the same phase relationships. However, the rectified voltages across the two diodes will be the sum of the voltages. A large electrolytic capacitor C_6 connected across XY will get charged to this sum voltage and keep it constant. As the frequency of the IF signal varies above and below the center frequency, different voltage will be developed across the two diodes in such a way that when one voltage increases the other voltage will decrease but the sum of the two voltages will be substantially constant. An output voltage will be obtained from the circuit only when the ratio of the two voltages developed by the diodes varies and hence the name of the circuit as ratio detector. Because of its high capacitance value (8-10 μF) and large time constant, the capacitor C_6 does not allow any voltage to be built up across the output due to amplitude variations of the IF signal caused by noise or any other interferences.

A practical circuit normally used in TV receivers is shown in Fig. 14.40. In this circuit, a tertiary (third) winding L_3 in the transformer T applies the primary voltage to the two halves of the secondary winding by inductive coupling.

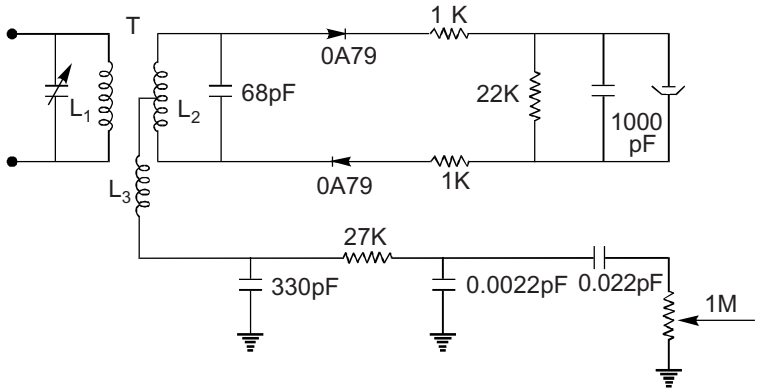


Fig. 14.40 Practical Circuit of a Ratio Detector

Series resistances in each diode lead are meant to improve linearity at higher signal conditions. The filter circuit consisting of a resistor 27K and capacitor 0.0022 pF which attenuates higher frequency provides what is known as de-emphasis. This is necessary to compensate for the boost given to the high frequencies at the transmitter in FM by a process called pre-emphasis. This is the de-emphasis circuit shown in the block diagram.

14.12 AF AMPLIFIER

The audio stage consists of an AF amplifier which includes a preamplifier and a Power amplifier to feed a loudspeaker or two speakers if FM stereo system is in use.

14.13 FM RADIO PAGING SERVICE

Paging means to "alert" or to 'summon' a person who is on the move. When paging is done with the help of radio it is called radio paging or FM radio paging when FM transmissions are used for radio paging. It is a one way transfer of coded messages to the person concerned only. In other words when a person is called or "paged", other subscribers do not know what message was sent to him. The radio paging is a specialised value added service with built in selective calling features. When radio paging is carried out in accordance with special Radio Data System (RDS) standards laid down by CCIR, it is called FM-RDS-Paging Service.

FM-RDS Paging service is a wireless way of communicating with a person who is in possession of a tiny set called "pagers". A group of persons or subscribers who have pager can avail of this facility of getting messages from all over the city. Each pager has a 6 or 7 digit code number called "cap code" and the messages are radiated by FM transmitter along with the capcode of the pager as shown in Fig. 14.41.

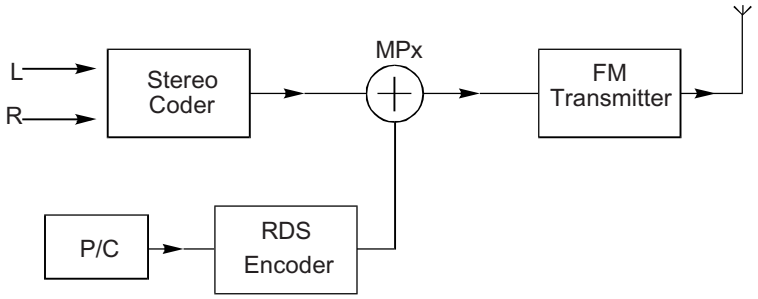


Fig. 14.41 FM-RDS-Paging Service

In paging network, paging control room collects the messages from the paging subscribers manually or automatically through telephone network, stores and forwards them to a FM transmitter where RDS encoder takes over the task of transmitting the messages on a 57 kHz RDS sub-carrier.

RDS paging service is being used in most advanced countries of the world. All India Radio has also a network of 85 FM transmitters and more transmitters are being installed. All India Radio has introduced FMRDS paging service at Major FM centres at Delhi, Bombay, Calcutta and Madras and this service will soon be extended to other important centres, like Pune, Hyderabad, Patna, Jallandhar and Nagpur etc.

Full details of the FM stereophonic service mentioned above are available in Chapter 22 under stereophonic sound systems.

14.14 COMMUNICATION RECEIVERS

A communication receiver is used for the reception of communication signals rather than for entertainment and broadcast purposes. The receiver is designed to receive low and high frequency signals (short wave reception) better than an ordinary domestic broadcast receiver. Its sensitivity makes it suitable for many other purposes like detector of signals in h.f. impedance bridges and signal strength measurements.

A communication receiver shown in Fig. 14.42 is basically a superheterodyne receiver which contains certain special features normally not included in a domestic broadcast receiver. Some of these features are described below:

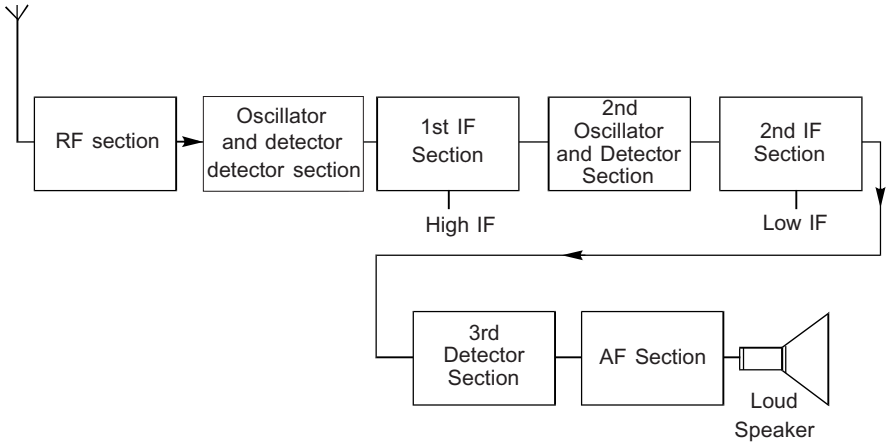


Fig. 14.42 Block Diagram of a Communication Receiver

1. Double Conversion or Triple Detection Here the incoming signal is first transformed to a relatively high IF in the usual manner. After one or two stages of amplification at this frequency the signal is then transformed to a second and low IF (200 kHz) by means of a second mixer in association with a local oscillator of fixed frequency (crystal oscillator).

Such an arrangement effectively suppresses image signals, since the high value of first IF provides the desired image rejection ratio and the low value of second IF (200 kHz) makes it possible to obtain high amplification per stage as well as sharp discrimination against adjacent channel signals. The triple action

receiver provides a combination of greater image suppression and higher adjacent channel selectivity compared to an ordinary superhet receiver.

It is essential that high IF must come first otherwise the image frequency will be insufficiently rejected at the input and will become so inter mixed with the proper signal that no amount of high IF stages will make any difference afterwards.

B.F.O. (Beat Frequency Oscillator) This required for reception of morse code or pulse modulated RF carrier. This BFO is only an LC Hartley oscillator operating at 1 kHz or 400 Hz above or below the last IF to produce a beat note of that frequency. Since the signal is present only during a dot or a dash, only then their beats are produced in morse code and the code can be heard satisfactorily. To prevent interference the BFO is switched off when normal reception is resumed.

Squelch (Muting) System The AGC disappears in the absence of a carrier and the receiver aquires its maximum sensitivity and disagreeable amount of noise is heard in the output of the receiver. This happens particularly when waiting for a station to come up on the air or while tuning from one station to the other. The "squelch" current enables the receiver output to remain cut off during the absence of the carrier. Systems such as Police wireless, ambulances and Coast guard radio stations in which a receiver must be kept tuned and manned all the time, but transmissions are sporadic, are the principal beneficiaries of the squelch system. Such devices are also sometimes referred to as "Muting" or Quieting systems, tuning silencers and "inter channel noise suppressors". The system is also known as "Codans" which is formed from the first letters of the phrase "carrier operated device antinoise" systems.

The circuit actually consists of a dc amplifier to which AGC is applied and which operates on the first audio amplifier of the receiver. When the AGC is low or zero the dc amplifier draws current and the voltage drop across its load cuts off the audio amplifier. When the AGC voltage becomes sufficiently negative the dc amplifier draws no current and the bias to dc amplifier is restored to its normal value, the audio amplifier now functions as if the squelch circuit is not there.

Automatic Frequency Control (AFC) This is an arrangement for automatically keeping the frequency of the local oscillator of the superheterodyne receiver at the value required to produce the desired IF, in spite of the normal tendency of the local oscillator of the superhet receiver to drift-with the line voltage changes, component ageing etc.

A block diagram of the receiver AFC system is shown in Fig. 14.43.

The heart of the AFC system is a frequency sensitive device such as a phase discriminator which produces a dc voltage whose amplitude and polarity are proportional to the amount and direction of the local oscillator frequency error.

This dc control voltage is then used to vary automatically, the bias on a variable-reactance device (Varactor) whose output capacitance is then changed.

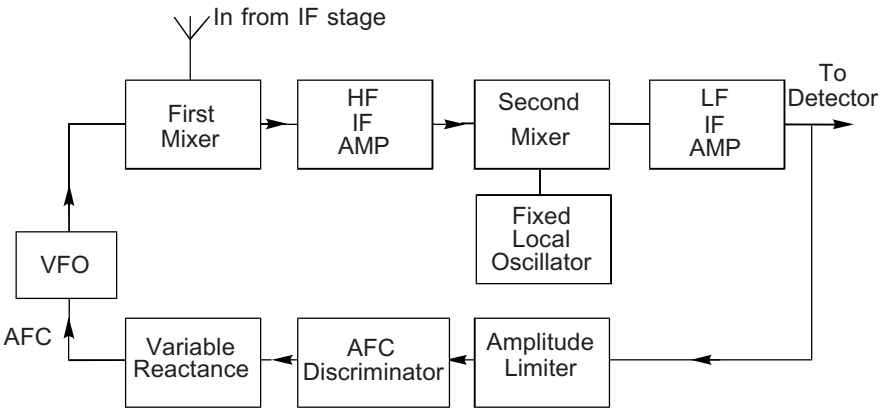


Fig. 14.43 AFC System

The variable capacitance appears across the local oscillator and the frequency of the VFO (Variable Frequency Oscillator) is prevented from drifting. AFC is extensively used in high frequency receivers where a small percentage drift will affect the operation of the receiver. It is a feature of better Broadcast FM equipment and is also used in military equipment.

Summary

The three main functions performed by a radio receiver are the selection of the desired station, the detection or demodulation of the RF signal and the conversion of the AF currents into sound waves that can be easily heard by the human ear.

A receiver can be an AM receiver or an FM receiver depending upon the type of modulation used for the received signals.

Selectivity, sensitivity and fidelity are the three main characteristics of a receiver. These characteristics depend on the type of circuit used for the receiver.

A straight receiver uses only selection, detection and audio amplification for its working. A crystal set is the simplest form of straight receiver which has a tuned *LC* circuit, a crystal detector and a pair of headphones. It has no amplification and does not need any power supplies. A TRF receiver uses selection, RF amplification, detection and AF amplification. It is a complete receiver. A TRF receiver has been largely replaced by the modern superheterodyne receiver which has many advantages over all other types of receivers.

In a superheterodyne receiver, the frequency of the incoming signal is converted into a lower fixed frequency called the intermediate frequency (IF). This is done in the mixer or convertor stage where the incoming frequency is allowed to beat or heterodyne with the unmodulated frequency produced by a local oscillator. The difference frequency in the output of the mixer tube is the IF which is selected by a fixed tuned IF transformer. This IF which is modulated and has a fixed frequency of 455 kHz is amplified by one or two stages of

amplification and then detected by the detector to produce audio frequencies which are amplified by the audio output stage to drive a loudspeaker. A superheterodyne receiver basically consists of the antenna, the RF amplifier, the mixer or the frequency convertor consisting of a mixer and a local oscillator, the IF amplifier, the detector, the AF amplifier and the loudspeaker. The detector also produces the AVC voltage for controlling the gain of the RF stages so that the output does not vary with the strength of the received signal.

Transistors are rapidly replacing vacuum tubes in radio receivers. A transistor receiver is more compact, portable and less costly. Most transistor receivers are also superheterodyne receivers which function in a manner similar to that of tube type superheterodyne receivers. There are some differences in the circuitry used in two types of receivers because of certain differences in the properties of the vacuum tubes and transistors. Some of the transistor receivers employ push-pull output stages without the output transformers thereby reducing the cost and the weight of the receiver.

Multiband receivers have one or two SW/MW band. Each band requires a different set of coils and trimmers in the antenna and the oscillator section. This change of coils is done with a band change switch of which there are many types, the most common being the rotary switch.

FM receivers operating in the VHF range require special methods of FM detection like phase shift discriminator and radio detection. Radio paging service is a special feature of FM radio. Communication receivers are highly sensitive receivers having special features like double conversion, AFC and squelch (muting) system.

Review Questions

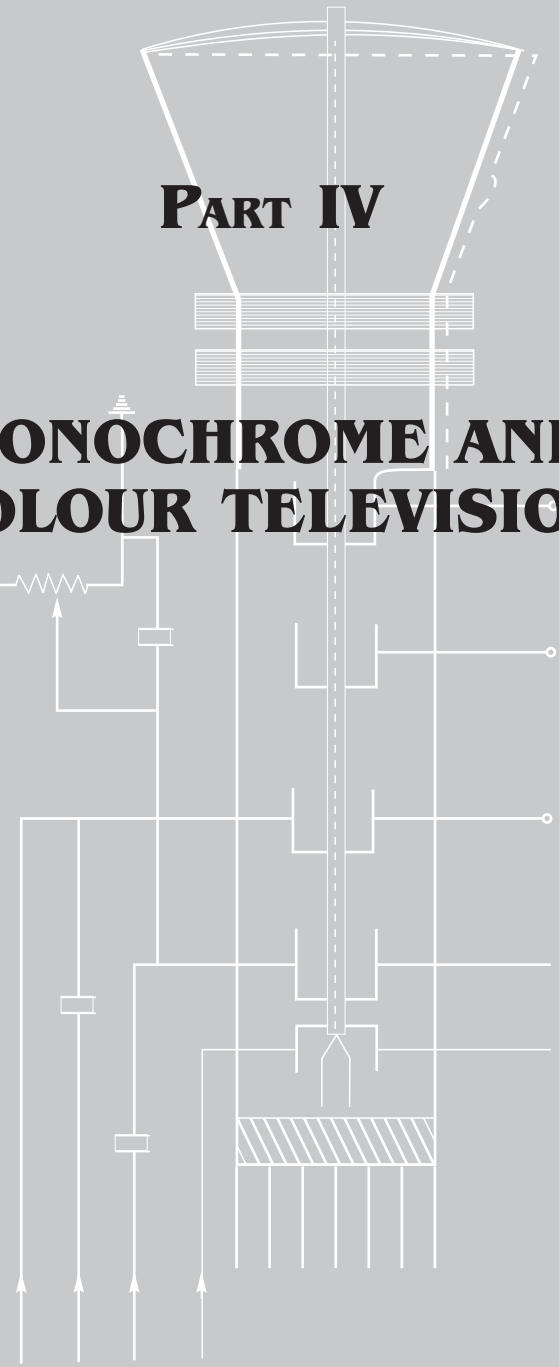
1. What are the main functions of a radio receiver?
2. Draw the circuit diagram of a crystal receiver set and explain its working.
3. What are the three main characteristics of a radio receiver? Explain the meaning of each of these characteristics.
4. What is a TRF receiver? How is selectivity obtained in a TRF receiver?
5. What is a superheterodyne receiver? What are the advantages of a superhet receiver over a. TRF receiver?
6. Draw a block diagram of a superheterodyne receiver and explain the function of each stage.
7. What is a converter stage in a radio receiver? Explain its working with a block diagram.
8. What is IF in a superheterodyne receiver? Why does a transistor receiver require more IF stages than a valve type receiver?
9. Draw the block diagram of a transistor radio receiver. What advantages does it possess over a tube type radio receiver?
10. What is AVC or AGC in a radio receiver? What stages of the radio receiver are controlled by AVC?
11. What is a multiband receiver? Explain the use of a rotary type band-change switch in a multiband radio receiver.

12. Draw the diagram of a wafer of a rotary band-change switch with two poles and three positions.
13. Explain the difference between negative peak clipping and diagonal clipping in AM receivers.
14. What is image frequency?
    Explain the steps taken for image frequency rejection in a radio receiver.
15. Give the block diagram of an FM receiver and explain its salient features.
16. What is FM radio paging service? Explain its working with a block diagram.
17. Explain with a block diagram the main features of a communication receiver.
18. State whether TRUE or FALSE with brief justification.
    (a) A crystal receiver needs no power supply.
    (b) The more the number of stages the less selective a receiver is
    (c) The higher the IF, the lower is the image frequency.
    (d) A communication receiver is more sensitive because it uses double detection.
    (e) FM radio has higher range than AM radio.

    Ans. (a) True, (b) False, (c) False, (d) True, (e) False

PART IV

MONOCHROME AND COLOUR TELEVISION



chapter 15

Basic Principles of Television

15.1 WHAT IS TELEVISION?

The word "television" is a combination of two words "tele" meaning "far" and "vision" meaning "to see". Thus, television means "seeing from a distance". *Doordarshan* is an apt Hindi translation of the word "television" which today has come to mean "viewing of distant objects or events by means of electrical transmission of radio waves".

Basically, television broadcasting, known as *telecasting* is very much similar to sound or radio broadcasting. In radio broadcasting sound waves are converted into electrical signals by a microphone and these electrical signals are transmitted through space as modulated radio carrier waves. On reception at the distant receiving end, the electrical signals are separated from the carrier waves by an ordinary broadcast receiver and converted into audible sound waves by a loudspeaker. In television, light signals from the object being televised are converted into electrical signals by a television camera and transmitted to distant points by radio carrier waves. The television receiver separates the television signals from carrier waves and converts them into light signals which form a picture of the televised object on the screen of the picture tube. However, in the television system sound has also to be transmitted along with the picture. Separate carrier waves are used for the transmission of picture signals and sound signals but they are radiated by the same transmitting antenna. At the receiving end, the same receiving antenna receives both carrier waves but the television receiver converts these signals separately into sound waves which drive a loudspeaker and light waves which produce a picture on the screen of the picture tube. For the proper display of the picture and the reproduction of accompanying sound, several controlling signals have also to be transmitted. The details of these signals will be given in the following sections.

15.2 TELEVISION BROADCASTING SYSTEM

A block diagram of a complete TV system for the transmission and reception of picture and sound signals is given in Fig. 15.1.

At the TV studio, the TV camera focuses an optical image of the scene on a photosensitive plate in the camera and the picture elements of varying light intensity are converted into correspondingly varying electrical signals by a process of electronic *scanning*. The electrical signals so formed by scanning the picture image by an electric beam are called *video* signals, "Video" is a Latin word meaning "to see". At this stage, certain *synchronising* signals meant to keep the reassembly of the picture at the receiver in step with the scanning at the studios are also added to the video information. The composite *video signal* so formed is amplified by video-amplifiers and made sufficiently strong to *amplitude-modulate* a picture carrier wave which is transmitted by the transmitting antenna.

The sound picked up by the microphone is converted into electrical currents at audio frequencies (AF) and is strengthened by the audio amplifier which frequency-modulates a separate RF carrier whose frequency is generally 5.5 MHz above the frequency of the video carrier. The frequency modulated (FM) sound carrier is radiated by the same transmitting antenna as used for the transmission of the video or picture carrier. Thus, at the TV transmitting station, two separate RF carriers, one for the transmission of picture signals and the other for sound signals are radiated by a common transmitting antenna. The picture (video) carrier is amplitude-modulated (AM) and the sound carrier is frequency-modulated (FM).

At the receiving end, both the picture and sound carriers are intercepted by the same receiving antenna and passed onto a wideband circuit called the *tuner*. In the tuner two separate IFs for picture and sound signals are formed by heterodyning with a local oscillator as in a superheterodyne receiver. The picture and sound IF frequencies are amplified by a common IF amplifier and then detected by the video detector. At this stage, the sound IF of 5.5 MHz (the difference between video and sound IF from the tuner) is separated and fed into the sound channel where it is detected by a method of FM detection and the AF is amplified and fed into the speaker to produce the sound as in a normal FM receiver.

The video signal from the video detector stage is amplified by a video amplifier and is used to modulate the electron beam in the picture tube to produce a picture of the television scene. A portion of the composite video signal is also fed to a *synchronising separator* where the synchronising signals are separated from the video signal and applied to the deflection circuits to keep the electronic scanning beam in the picture tube in step with the electronic beam at the transmitter.

The above method of obtaining the sound IF (5.5 MHz) by beating or heterodyning the video and sound carriers is known as the *inter-carrier system* and is the modern system used in TV receivers. The older method, where the

sound signal is separated at the mixer stage itself and then handled separately is known as the *split sound system* and is no longer in use.

15.3 SCANNING

Scanning is the process by which the optical image of the televised object formed on the photosensitive plate of the TV camera is dissected or broken into a series of horizontal lines by an electron beam. This electron beam sweeps across each line at a uniform rate, then flies back to scan another line directly below the earlier one and so on till the horizontal lines into which it is desired to break or split the picture have been scanned in the desired sequence. After this the electron beam flies back to its original position and starts the scanning sequence again. The scanning process is repeated again and again. As the electron beam sweeps across a line, it falls over portions of different light intensities and is accordingly converted into electrical currents of different amplitudes. The higher the illumination of a particular spot on the picture, the greater is the amplitude of the corresponding electrical currents produced by the TV camera. In this way current pulses are produced which correspond in time sequence to bright and dark areas of the televised picture as they are scanned by the electron beam. This electrical signal which corresponds to variations of illuminations in the televised scene is the video signal which is used to modulate the picture carrier for transmission to distant places.

At the receiving end also, a similar electronic beam traces out horizontal lines on the fluorescent screen of the picture tube and by horizontal and vertical scanning produces a uniformly lit rectangular area called the *raster*. When the scanning electron beam of the TV receiver is modulated with the video signal received from the transmitter, the raster is converted into a picture. In order that the picture formed at the picture tube corresponds to the televised scene, it is necessary that the scanning at the transmitter is completely in step or in synchronisation with the scanning at the TV receiver. This is achieved by sending certain command signals called *synchronising signals* or simply *sync signals* along with the video signal. Any time the sync signals are not properly received or applied to the scanning electron beam in the picture tube, the picture gets distorted and changes into patterns of dark and bright horizontal lines. We then say that the sync is out.

The scanning process in TV is very much similar to reading the page of this book. The eye starts reading at the upper left-hand corner and travels to the right across the first line of words. At the end of the first line, it quickly returns to the left-hand side of the page and starts reading the second line and so on. This process is repeated for each line, till the end of the bottom line of the page is reached when the eye returns to the top and starts the same process for the next page.

If it is required to describe the page to somebody on the telephone for preparing a copy, then besides just reading out the wording in the manner described above, it will also be necessary to announce the beginning and end

of each line and each page to enable the person at the other end of the telephone line to prepare an exact copy of the book, word by word, line by line and page by page. These announcements or signals will then in a way correspond to the sync signals in TV scanning.

15.3.1 Persistence of Vision

A TV picture cannot be transmitted as a whole like the motion picture where the projector enables the complete picture to be projected on the cinema screen by optical means. It, therefore, becomes necessary to transmit the TV picture by scanning it bit by bit and line by line and assemble it at the picture tube again by the same process. Even with this method of transmitting and assembling the TV picture piecemeal, we get the impression of completeness and continuity due to a phenomenon called the *persistence of vision* of the human eye. It is a property of the retina of the human eye that any impression produced on the retina by a light ray will persist for a fraction of a second even after the light source is removed. If within this short interval of persistence of vision, which is generally about 1/16th of a second, a series of images are presented to the eye, the eye will see all the images without any break and will get the impression of continuity. When the electron beam strikes the face of the picture tube at a particular point, this point continues to glow for a short period even after the beam has moved to the next point and the persistence of vision makes it possible to televise the picture element by element and when these elements are scanned rapidly enough, they appear to the eye as one complete picture.

In motion pictures the persistence of vision helps create and illusion of continuous motion when the picture is projected at a repetition rate of 24 picture frames per second. This repetition rate, however, produces *flicker* when one picture does not completely blend into the other. This flicker effect is considerably reduced by using a rotating shutter to present each picture frame twice which virtually makes the picture repetition rate 48 frames per second and thereby eliminate flicker.

In TV, the persistence of vision not only helps produce a complete picture from separate elements but also helps create the impression of continuous motion the same way as is done in motion pictures. The picture repetition rate used in TV is 25 per second in India (30 per second in America).

15.3.2 Flicker

For reducing flicker in television pictures, a special type of scanning called *interlaced scanning* is used.

15.3.3 Interlaced Scanning

The scanning process described earlier in which the scanning electron beam sweeps across each horizontal line in regular succession from top to bottom of the picture is called *progressive scanning*. For continuity of motion each picture frame is scanned 25 times per second in India (30 times per second in

America). The total number of horizontal lines into which a picture frame is divided depends on the standard adopted. In India, the number of lines per picture is 625 compared to 525 lines in America. This means that the total number of lines scanned per second in India is $625 \times 25 = 15625$ lines ($525 \times 30 = 15750$ lines in America).

The repetition rate of 25 pictures per second in television again creates the problem of flicker. This problem is solved by a special method of scanning called interlaced scanning as shown in Fig. 15.2.

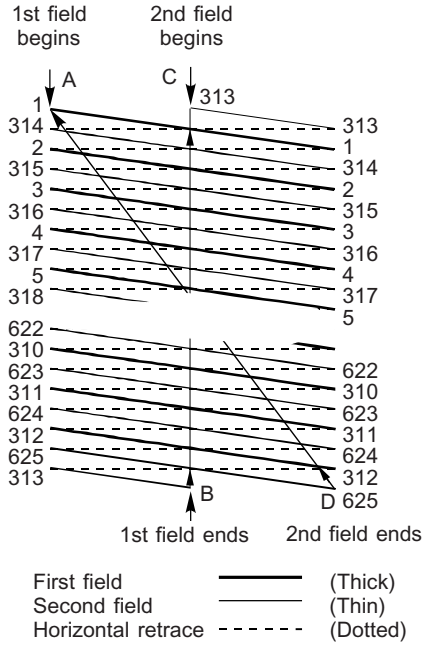


Fig. 15.2 Interlaced Scanning

In interlaced scanning all the 625 lines are not scanned at a stretch but the scanning process is divided into two stages called fields. Each field will contain only half the total number of lines contained in one frame (i.e. $312\frac{1}{2}$ lines). The scanning beam which now moves down at double the rate will scan only alternate lines.* The scanning process starts at *A* when the first set of $312\frac{1}{2}$ lines is scanned sequentially in the first (odd) field. Since each field contains a half line, the scanning of odd lines will end at *B* when the beam will suddenly flyback to point *C* and start scanning the even set of lines in the second (even) field which will end at *D*. From here the beam will flyback to *A* and the scanning process will start all over again. It can be seen that each picture frame

* The scanning lines on the screen appear slanting because the scanning beam slowly moves down while sweeping horizontally across the screen.

has been divided into two fields of $312\frac{1}{2}$ lines each, thereby making the picture repetition rate double the frame repetition rate, i.e. 50 fields per second, whereas the number of lines scanned per second remains the same. Thus, in interlaced scanning the problem of flicker is overcome without increasing the bandwidth which goes up with the increase of total lines scanned per second. The interlaced scanning used in India consists of 25 frames and 50 fields per second. In America it is 30 frames and 60 fields per second.

In brief, the frequency of horizontal scanning is 15625 Hz and the frequency of vertical scanning is 50 Hz. In other words, the horizontal and vertical scanning oscillators must oscillate at these frequencies.

Synchronisation For proper reproduction of the televised picture at the receiving end, it is essential that the scanning sequences at the transmitter (camera) and receiver are completely in step with each other. In other words, the scanning of each line and each field should start as well as finish simultaneously at the transmitter and TV receiver. The process by which the horizontal and vertical sweeps at the camera and TV receiver are kept in step with each other is known as *synchronisation.* This is achieved by sending sync signals from the transmitter. These timing signals are in the form of rectangular pulses used to control both the transmitter and receiver scanning.

Separate sync signals are required for the horizontal or line sweep and vertical or frame sweep. Since the frequency of horizontal line scanning is 15625 Hz, this will also be the frequency of the horizontal sync signals. Similarly, the frequency of the vertical sync signals will be the same as the field-scanning frequency which is 50 Hz. To distinguish the line sync signals from vertical or field sync signals at the receiver, their duration is kept different. Whereas the duration of the line sync pulse is about 5.8 μs, the duration of each field sync pulse is about 26 μs. To eliminate the difference between odd and even fields and to help maintain proper interlace between the two fields, *equalising* pulses are also transmitted before and after the frame sync signals.

Blanking As explained earlier under the section on scanning, at the end of each field, the scanning beam quickly returns to start the scanning of another line or field. The path followed by the returning electron beam is called *retrace*. The retrace is made invisible by a process known as *blanking.*

The scanning electron beam is suppressed or cut off during the retrace period by *blanking pulses* transmitted with the picture. Horizontal blanking pulses at 15625 Hz blank out the retrace from right to left for each line and vertical blanking pulses at 50 Hz will blank out the retrace from bottom to top, for each field. The scanning synchronisation is so arranged that the retraces will occur during the blanking time only.

15.3.4 Video Signal

For the transmission of TV pictures by electromagnetic or radio waves, the optical image of the object must first be converted into an electrical image.

This is done by the TV camera. The optical system of the camera focuses the image of the object on a photosensitive plate where, an electric charge pattern corresponding to the light image of the object is formed. The electric charge pattern or the electric image so formed is then scanned by the scanning process already described to convert the charge pattern into an electric current whose instantaneous value corresponds to the amount of light falling on the area being scanned. The picture information obtained by scanning in the TV camera, also known as the video signal, is transmitted as a modulation of the picture RF carrier. The video signal is not transmitted alone but is accompanied by a number of other signals or pulses which help in the proper reconstruction of the picture at the receiving end, these pulses are blanking pulses, line synchronising (horizontal) pulses, field synchronising (vertical) pulses, equalising pulses, etc. as described earlier. A combination of picture signals and all the controlling signals or pulses mentioned above is called *a composite* video signal.

For the formation of a composite video signal, a blanking signal is imposed on the electron beam at the end of each horizontal line. This completely cuts off the electron beam from reaching the photosensitive plate at the TV camera or the fluorescent screen at the receiver. A synchronising pulse, which is of a shorter duration than the blanking pulse is then made to ride on top of the blanking pulse as shown in Fig. 15.3. The horizontal sync pulse starts the retrace process and causes the position of the scanning beam to shift from the right to the left of the picture. A fraction of a second later, the blanking pulse releases its hold on the picture tube and the electron beam starts scanning again.

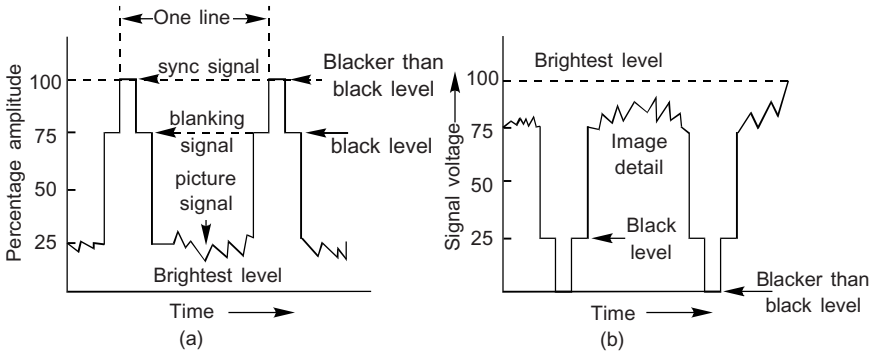


Fig. 15.3 Video Signal (Horizontal Scanning) (a) Negative Transmission, (b) Positive Transmission

Figure 15.3 indicates the relative amplitudes of the picture signal, blanking signal and sync signal in the case of horizontal scanning. Of the 100% amplitude available, 75 to 80% is used for picture information. The blanking signal voltage is introduced at a level which completely cuts off the electron beam in the picture tube and no light is produced on the picture screen. This level is called the *black level.* The sync pulse which has an amplitude even greater

(100%) than the blanking voltage takes the picture tube even below the cut-off region and this is known as the *blacker-than-black level.*

Figure 15.3(a) shows that the brightest portions of the picture correspond to the least amount of current flow. This is, of course, the reverse of what happens at the camera tube where the brightest portions of the scene produce the maximum current. The blanking voltage, which should be more negative than any portion of the camera signal is actually more positive and the synchronising signals produce the largest current value. This type of video signal transmission is called *negative picture transmission* or *negative modulation*. The video signal is in the reverse phase. It is like the negative of a photographic film in which bright portions appear dark and vice versa. The video signal must be reversed in phase before it is applied to the picture tube in the TV receiver. Figure 15.3 (b) is a representation of the *positive picture transmission* in which the phase of the video signal has been reversed. Negative modulation adopted by India has the advantage that visible interference on the picture tube screen produces black spots which are less annoying than the white blobs in positive modulation system.

As far the vertical scanning, the scanning beam has to be shifted quickly to the top of the picture after it has completed scanning the last line in a particular field to enable it to start scanning another field. Since the distance to be travelled is greater, a longer blanking signal is required to avoid the vertical retrace being shown on the screen. The vertical synchronising pulses again ride on top of the vertical blanking pulses. To avoid the horizontal oscillator control getting out of step during the long duration of the vertical pulses, the vertical pulse is broken up into smaller intervals called *serrations*. With the serrated vertical pulse, both horizontal and vertical synchronising actions can go on simultaneously. The waveforms of the two sync pulses being different, they can be easily separated in the TV receiver.

15.3.5 Equalising Pulses

In interlaced scanning, the vertical pulse is inserted into the video signal once when a horizontal line is half completed (first field) and again at the end of a complete line (second field). In order that the vertical pulse oscillator receives the sync voltage at the same time after every field, a series of 5 *equalising pulses* is inserted into the signal immediately before and after the vertical sync pulses and are known as *pre-equalising pulses* and *post equalising pulses* respectively.

A composite video signal (positive picture transmission) is shown in Fig. 15.4.

15.3.6 Aspect Ratio

The ratio of the width to height of the picture frame in a TV is called the *aspect ratio*. The standardised aspect ratio is 4:3 which means that the width of the picture will be 4/3 or 1.33 times larger than the height. Making the frame wider

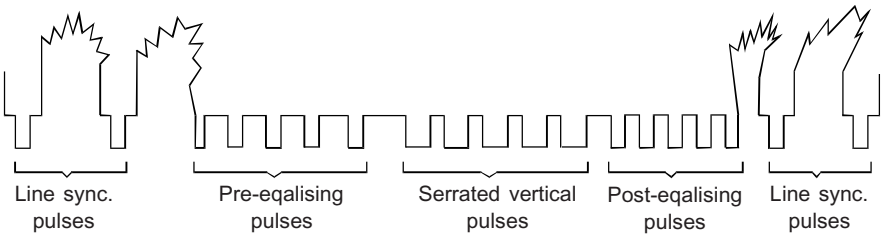


Fig. 15.4 A Composite Video Signal

than the height allows for motion in the scene which is usually in the horizontal direction. The aspect ratio only fixes the proportions and the actual size can be anything so long as the correct aspect ratio of 4:3 is maintained. For example, for a picture frame which is 80 cm wide, the height should be 60 cm for the correct aspect ratio to be maintained. If the correct aspect ratio is not maintained in the picture frame, the persons in the picture will look either too thin or too broad.

15.4 BANDWIDTH REQUIRED FOR TV SIGNALS

As already explained, the conversion of light into electrical signals is done by the TV camera at the transmitter. As the scanning beam moves over the photosensitive plate, the portions affected by strong light produce higher current amplitudes as compared to the portions affected by weak light. If the photosensitive plate is broken up into a series of white and black squares called *elements* as in Fig. 15.5 (a), the scanning beam will produce a pulse of current as it passes over a white element and this current will drop down to zero as the scanning beam moves over a dark element. The waveform of the current obtained by the scanning of one horizontal line is shown in Fig. 15.5 (b). One maximum point combined with its successive minimum point will constitute one complete cycle as in the case of the sine wave shown in Fig. 15.5 (c).

Since one pair of elements (white and dark) produces one cycle of the electrical current, the number of cycles produced in one second or the fundamental frequency of the camera signal will be given by

$$\text{Fundamental frequency} = \frac{\text{Total no. of picture elements scanned/s}}{2} \qquad (15.1)$$

The fundamental frequency, which will decide the bandwidth of the video signal can be calculated as follows: For equal resolution of the picture, both in the vertical and horizontal directions, the picture elements in the horizontal and vertical directions should be of the same width, but with an aspect ratio of 4 : 3, the number of picture elements in the horizontal direction will be 4/3 times the number of picture elements in the vertical direction. As shown in Fig. 15.5(a), the number of picture elements in the vertical direction is nine but the number of picture elements in the horizontal direction is $9 \times 4/3 = 12$. The width of a picture element in the vertical direction is the distance between two

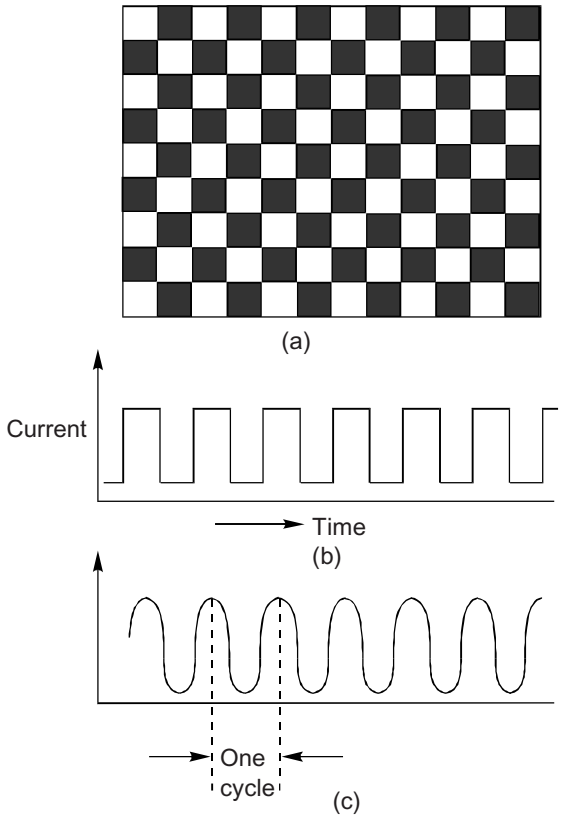


Fig. 15.5 Video Signal Bandwidth Requirement (a) White and Black Elements on Photosensitive Plate (b) Camera Signal, (c) Sine Wave

successive horizontal lines. Therefore, the number of picture elements in the vertical direction will be equal to the number of horizontal lines in one picture frame and with an aspect ratio of 4:3, the number of picture elements in the horizontal direction will be 4/3 times the number of lines in one picture frame. If n is the number of lines in one picture frame, the total number of picture elements scanned in one picture frame is given by $n \times 4/3\ n = 4/3\ n^2$. Assuming the number of picture frames scanned per second to be m, the total number of picture elements scanned per second will be $4/3\ n^2.m$.

As per Eq. (15.1), the fundamental frequency of the picture signal will be:

$$\text{Fundamental frequency} = \frac{4/3n^2 \cdot m}{2} \tag{15.2}$$

Taking the practical example of the scanning system followed in India, i..e. 625 lines per frame and 25 frames per second, we have in Eq. (15.2).

$$n = 625 \quad \text{and} \quad m = 25$$

The highest frequency of the video signal is

$$\frac{4}{3}(625)^2 \times \frac{25}{2} = 6.5 \text{ MHz (approx.)}$$

Thus in the system followed in India, a bandwidth of about 7 MHz is theoretically required for the transmission of the picture signal. It may, however, be mentioned that the effective number of lines per frame is less than 625 because of the lines lost during the vertical blanking periods and certain other factors. The highest video frequency in the signal works out to be 5 MHz which provides adequate bandwidth for the video signal.

15.4.1 Factors Governing Bandwidth of TV Signals

Two factors that must be taken into consideration for working out the effective number of lines in a frame are, the number of lines lost in the vertical blanking period and the kell factor, as discussed below.

(a) Number of Lines Lost In the 625 line scanning system the total number of line per frame is

$$625 \times 25 = 15625$$

Scanning time for one horizontal line

$$= \frac{1}{15625} = 64\ \mu s.$$

Out of this normal time scanning of 64μs, 52μs is the actual scanning time for a line and 12μs is the line blanking period as shown in Fig. 15.6.

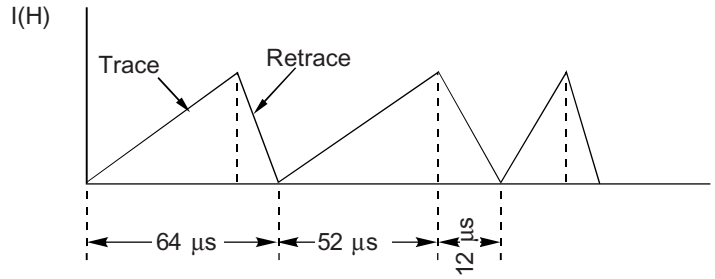


Fig. 15.6 Horizontal Scanning

Vertical scanning period per field is

$$\frac{1}{50} \text{ seconds} = 20 \text{ mili seconds}$$

As per the standards laid down for this type of scanning, this period of 20 ms is again divided into two parts. The actual time taken by the scanning beam to travel from top to bottom of the field is 18.720 ms and the remaining period of 1.280 ms is the time taken by the beam to flyback to the top to commence the next cycle. This is called the blanking period as shown in Fig. 15.7

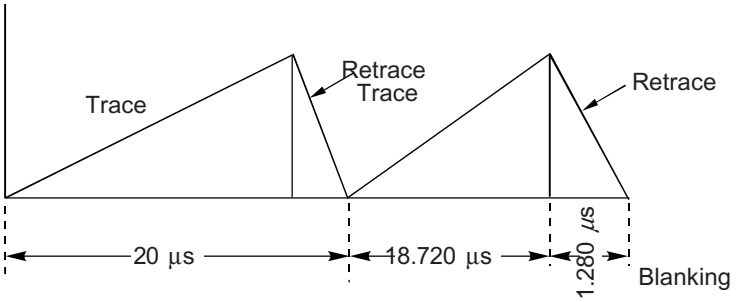


Fig. 15.7 Vertical Scanning

No. of lines traced during the retrace period = $\frac{1.28 \times 10^{-3}}{64 \times 10^{-6}} = \frac{1280}{64}$ = 20 lines

No. of lines traced in the two fields scanned in the interlaced scanning system is 20 × 2 = 40 lines

These lines will not be visible as these are traced during the blanking period and will not contribute towards picture formation on the screen. Therefore the no. of active lines available per frame is 625 – 40 = 585 lines instead of the actual number of 625 lines used per frame.

The number of lines scanned in the vertical direction is the vertical resolutions.

(b) Kell Factor In the earlier calculations, we have assumed that the picture elements are equally spaced and that there is a uniform distribution of light all over the frame. Actual distribution of light is not uniform but depends on the nature of the picture. Statistical analysis and objective tests show that only about 70% of the total lines are actually scanned and the rest 30% get merged into each other due to the finite width of the scanning beam which does not fall equally on two consecutive lines. Thus the actual number of lines resolved or traced is obtained by multiplying the original number of lines by a factor known as the *kell factor* or the resolution factor. The value of kell factor (k) lies between 0.65 and 0.75. Different systems assume different values for the kell factor. In our case, if we assume $K = 0.7$, the effective number of lines $Ne = 585 \times 0.7 = 409.5 \approx 410$.

If the horizontal and vertical resolution is to be the same, the total no of lines scanned will be $\frac{410 \times 4}{3} \approx 576$

Fundamental frequency developed by scanning $= \frac{576}{2} = 273$ cycles/s

Total time for covering these cycles = 52 μs

Time for each cycle (T) is given by,

$$(T) = \frac{52 \times 10^{-6}}{273}$$

and the frequency of the square wave developed

$$\frac{1}{T} = \frac{273}{52 \times 10^{-6}} = 5.2 \text{ MHz} \approx 5 \text{ MHz}$$

The above value of the highest frequency f_h can be found by the formula

$$f_h = \frac{\text{No. of activeness} \times \text{aspect ratio} \times \text{kell factor(k)}}{2 \times \text{time for tracing one line}}$$

ln the case of 625 line system

$$f_h = \frac{585 \times \frac{4}{3} \times 0.70}{2 \times 52 \times 10^{-6}} \approx 5 \text{ MHz}$$

The American system which uses only 525 scanning lines per frame, the highest frequency developed is 4 MHz. This explains the fact the allocation of Channel bandwidth in the case of American system (525/60) is only 6 MHz compared to a bandwidth allocation of 7 MHz in the case of 625/50 system used in India.

Example: Calculate the maximum frequency developed in the American 525/60 TV system, if the number of active lines is 485 and the duration of one active line is 57μs.

As per the above formula

$$f_h = \frac{485 \times \frac{4}{3} \times 0.70}{2 \times 57 \times 10^{-6}} \approx 4 \text{ MHz}$$

15.5 VESTIGIAL SIDE BAND SYSTEM

If the picture signal of 5 MHz bandwidth is allowed to amplitude modulate the picture RF carrier, the upper and lower side bands produced will be of frequencies equal to the carrier plus 5 MHz and the carrier minus 5 MHz. The total bandwidth occupied by this amplitude modulated wave will be 10 MHz. In addition, some bandwidth will also be required for the transmission of the sound signal. With a bandwidth of more than 10 MHz for each channel, the number of channels that can be accommodated in a frequency band will be considerably reduced. Since both the side bands contain the same information, the above difficulty can be solved by producing a modulated carrier with only one complete sideband, the other sideband having been partially but not wholly suppressed by a specially designed filter. Such a system of transmission is termed as the *vestigial* or asymmetrical sideband system and is the standard type of transmission system used in TV transmitters. This system has the advantage of requiring much less total bandwidth than does the amplitude modulated wave system with its two full side bands.

A frequency spectrum of a complete TV channel employing the vestigial side band system is shown in Fig. 15.8.

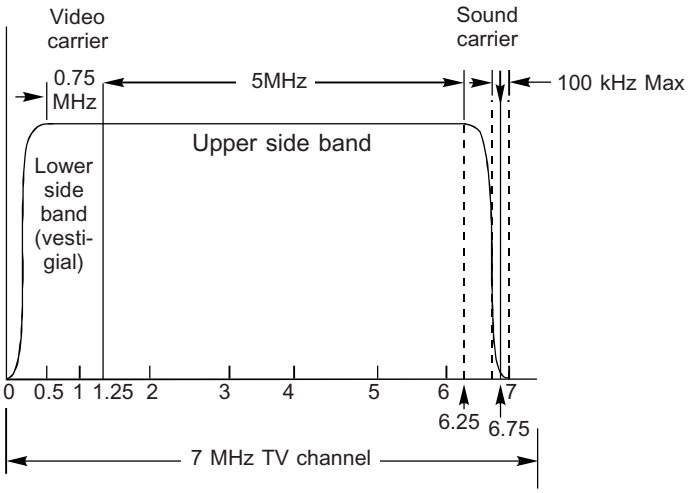


Fig. 15.8 A Standard TV Channel

Taking the limit of the lower side band as zero frequency, the vestigial side band extends up to 1.25 MHz, where the *video carrier is* located. The upper video side band extends at full amplitude up to 6.25 MHz and with reduced amplitude up to 6.75 MHz. The sound carrier is located 5.5 MHz above the video carrier with a centre frequency of 6.75 MHz and has two side bands extending up to a maximum of 100 kHz. For this small bandwidth, narrow band FM with a maximum frequency deviation of ±25 kHz is used.

15.6 TELEVISION BANDS AND CHANNELS

Television signals are radiated at frequencies much higher than those used for radio broadcasts. Frequencies above 40 MHz are used for broadcasting television signals. The frequency bands that have been assigned for the use of TV stations are as follows:

Band I	41-68 MHz	Known as VHF (Very High Frequency) band
Band III	174-230 MHz	Also a VHF band
Band IV	470-790 MHz	Known as UHF (Ultra High Frequency) band

Band II 88-108 MHz for FM broadcasts is not used for TV broadcasts. Only Band I and Band III are used for TV transmitters in India. Each band is divided into a number of channels. According to the standards adopted in India, a channel is 7 MHz wide, whereas the American standards make use of 6 MHz wide channels. The low frequency end of the channel is used for the AM picture carrier and the HF end of the channel is used for the FM sound carrier. The separation between the vision carrier and the sound carrier in India is 5.5 MHz. To take an example, the Delhi TV station operates in channel 4 of Band I with a picture carrier of 62.25 MHz and sound carrier of 67.75 MHz. The same channel can be shared by a number of stations provided these stations are sufficiently wide apart to avoid any interference between them. The allocation of TV channels to Indian stations in the two frequency bands used in India is given in Table 15.1.

Table 15.1 Allocation of TV Channels to Indian Station (Courtesy: Doordarshan)

Band	*Channel*	*frequency*	*Picture carrier MHz*	*Sound carrier MHz*	*Allotted to*
I	1	41-47	Not used for TV		—
	2	47-54	48.25	53.75	—
	3	54-61	55.25	60.75	—
	4	61-68	62.25	67.75	Delhi, Bombay, Calcutta, Madras, Srinagar, Lucknow
III	5	174-181	175.25	180.75	Pune, Kanpur, Jaipur, Bangalore
	6	181-188	182.25	187.75	Trivandrum, Muzzaffarpur
	7	188-195	189.25	194.75	Amritsar, Gulbarga, Sambhalpur, Ahmedabad, Madurai, Asansol, Panaji
	8	195-202	196.25	201.75	—
	9	202-209	203.25	208.75	Jullundur
	10	209-216	210.25	215.75	Mussoorie
	11	216-223	217.25	222.75	—
Additional Channel.	12	223-230	224.25	229.75	—

Summary

Television means seeing from a distance and doordarshan is an apt translation of the word television.

Television broadcasting or telecasting is similar to Radio broadcasting except that in Television both sound and picture signals are transmitted through separate carrier waves and received by a single receiving antenna. These are reconverted into sound and video signals with the help of a loudspeaker and picture tube.

Scanning is the process by which the televised scene is converted into electrical signals for transmission and then reconverted into the picture with the help of the picture tube. Interlaced scanning is used to avoid flicker in the picture. In interlaced scanning each frame is divided into two fields with half the number of lines ($312^1/_2$) in the Indian TV. Synchronising pulses are used both for the line scanning and the field scanning. To avoid visible retrace on the screen, blanking pulses are used both for lines and the fields.

With 625 lines and 50 fields, the fundamental frequency of the video signals works out to be about 6.5 MHz. However, keeping in view the kell factor and the lines lost during vertical blanking the actual frequency required is about 5 MHz. Double side band will need 10 MHz. To conserve bandwidth vestigial sideband is used and with this system a standard TV channel for Indian TV will be 7 MHz wide.

Review Questions

1. Explain the difference between Radio Broadcasting and TV Broadcasting. Why are two separate RF carrier waves required for TV broadcasting?
2. What is composite video signal? Draw the block diagram of a composite video signal with negative modulation and explain its working.
3. What is interlaced scanning in TV? What are its advantages over progressive scanning?
   Calculate the horizontal and vertical frequency of interlaced scanning in the following systems.
   (i) 405 lines, 50 fields/s
   (ii) 525 lines, 30 frames/s and
   (iii) 625 lines, 25 frames/s
   (*Ans.* (i) 10125 Hz, 50 Hz, (ii) 15750 Hz, 60 Hz (iii) 15625 Hz, 50 Hz).
4. Explain the terms Raster and Aspect ratio as applied to television.
5. What are sync pulses and blanking pulses? Describe the use of these pulses in TV.
6. Calculate the bandwidth required for the video signal formed by the scanning system with 525 lines per picture and 30 pictures per second. The aspect ratio is 4 : 3 and Kell factor = 0.7. (*Ans.* 4 MHz)
7. What is flicker? How is it removed in TV and in cinema.
8. Explain the use of Vestigial Sideband Transmission in TV. How much is the saving in spectrum effected by use of vestigial side band system.
9. Draw a neat sketch of a 7 MHz TV channel and explain its working.

chapter 16

Colour Television—Fundamentals

16.1 INTRODUCTION

Reproduction of a televised scene in colour is like viewing the object directly in its natural colours. Nature abounds in colour and only colour television can reproduce natural colours on the TV screen. It is the naturalness of colour TV that appeals to the human eye more than the black-and-white or monochrome TV. Although the black-and-white TV can provide sufficient information and has full entertainment value, it has its limitations and fails to be natural and true to life. The originality and contrast provided by colour television makes it particularly suitable for televising educational programmes like surgical operations, and programmes on art exhibitions in picture galleries. It is because of these obvious advantages that more and more countries are switching over to colour television in spite of the system being a little more expensive than the black-and-white system.

The colour television system is very similar to the monochrome system except for a few additional controls and circuits for the colour section. The colour TV camera produces three video signals corresponding to the red, blue and green primary colours. The three video signals are transmitted as modulation of the picture carrier in the standard TV channel. The colour TV receiver merely separates out the three vedio signals to intensity modulate three electron beams in a tri-gun colour picture tube. The three electron beams strike a specially coated phosphor screen to produce the red, blue and green colours which are mixed together or integrated by the human eye to reproduce the natural colours of the televised scene.

The technique employed for modulating the picture carrier with the three video signals is such that the modulated signal can be received by a black-and-white TV receiver to produce a monochrome picture without any circuit modifications. This is known as *compatibility*.

16.2 COMPATIBILITY

Compatibility is the phenomenon by which a colour television system produces a normal black-and-white picture on a monochrome receiver without any modifications to the existing circuitry.

Also, a normal monochrome transmission system should be able to produce a black-and-white picture on a colour television screen. This is known as *reverse compatibility*.

For complete compatibility between the colour television system and the monochrome system, the colour TV system must satisfy the following conditions:

1. It must use the same 7 MHz standard TV channel.
2. It must have the same bandwidth of 5.5 MHz.
3. The location and separation of the sound and picture carriers in the standard channel must be the same for the two systems.
4. It must employ the same line and frame synchronising pulses.

It will thus appear that the requirements of compatibility impose certain limitations on the colour TV system to make it more complicated than the monochrome system. However, compatibility is necessary because colour television was developed after the black-and-white system was fully in vogue. In fact, colour television would have been much simpler if we had started with colour television right from the beginning instead of changing over from monochrome to colour TV system.

16.3 PROPERTIES OF COLOURS

To understand colour TV, it is necessary to know something about colours and their fundamental properties.

It is a well known fact that light rays are electromagnetic waves whose properties are governed by their frequency or wavelength. Light waves form only a small part of the spectrum of electromagnetic waves shown in Fig. 16.1. The entire spectrum which extends from about 10^5 to 10^{25} Hz is divided into sections like radio waves, infrared waves, ultraviolet rays, X-rays and cosmic rays. Between infrared rays and ultraviolet rays there is a range of frequencies centered around 5×10^{14} Hz which is known as the *visible spectrum*. Each frequency or wavelength in this range produces a visible effect on the human eye and is perceived by the eye as a definite *hue* or *tint*. When all the wavelengths from the visible spectrum reach the human eye simultaneously, the eye sees *white light*. If, however, a part of the visible range of the frequency spectrum is flitered out and only the remaining part reaches the eye we see colour.

White light is actually a combination of seven different colours. These are labelled as Violet, Indigo, Blue, Green, Yellow, Orange and Red (VIBGYOR). Accordingly, a beam of white light can be split into these seven colours when passed through a prism. These colours are also seen as rainbow colours when

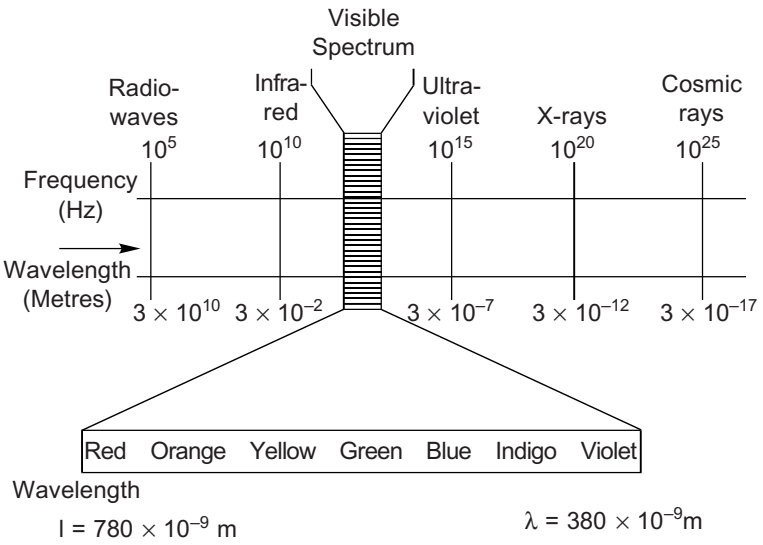


Fig. 16.1 Spectrum of Electromagnetic Waves

white light is split by refraction through rain drops. In Fig. 16.1, the visible part of the electromagnetic spectrum has been expanded to show the range of colours it contains. These colours, known as spectral colours, correspond to a definite frequency or wavelength to which the eye is responsive.

When white light falls on an object, it absorbs certain wavelength and reflects others. The colour of the object is the colour of the unabsorbed light reflected by it. This is the reason why an object seen in artificial light appears to possess a different colour from what it shows under sunlight. In the case of transparent objects like glass, it absorbs certain wavelengths and transmits others. It acts like a band-pass filter. The colour of a transparent object is the colour of the light it transmits.

16.4 PRIMARY COLOURS AND COLOUR MIXING

It is now clear that white light is a mixture of several spectral colours which are distinctly seen by the eye as red, orange, yellow, green, blue, indigo and violet. The three spectral colours represented by red, blue and green (RBG) wavelengths are known as *primary colours* because these can be combined or added together in different proportions to produce almost all other colours perceived by the eye. For example, white light can be produced by mixing 30% red, 59% green and 11% blue. Similarly, yellow, magenta and cyan can be produced by suitably mixing the primary colours as shown below.

All these results are summed up in Fig. 16.2 (a). Red, blue and green are also called *additive primaries* because of their property of producing a wide range of colour mixtures. In fact, it is this additive property of the three primaries that is made use of in producing a coloured picture on the TV screen. Red,

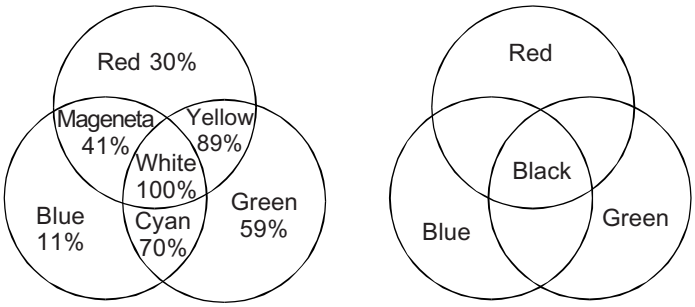


Fig. 16.2 Mixing of Primary Colours (a) Additive Mixing (b) Substractive Mixing

blue and green colours produced by the tri-colour phosphor screen are integrated by the human eye to reproduce all the natural colours of the televised scene.

$$\text{Red} + \text{blue} + \text{green} = \text{white}$$
$$\text{Red} + \text{green} = \text{yellow}$$
$$\text{Red} + \text{blue} = \text{magenta}$$
$$\text{Blue} + \text{green} = \text{cyan}$$

Another property attributed to the primary colours is that no two primaries can be combined to produce the third primary. White light can also be produced by mixing a primary colour with some other colour known as its *complementary* colour. Thus, yellow added to blue can produce white. So yellow is the complementary colour for blue. Similarly, magenta is complementary colour for green and cyan is complementary to red. Complementary colours are also called *subtractive primaries*. Any of these colours can be produced by subtracting from white light its complementary colour through a colour filter as shown below:

$$\text{Yellow} = \text{white} - \text{blue}$$
$$\text{Magenta} = \text{white} - \text{green}$$
$$\text{Cyan} = \text{white} - \text{red}$$

Figure 16.2(b) shows the result of subtractive mixing of primaries. Additive mixing as well as substractive mixing are both useful in colour television.

16.4.1 Colour Specifications

Any colour can be completely described or specified by three characteristics known as *hue or tint, saturation* and *luminance.*

Hue or tint is the actual colour seen by the eye. Red, green, blue, yellow, magenta represent different hues in the visible spectrum and result from the effect produced on the eye by wavelengths corresponding to these colours.

Saturation represents the purity of a colour. It is the amount of white light that is mixed with a colour. A fully saturated colour will have no white light mixed with it. A colour which is diluted with white is said to be *desaturated.*

For example, pure red without white is a saturated colour but red mixed with white is a desaturated colour.

Luminance is the amount of light intensity or the total amount of light energy that is received by the eye irrespective of the colour of light. Luminance is also called brightness. It is because of this property that certain colours appear to be brighter than others.

Chrominance is a term used to describe all colour information except brightness. It is a combination of the hue and saturation in a colour.

As explained earlier, different colours only represent waves of different frequencies or wavelengths in the visible electromagnetic spectrum. When treated as an electromagnetic wave, a colour will have a frequency and an amplitude. The frequency corresponds to hue, the amplitude corresponds to brightness. The saturation or purity of the colour is analogous to signal-to-noise ratio of the electromagnetic wave or signal.

16.5 PRODUCTION OF COLOUR TV SIGNALS— COLOUR TV CAMERA

For televising a scene in colour, the light originating from the scene is first separated out into red, blue and green primary colours. This is done with the help of special colour filters which allow only one colour to pass through. The three primary colours are then converted into video signals with the help of a colour TV camera shown in Fig. 16.3.

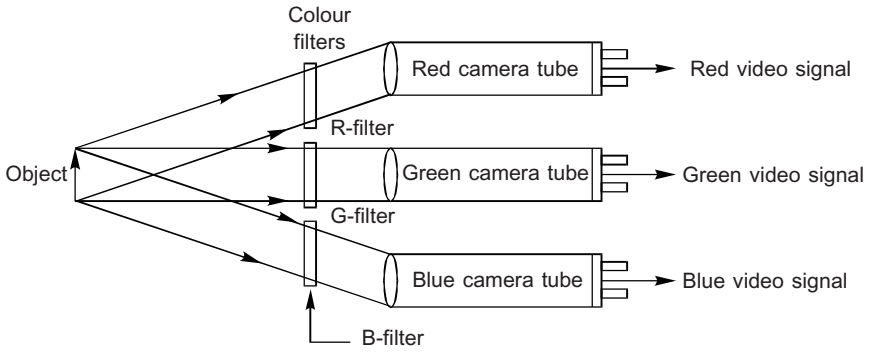


Fig. 16.3 Colour TV Camera

A colour TV camera actually consists of three camera tubes, one each for red, blue and green colours. Special colour filters are used to ensure that the red camera tube receives only red light, the blue camera tube only blue and the green camera tube receives only green light from the coloured scene. Each camera tube has its own scanning system to convert light into corresponding video voltages normally denoted by the symbols *R*, *B* and *G*. These video voltages are then combined in specific proportion or *encoded* to produce two main signals–the *luminance* signal and the *chrominance* signal.

16.5.1 The Luminance Signal

Luminance signal also known as the *Y*-signal is obtained by mixing 30% red, 50% green and 11% blue video signal from the colour TV camera as below:

$$Y = 0.30R + 0.59G + 0.11B$$

The proportion of different colours is chosen keeping in view the colour sensitivity of the eye. The eye is much more sensitive to green than to red or blue colours. Luminance signal modulates the carrier to provide compatibility by reproducing a black-and-white picture in a monochrome receiver from a colour transmission. In colour TV, it helps to decode the *R*, *B* and *G* signals for the colour picture tube.

16.5.2 The Chrominance Signal

Chrominance signal contains all the colour information. It indicates both the hue and saturation of a colour. Chrominance signal is also called the *C*-signal. Chrominance signal is obtained from what are called the colour difference signals $R - Y$ and $B - Y$. A colour difference signal can be produced by adding the *Y*-signal with its phase reversed ($-Y$) to any of the signals *R, B* or *G*. Of the three possible colour difference signals, $R - Y$, $B - Y$ and $G - Y$, only $R - Y$ and $B - Y$ are used for the chrominance signal as shown in Fig. 16.4. The third colour difference signal $G - Y$ is not required because the *G* information is already contained in the *Y*-signal.

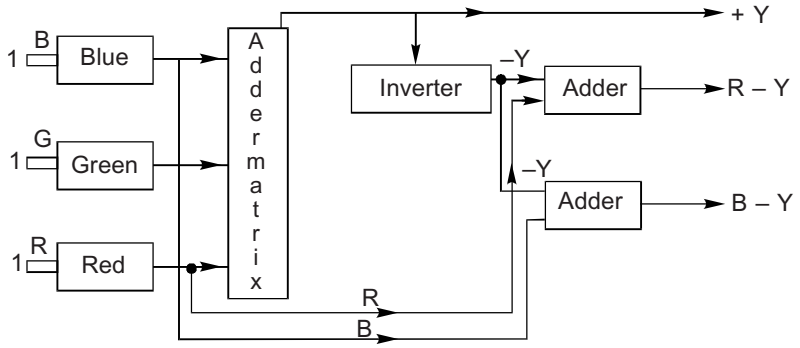


Fig. 16.4 Colour Difference Signals $R - Y$ and $B - Y$

The two colour difference signals $R - Y$ and $B - Y$ carrying all the colour information are transmitted as modulation of a colour sub-carrier (4.43 MHz).

At the receiver, the process is reversed. The *R*, *G* and *B* signals are recovered to control their respective beam currents in the tri-colour picture tube.

16.5.3 Matrix

A circuit in which signals are combined in a given proportion is called a matrix. It may consist of a network of resistors to act as an attenuator or it may be a non-inverting or an inverting amplifier.

16.6 QUADRATURE AMPLITUDE MODULATED (QAM) SIGNAL

The chrominance signal C is actually obtained by the amplitude modulation of the colour sub-carrier by the two colour difference signals $R - Y$ and $B - Y$. The colour difference signals modulate two sub-carriers which have the same frequency of 4.43 MHz (PAL system) but which are in quadrature and have a phase difference of 90°. The two modulated sub-carriers are added vectorially to produce a resultant which is the C-signal as shown in Fig. 16.5. It is this C-signal which modulates the main picture carrier in the standard TV channel. The frequency of the C-signal is the same for all TV channels. At any instant the amplitude and phase of the C-signal depends on the relative amplitudes and phases of the colour difference signals. In fact, the amplitude of the C-signal represents the saturation and its phase represents the hue of the chrominance signal. Since C-signal represents the vector sum of the two signals mutually at right angles or in quadrature, it is known as the *quadrature amplitude modulated* (QAM) signal.

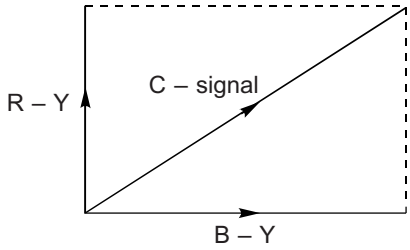


Fig. 16.5 Quadrature Amplitude Modulated *C*-Signal

16.7 COLOUR TELEVISION SYSTEMS

The three main colour television systems in use all over the world are the NTSC, PAL and SECAM. All these systems use the luminance signal and the colour difference signals to produce a coloured picture but they differ in the way the colour difference signals are used to modulate the colour sub-carrier. *NTSC* system is the earliest developed by the National Television Systems Committee (NTSC) of USA. It makes use of the colour difference signals with limited bandwidth to modulate the 3.58 MHz colour sub carrier. This system is highly sensitive to phase errors in the chrominance signal.

PAL system is the system of Phase Alternation by Line (PAL). It was developed in Germany and is very much similar to the NTSC system. In the PAL system the phase of the 4.43 MHz colour sub-carrier is reversed on every other line, thereby making the system less sensitive to phase errors.

Both the NTSC and the PAL systems employ the QAM chrominance signals. PAL system is used in most countries of the World, including India, that employ the 625-line scanning system.

SECAM (Séquential Colour á mémoire) system, which was developed in France, employs an entirely different method of modulation by colour difference signals. The system uses frequency modulation of the colour sub-carrier. Instead of both the colour difference signals modulating the colour sub-carrier simultaneously, the sub-carrier is modulated by ($R - Y$) on one line and by

($B - Y$) on the next. This avoids errors in colour reproduction caused by errors introduced in the phase of the sub-carrier during transmission.

16.8 FREQUENCY INTERLEAVING AND CHOICE OF SUB-CARRIER

The modulated picture sub-carrier containing the colour information is to be accommodated in the same standard television channel which is almost fully occupied by the luminance signal which has a bandwidth of 5 MHz. The problem is solved by making use of an interesting property of the amplitude modulated television signal.

It is found that when the picture carrier is modulated by the luminance signal at line frequency of 15625 Hz (PAL), the video signal is not a continuous one. It consists of "clusters" of energy located around the harmonics of the line frequency of 15625 Hz as shown in Fig. 16.6. Each cluster consists of a harmonic of the line frequency ($2H$, $3H$, $4H$,..., nH), surrounded on both sides

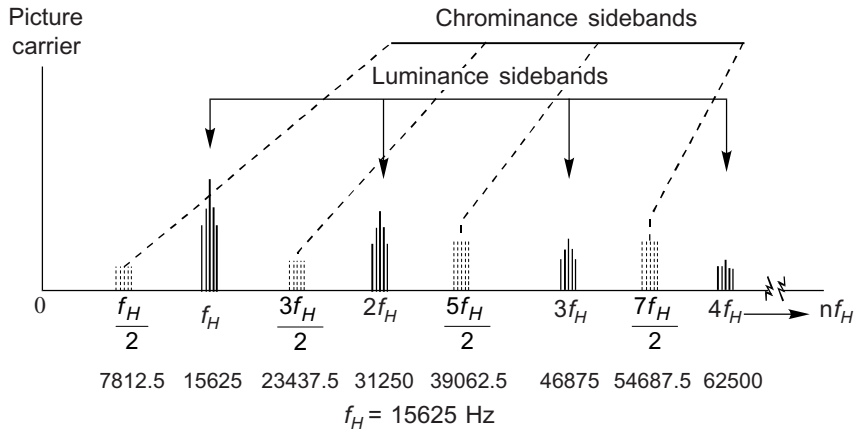


Fig. 16.6 Frequency Interleaving of Luminance Sidebands and Chrominance Sidebands

by a group of harmonics of the frame frequency, i.e. 25 Hz, 50 Hz, 75 Hz and so on. These clusters are separated by wide gaps or vacant spaces which can be utilised to accommodate the colour information. This process of accommodating information from one signal in the gaps occurring in the spectrum of another signal is called *frequency interleaving*. It is further observed that the gaps occur at odd multiples of one half the line frequency ($f_H/2$, $3f_H/2$, $5f_H/2$, etc.). If the sub-carrier frequency is chosen to be a multiple of half the line frequency, the harmonics of the sub-carrier will easily fit into the gaps existing between harmonics of the line frequency. Moreover, the luminance side-bands go on becoming weaker and weaker as they move away from the picture carrier. If the frequency of the picture sub-carrier is chosen to be near the upper end of the channel bandwidth, the strongest sidebands of the colour sub-carrier

will occur where the weakest sidebands of the monochrome luminance signal exist. This will eliminate all chances of interference with the monochrome signal and the colour signal. Based on these considerations, 567th harmonic of the half-line frequency is chosen as the sub-carrier frequency in the PAL system. The value of the sub-carrier frequency is calculated as below.

Colour sub-carrier frequency $f_{sc} = 567 \times 15625/2 = 4{,}429{,}687.5$ Hz

$= 4.43$ MHz

16.9 COLOUR SIGNAL

The total colour signal consists of luminance Y signal and the chrominance C signal. The luminance signal modulates the picture carrier directly and corresponds to the video signal in black-and-white TV system. It is formed out of the R, B and G signal voltages as follows:

$$Y = 0.59G + 0.30R + 0.11B$$

The colour sensitivity of the human eye has been kept in view while choosing the colour proportions for the Y-signal.

As regards the chrominance signal, it consists of two colour difference signals $R - Y$ and $B - Y$ which modulate the colour sub-carrier (4.43 MHz) in quadrature. The $R - Y$ and $B - Y$ colour-difference signals are known as V-signal and U-signal respectively in the PAL system. These colour difference signals can be obtained by adding $-Y$ to R and B signals as follows:

$$V = R - Y = R - (0.59G + 0.30R + 0.11B)$$

$$= 0.7R - 0.59G - 0.11B$$

$$U = B - Y = B - (0.59G + 0.30R + 0.11B)$$

$$= 0.89B - 0.59G - 0.30R$$

The U-signal modulates the in-phase sub-carrier whereas the V-signal modulates the sub-carrier which is in quadrature (90° phase difference). The resultant vector represents the chrominance signal both in amplitude and phase.

In the NTSC system, the $R - Y$ and $B - Y$ colour-difference signals are substituted by I and Q. The I-signals is the in-phase signal and Q-signal is the quadrature signal. The I-and Q-signals are formed out of colour difference signals as below:

$$I = 0.74\ (R - Y) - 0.27\ (B - Y)$$

$$= 0.60R - 0.28G - 0.32B$$

$$Q = 0.48\ (R - Y) + 0.41\ (B - Y)$$

$$= 0.21\ R - 0.52G + 0.31B$$

It is clear that I-and Q-signals can be formed directly from R, B and G outputs from the respective camera tubes. Also, compared to standard white in which $R = B = G = 1$, the luminance of these signals is zero because

$$I = 0.60 - 0.28 - 0.32 = 0$$

$$Q = 0.21 - 0.52 + 0.31 = 0.$$

In fixing the bandwidths of the *I*-and *Q*-signals, colour sensitivity of the eye is kept in view. The eye cannot distinguish between different colours of an object when the size is very small. Colour details are seen by the eye only in the case of medium sized or big objects. Details of small objects is best rendered in black-and-white. The maximum bandwidth used in colour signals is 1.5 MHz. The *I*-signal contains frequencies from 0 to 1.5 MHz and the *Q*-signal has colour frequencies extending from 0 to 0.5 MHz only. As such, the *I*-and *Q*-signals are passed through low-pass filters to limit the sidebands to the ranges mentioned above. This also helps in avoiding interference between the *I*-and *Q*-signal and the lower sideband frequencies of the luminance signal. Figure 16.7 gives a block diagram showing the actual formation of the *I*-and *Q*-signals and the method of encoding these signals at the transmitter.

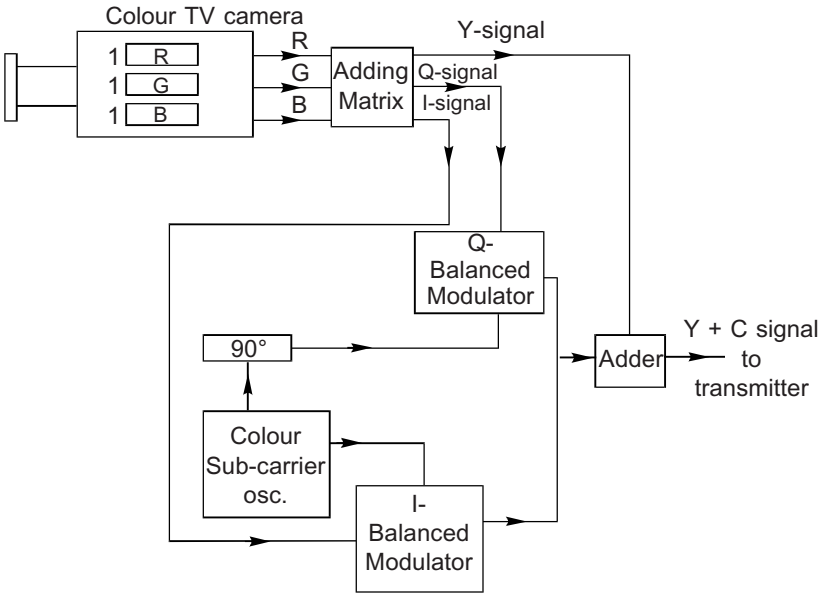


Fig. 16.7 Encoding *I* and *Q* Signals in NTSC System

16.10 COLOUR BURST SIGNAL

The balanced modulators in Fig. 16. 7 are meant to suppress the sub-carrier and produce only sidebands of the *I*-and *Q*-signals. This is done to reduce interference, in the reproduced picture, due to the high level of the sub-carrier. But the presence of the sub-carrier at the receiver is necessary for the demodulation of the sidebands of *I* and *Q*. The sub-carrier is, therefore, generated at the receiver by a local oscillator. It is essential that the phase of the sub-carrier generated at the receiver should be exactly the same as the phase of the sub-carrier suppressed at the transmitter. To ensure this, a wave train of 8 to 11 cycles of the colour sub-carrier is transmitted along with the colour signal. This is called the *colour burst*. It is located on the back porch of the blanking signal as shown in Fig. 16.8.

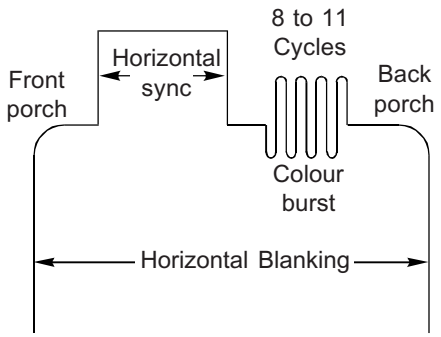


Fig. 16.8 Colour Burst on the Back Porch of Horizontal Sync Pulse

The front portion of the blanking pulse immediately before the sync pulse is called the *front porch* and the portion immediately following the sync pulse is called the *back porch*. The back porch is three times longer than the front porch.

The colour burst signal is the sync signal which controls the tint or hue of the reporduced colour picture. It is an additional sync signal for colour television besides the normal horizontal and vertical sync signals. If the colour burst is absent in the transmitted signal, the reproduction of the picture in the receiver is a monochrome one.

16.11 COMPOSITE COLOURPLEXED VIDEO SIGNAL

Methods for obtaining luminance signal Y and the chrominance signal C from the R, B and G outputs from the colour TV camera have already been explained in earlier sections. These two signals modulate the picture carrier of the standard TV channel. However, the actual waveform of the composite colour signal that modulates the picture carrier also contains a number of other pulses. These pulses are the blanking pulses, the sync pulses and the colour-burst pulses which serve as the sync pulses for the colour signal. All these pulses are necessary to produce a steady picture on the TV screen and to reproduce properly the hues of various colours. Figure 16.9 shows the waveform of the actual composite video signal transmitted by a colour TV system.

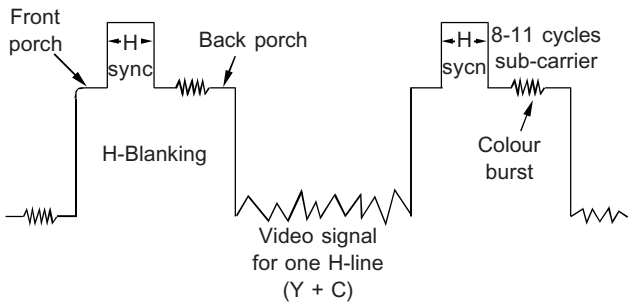


Fig. 16.9 Composite Video Signal Waveform for Colour TV

16.12 COLOUR AND THE PHASE ANGLE

It has been explained earlier that in Quadrature Amplitude Modulator the two colour difference signals $(R - Y)$ and $(B - Y)$ modulate their individual sub-carriers (4.43 MHz) which are at right angles to each other. The resultant is the C-signal or the chrominance signal whose amplitude and phase depend on the relative amplitude and phases of the two colour difference signals as shown in Fig. 16.10.

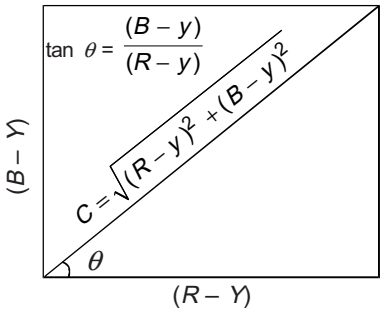


Fig. 16.10 Formation of C-signal

In this figure

$$C = \sqrt{(R-Y)^2 + (B-Y)^2}$$

and the phase angle θ is given by

$$\tan\ \theta = \frac{(B-Y)}{(R-Y)}$$

The amplitude or length of the resultants determine the saturation of the colour signal and the phase angle θ is governed by the hue or colouring of the picture being televised.

This brings out an important fact that for correct reproduction of the colours at the receiver, the phase angle θ of the resultant with respect to $(B - Y)$ and $(R - Y)$ should not change during transmission and for this purpose a special phase control is provided in the receiver to compensate for any phase shift that might reoccur.

To understand how the relative values of $(R - Y)$ and $(B - Y)$ determine the value of the resultant vector C and the phase angle θ we consider the following:-

We know that,

$$Y = 0.59G + 0.30R + 0.11B \tag{16.1}$$

$$\therefore \quad R - Y = R - 0.59G - 0.30R - 0.11B \tag{16.2}$$

$$\text{or } R - Y = 0.70R - 0.59G - 0.11B \tag{16.3}$$

and

$$B - Y = B - 0.59G - 0.30R - 0.11B \tag{16.4}$$

or

$$B - Y = 0.89B - 0.59G - 0.30R \tag{16.5}$$

This means that the $(R - Y)$ and $(B - Y)$ vectors contain R, B and G voltage in the proportion indicated.

Let us suppose that the colour camera is scanning a scene which contains only red, i.e. no green or blue voltages are produced and from equation (16.3) $R - Y = 0.70R$ $(B = 0,\ G = 0)$ and $(B - Y)$ from equation (16.4) is $B - Y = -0.30R$. This position is represented by the following vector diagram (Fig. 16.11).

By the same process we can draw a diagram showing the portion of the resultant when only blue or green is being scanned or any other colour formed

by containing the three primary colours. A phasor diagram given in Fig. 16.11 shows how the phase of the sub-carrier varies as the colour to be transmitted.

Another fact that is clearly brought about by the colour phasor diag. of Fig. 16.12 is that both $(R - Y)$ and $(B - Y)$ can be either positive or negative depending on the hue they represent. The reason being that for any

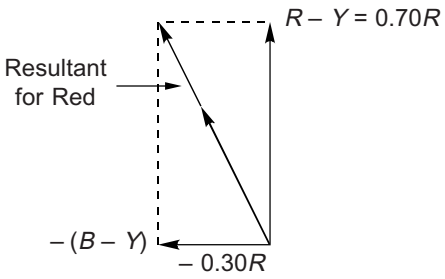


Fig. 16.11 When Only Red is being Scanned

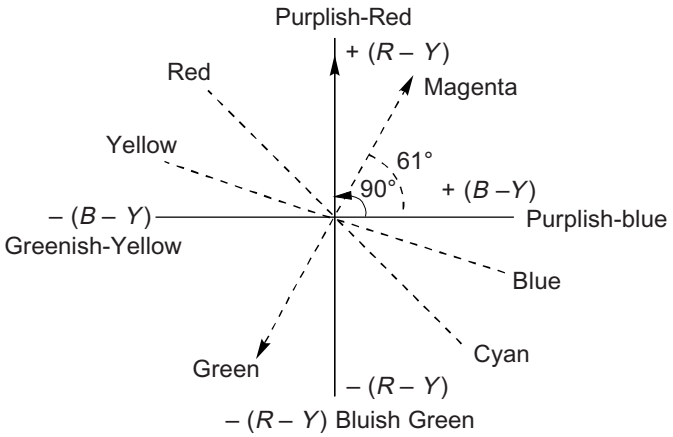


Fig. 16.12 Phasor Diagram for Colour Sub-carrier

primary, its complement contains the other two primaries. Thus a primary and its complement can be considered as opposite to each other and hence the colour difference signals turn out to be of opposite polarities.

To calculate the actual value of the resultant C and the phase angle θ as shown in Fig. 16.11 we proceed as below

$$C = \sqrt{(R-Y)^2 + (B-Y)^2}$$

$$R - Y = 0.7R \quad \text{and} \quad B - Y = -0.3R$$

$$C = \sqrt{(0.7R)^2 + (-0.3R)^2}$$

$$= \sqrt{0.49R^2 + 0.09R^2} = \sqrt{0.58R^2} = 0.76R$$

$$\tan\theta = \frac{0.7}{0.3} = 2.3$$

$$\theta = 67°$$

θ is the angle made by the resultant C with $(R - Y)$. This angle θ determines the hue and must not change for correct reproduction of the colour at the receiver.

To summarise, the phase angle of the resultant is governed by the colouring of the picture whereas the amplitude or length of the vector determines the saturation of the colours.

16.13 TRANSMISSION OF COLOUR DIFFERENCE SIGNALS

As stated earlier, out of the three possible colour difference signals only two viz $B - Y$ and $R - Y$ are used for transmitting colour information to the receiver and the third colour difference signal $G - Y$ is not considered suitable for transmission because G information can be obtained from $B - Y$ and $R - Y$ at the receiver as explained below:

$$Y = 0.30R + 0.59G + 0.11B \tag{16.6}$$

also
$$Y = 0.30Y + 0.59\ Y + 0.11Y \tag{16.7}$$

$\therefore$
$$0 = 0.30\ (R - Y) + 0.59\ (G - Y) + 0.11\ (B - Y) \tag{16.8}$$

$$0.59\ (G - Y) = -\ 0.30\ (R - Y) - 0.11\ (B - Y) \tag{16.9}$$

or
$$(G - Y) = -\frac{0.30}{0.59}(R - Y) - \frac{0.11}{0.59}\ (B - Y) \tag{16.10}$$

$$= -\ 0.51(R - Y) - 0.188(B - Y) \tag{16.11}$$

This shows that amplitude of both $(R - Y)$ and $(B - Y)$ is less than unity and this can be derived by use of simple resistor attenuation put across the paths of the respective signals. On the other hand, we find that

(i)
$$(R - Y) = -\frac{0.59}{0.30}\ (G - Y) - \frac{0.11}{0.30}\ (B - Y) \tag{16.12}$$

$$= -\ 1.97\ (G - Y) - 0.37\ (B - Y) \tag{16.13}$$

The factor 1.97 implies a gain in the matrix and would need an extra amplifier in the circuit, leaving $(R - Y)$ is not easy

(ii) Similarly
$$(B - Y) = -\frac{0.59}{0.1}\ (G - Y) - 0.3\ (R - Y)$$

$$= -\ 5.4\ (G - Y) - 2.7\ (R - Y)$$

Both the factors 5.4 and 2.7 are greater than unity which means two extra amplifiers will be needed in the matrices. This is even more complicated than missing out $(R - Y)$. Moreover, proportion of G is relatively large (0.59) in Y and thus $G - Y$ is small. This being the smallest of the three colour difference signals will need amplification and create S/N problems.

Thus transmitting $(G - Y)$ is both technically and economically less convenient than transmitting the other two colour difference signals.

16.14 COMPATIBILITY CONDITIONS PRODUCED BY COLOUR DIFFERENCE SIGNALS

As already stated Y signal is formed by adding the colour camera outputs in the following ratio.

$$Y = 0.3R + 0.59G + 0.11B$$

These percentages correspond to relative brightnesses of the three primary colours.

The colour difference signals equal zero when white or grey shades are being transmitted as shown below

(a) On peak white let $R = G = B = 1$ volt

Then $Y = 0.3 + 59 + 0.11 = 1$ volt

$R - Y = 1 - 1 = 0$ volt and $B - Y = 1 - 1 = 0$ volt

Both $R - Y$ and $B - Y$ vanish.

(b) When grey shades (not white) are being transmitted let $R = G = B =$ v volts where v < 1

Then $Y - 0.59$ v + 0.3 v + 0.11 v = v

$\therefore$ $R - Y =$ v – v = 0 volt and $B - Y =$ v – v = 0 volt

It is thus clear that colour difference signals disappear completely during white or grey content of a colour scene or during transmission of a monochrome scene and thereby producing complete compatibility in the colour TV system.

Summary

Nature abounds in colour and only a colour TV can reproduce natural colours on the TV screen. Except for a few additional circuits, a colour TV set is the same as a black and white TV screens.

Requirements of compatibility make the colour TV a little more complicated than a monochrome TV.

A colour TV camera processes video signals corresponding to red, blue and green which are called primary colours because these can be combined to form any other colour. Thus white is a mixture of 30% red, 59% green and 11% blue.

These video signals are combined to form luminance signal if mixed in the proportion mentioned above. Chrominance signal which contains all the colour information is formed by quadrature Amplifier Modulation of colour difference signals (R-Y) and (B-Y). The two colour difference signals are transmitted by modulating a colour sub-carrier (4.43 MHz). At the receiver the R, B, G signals are recovered to modulate their respective beams in the colour picture tube. This process is used by the three main colour TV systems viz. NTSC, PAL and SECAM.

Colour burst signal is the sync signal that controls the hue or colour of the reproduced picture. The value of c and the phase angle between ($R - Y$) and ($B - Y$) determine the actual colour at the receiver. This is done with the help of phase control or hue control.

Review Questions

1. What are the advantages of Colour TV over Black and White TV?

2. What is COMPATIBILITY? State the conditions necessary for compatibility between colour TV and B/W TV.
3. (a) Name the three primary colours and state their properties.
   (b) What are the complementary colours for green red and blue?
4. Define the terms (a) luminance (b) chrominance and (c) colour difference signals.
5. Describe a colour TV camera. Give a block diagram for the production of (*R*-*Y*) and (*B*-*Y*) signals.
6. What is qudrature amplitude modulation and what are its advantages and explain its application in colour TV?
7. Write short notes on
   (a) colour burst (b) colour killer (c) composite colourplexed video signal.
8. Explain why colour difference signal (*G*-*Y*) is not normally transmitted for production of the colour picture at the receiver.
9. Explain how colour difference signals help in the production of compatibility conditions between colour TV and B/W TV.
10. State whether TRUE or FALSE.
    (a) A primary colour cannot be produced by combining the other two primary colours.
    (b) Cyan is complimentary to green.
    (c) Colourburst frequency for the NTSC system is 4.43 MHz.
    (d) (*G*-*Y*)is not suitable for transmission of colour signals.
    (e) Luminance is the actual colour seen by the eye.

chapter 17

TV Cameras and Picture Tubes

17.1 TV CAMERA

TV telecasting process actually begins at the TV Camera which is a device for conversion of the optical image of the scene to be televised into corresponding electrical signals to be transmitted to the receiver by modulation of radio waves.

A TV Camera chain consists of an optical system, a camera pickup tube and video processing circuits as shown in Fig. 17.1.

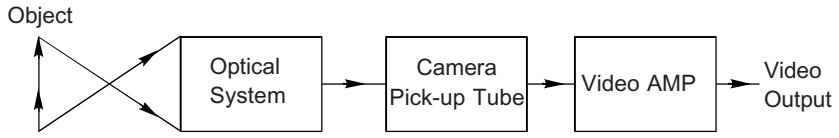


Fig. 17.1 TV Camera Chain

The optical system is very much similar to the optical system of photographic camera or a movie camera, so only properties of the camera tubes will be described in this chapter, the video circuits will be taken up in the receiver section.

17.2 CAMERA TUBE

A camera tube has been very appropriately described as the eye of the TV system. As such, it must possess characteristics similar to the human eye. These characteristics are:

(i) Sensitivity to the visible spectrum (ii) A wide dynamic range for light intensity and (iii) good resolution capability. Several types of camera tubes have been devised to meet the various requirements of operating conditions, physical size, and cost etc.

17.2.1 Basic Principle

The basic principle employed in the construction of camera tubes is the photo-electric effect which allows the conversion of light energy into electrical energy with the help of certain photosensitive materials like potassium, seetenium, lead and their oxides. Two types of photoelectric effects used in camera tubes are (i) Photoemission (ii) Photo-conduction. Besides the above two photo-electric effects, the solid-state image scanners based on the properties of charge coupled devices (CCD) are also coming into vogue.

TV tubes based on these principles will now be described.

17.2.2 Photoemission

Certain metals emit electrons when light falls on them. The emitted electrons are called photo-electrons and the surface emitting the electrons is called a *photo-cathode*. The photons (bundles of light energy) give energy to the outer valence electrons to allow them to overcome potential energy barrier at the surface. The number of electrons emitted depends on the light intensity. Alkali metals, which have a low work function, are normally used as photo-cathode materials. The metals commonly used as photo-cathodes are, Cesium-silver, bismuth-silver-cesium. These are preferred as photoemissive surfaces because they are sensitive to incandescent light and have a spectral response very much similar to the human eye.

The schematic diagram in Fig. 17.2 shows how a video signal is formed by photoemission:

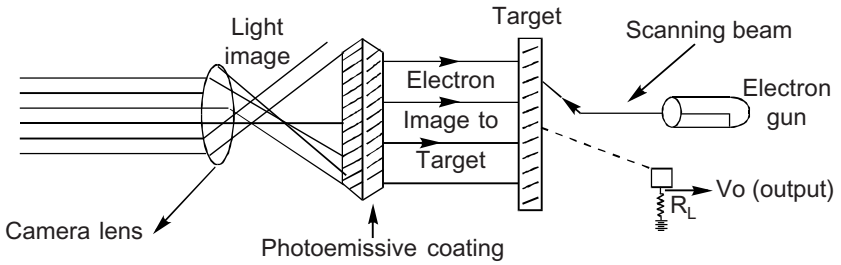


Fig. 17.2 Video Signal Formation by Photo-camera

In this process an optical image of the object to be televised is focused on a mica screen, on the front side of which are deposited millions of globules of silver with a surface coating of cesium oxide which is a photosensitive material. The mica plate coated with myriads of tiny photosensitive silver cesium oxide globules is called *mosaic*. The other side of the mica sheet is coated with a conducting film of graphite called the signal plate. The tiny globules, which are insulated from the signal plate by mica form separate capacitors having mica dielectric and the conducting signal plate in common. The photosensitive globules emit electrons when light falls on the mosaic and each capacitor gets positively charged, the amount of charge depending on the intensity of light

falling on the globule. The mosaic surface thus gets a charge distribution in accordance with the variations of light intensity in the original optical picture. Thus an electric image of the object is formed on the target plate by storage of electric charge on the globules.

The electric image is scanned in accordance with the standard interlaced scanning pattern by a narrow electron beam formed by the electron gun and deflected horizontally and vertically by the deflection coils. As the electron beam scans the positively charged surface elements, it restores to these elements the electrons lost by emission. The capacitors associated with the picture elements get discharged, the amount of discharge current being proportional to the intensity of the light with which the particular picture element has been affected. The capacitor discharge current passes through the load resistor R_L and a voltage called the video signal is developed across the load resistor. This video signal is then amplified by the video amplifier.

The original camera tube based on this principle was called iconoscope. The sensitivity of the *iconoscope* camera tube being rather low, it has been largely replaced by the more sensitive image orthicon camera tube described in the following section.

17.2.3 Image Orthicon

The image orthicon shown in Fig. 17.3 consists of three main sections—the image section, scanning and multiplier section.

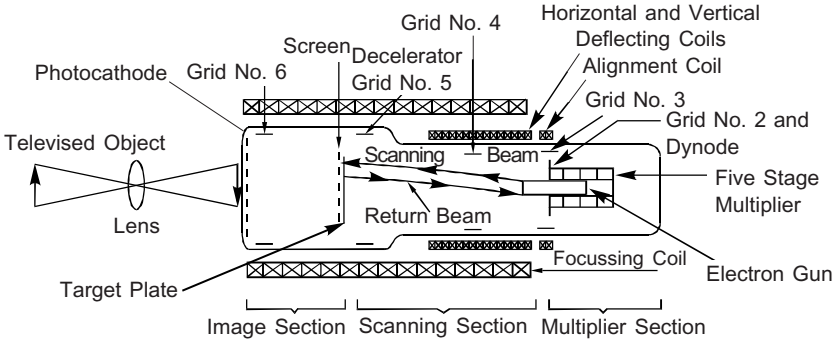


Fig. 17.3 Diagram of an Image Orthicon

In the image section an optical image of the televised scene is formed on the left-hand side of the photocathode, which is a photosensitive film so thin that electrons are emitted from the right-hand side in proportion to the intensity of light. These are accelerated towards the target, and passing through the screen mesh they strike the target and produce secondary electrons. These secondary electrons are drawn to the screen, leaving on the target a pattern of positive charges corresponding to the light intensity distribution in the televised scene.

In the scanning section, the back of the target plate is scanned with a low velocity scanning beam produced by the electron gun. The scanning beam

produces the electrons required to neutralise the positive charges on the electrical image. The electrons in excess of the amount needed to neutralise the positive charges turn back from the target and move towards the electron gun. On striking the electron gun aperture which is grid 2, it produces secondary electrons which are deflected into an electron multiplier system. This system consists of a succession of surfaces called dynodes which are at progressively higher positive potentials and are treated to produce secondary electron emissions. The secondary electrons are focussed on successive dynodes resulting in electron multiplication, and hence enough amplification to give a good voltage drop across the load resistor in the multiplier's anode circuit. This is the camera signal output which is fed into the video amplifier circuit.

The image orthicon produces less current corresponding to brighter portions of the image scanned which means that the signal is maximum for black.

17.2.4 Photo Conduction

In tubes employing photo conductive cathodes, the scanning electron beam causes a flow of current through the photo conductive material whose conductivity or resistivity varies in proportion to the amount of light falling on the photo cathode. The materials which are photo conductive include selenium and lead with their oxides. In effect, the photo electric effect in these photoconductive materials is a decrease in resistance with more light.

Figure 17.4 given below indicates the production of video signal by photo conductive effect.

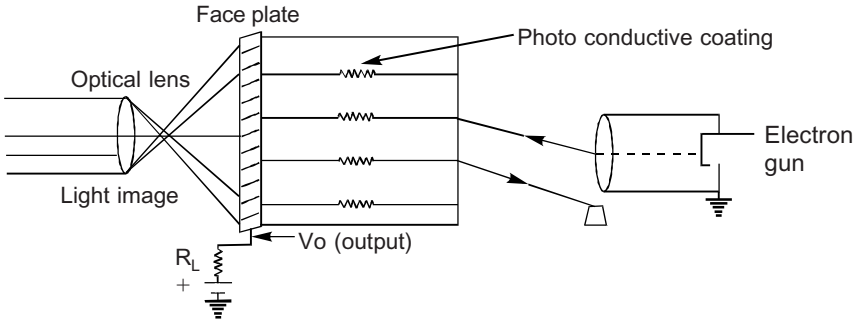


Fig. 17.4 Production of Video Signal by Photo Conduction

17.2.5 Vidicon-Camera Tube

Vidicon camera tube is of relatively simpler construction and makes use of the photoconductivity of certain semiconductor materials such as amorphous selenium, the resistances of which decreases on exposer to light. As shown in Fig. 17.5 the Vidicon tube consists of a target, a fine mesh screen (grid 4), a beam focusing electrode (grid 3) and an electron gun.

The target plate consists of a very thin and transparent film of a conducting material coated directly on the inside of the glass face plate. This is the signal plate for the camera output signal. The gun side of the signal plate is coated

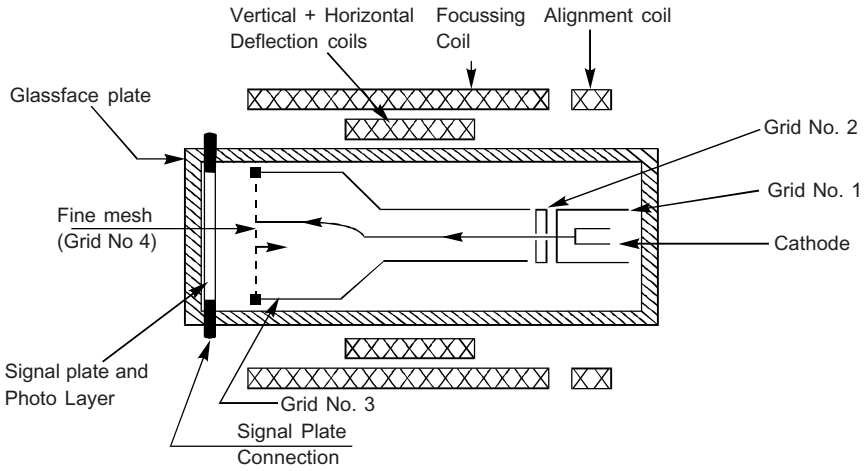


Fig. 17.5 A Vidicon

with an extremely thin layer of a photoconductive material such as selenium or antimony compounds. When illuminated with light from the optical image, the resistance of each element of the photoconductive layer decreases in proportion to the amount of light falling on it. The photoconductive layer is scanned by an electron beam from the electron gun. This electron beam originates from the cathode that is at a potential of about –30 V with respect to the signal plate. The scanning electron beam deposits enough electrons on each spot that it scans to reduce the potential. Excess electrons not deposited on the target are turned back but not used in the vidicon. Grid 4 is a wire mesh which provides a decelerating voltage for the electron beam so that the low velocity beam can deposit electrons on the charge image without producing secondary electrons from the photolayer.

The potential difference across a particular spot on the photoconductive material is 30 V before and after it has been scanned. This change in potential causes a capacitive signal current to flow in the signal plate circuit producing a voltage drop across the load resistance R_L. This is the video signal voltage.

The vidicon has the same sensitivity and resolution as the image orthicon but it is not able to reproduce rapid motion quite satisfactorily as the resistance of the photoconductive material does not vary instantaneously with changes in light intensity.

17.3 LAG

This refers to the time lag of the photo conductive layer in the vidicon which can cause smear with a tail or a comet following the fast moving objects in the televised scene. This photoconductive lag increases at high target voltages where the vidicon has its highest sensitivity. Because of this inherent lag, the use of earlier vidicon tubes was limited to only stationary objects like, slides, pictures or closed circuit TV. However, the present day improved versions of

vidicon having low lag are quite useful in applications in education, medicine, aerospace, oceanography etc. These special types are lead oxide vidicon, Silicon-diode vidicon named after the special types of target plates used in the respective types. However, most of the problems have been overcome with the development of PLUMBICON camera tube.

17.4 PLUMBICON

This is a small camera tube like the vidicon. It has overcome most of the drawbacks of the standard vidicon tube. It has a fast response and produces high quality pictures at low light levels. Because of its small size, light weight and low power operating characteristics, plumbicon is the most suitable camera tube for transistorised television cameras.

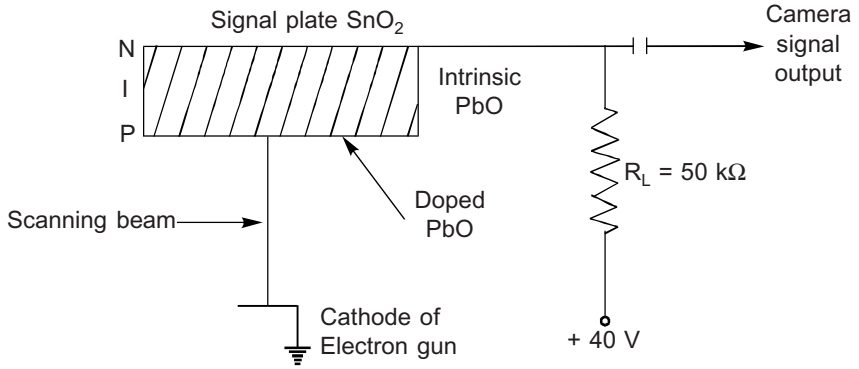


Fig. 17.6 Plumbicon Camera Tube

Except for the target, plumbicon is very much similar to standard vidicon. The target operates effectively as a PIN semiconductor diode coating on the inner side of the glass face plate. P-type semiconducter is doped to have an excess of positive charges, N-type has excess of electrons as negative charges. Intrinsic semiconductor, of I-type is pure Pbo to be neutral without doping. In the manufacture, an intrinsic (I) layer of pure PbO is sandwitched between the N-type SnO_2 signal plate to form a PIN-type semiconductor diode.

The functioning of the photoconductive target in plumbicon is similar to the photo conductive material in vidicon except that in standard vidicon, each element acts as a leaky capacitor, with leakage resistance decreasing with increasing light intensity. In the plumbicon, however, each element serves as a capacitor in series with a reverse biased light controlled diode. The incidence of light on the target results in photoexcitation of semiconductor junction between the pure PbO and doped layer. The resultant decrease in resistance causes signal current flow which is proportional to the intensity of light on each photo element. The average thickness of the target is 10 to 20 μm. The higher sensitivity of plumbicon compared to vidicon is due to much reduced recombination of photogenerated electrons and holes in the intrinsic layer which contains very few discontinuties.

The spectral response of the plumbicon is closer to that of the human eye except in the red colour region. This makes it one of the most useful camera tubes for TV and other video cameras.

17.5 SOLID STATE IMAGE SCANNERS

Solid state image scanners are based on the functioning of charge coupled devices (CCD) which is a new concept in the Metal-Oxide-Semiconductor (MOS) technology. This solid state image sensor consists of a flat silicon chip with an array of electrons. Unlike the camera tubes described earlier, it does not need an electron gun, scanning beam, high voltage or vacuum envelope. The entire image sensor assembly is contained in one chip of solid state semiconductor.

The constructional details of a solid state image scanner are shown in Fig. 17.7.

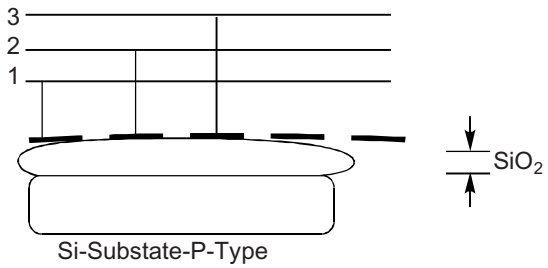


Fig. 17.7 MOS Charge Coupled Device

The chip consists of a p-type substrate, one side of which is oxidized to form a film of silicon dioxide, which is an insulator. Thus by a photolithographic process an array of metal electrodes, known as gates, is deposited on the insulator film. This results in the creation of a large numbers of tiny MOS capacitors on the entire surface of this chip.

The metal electrodes operates in groups of three with every third electrode connected to a common conductor. The spot under the center of each triplet serves as one light sensitive element or a resolution cell.

Operation When the optical image is focussed on to the chip, the light causes electrons to be produced within the silicon, the number of electrons generated depending on the intensity of light. In each triad, the center electrode is the most positive. The charges generated by that element collect at the surface of the silicon under the center electrode. As a result the pattern of collected charges represent the image. The charge at one element is transferred along the surface of the silicon chip by applying a more positive voltage to the adjacent electrode while reducing the voltage on the electrode over the charge packet. The charge is applied by voltage pulses in successive order to all the elements. This method of dissecting the image is equivalent to scanning. When the charge packets reach the output electrode they are collected to form the signal current. The potential required to move the charges is only 5 to 10 volt. This type of image sensor is called a charge coupled device (CCD).

Solid state image scanners have a bright future in TV. Full TV line scanners have already been constructed and chips capable of producing standard interlaced 525 line (or 625 line) pictures have already been produced by RCA. Although the quality of pictures produced by CCD scanner is not quite suitable for TV studio use, there are many other applications where these scanners can prove quite useful.

17.6 COLOUR TV CAMERAS

A colour TV Camera actually consists of three Camera tubes, one each for red, blue and green colours. Such a colour TV Camera has already been briefly described in Chapter 16, Section 16.5. Each camera tube develops a signal voltage proportional to the respective colour intensity received by it.

A Colour TV camera normally requires three individual camera tubes for separating the three primary colours. However, in some colour TV cameras a fourth camera tube is used to provide an improved monochrome picture and better detail for the colour picture.

Except for the number of camera tubes used, the distinguishing feature of a colour TV camera is its optical system. It is here that all light entering the camera lens is divided or 'split' into separate colour beams, one beam for each camera tube. This arrangement of mirrors, prisms and lenses is called a beam splitter. Figure 17.8 shows the arrangement for beam splitting in a 3-way beam split.

The white light from the televised scene enters the zoom lens which is an objective lens whose focal length can be varied either manually or with a motor to provide a choice of field of view without changing the lens.

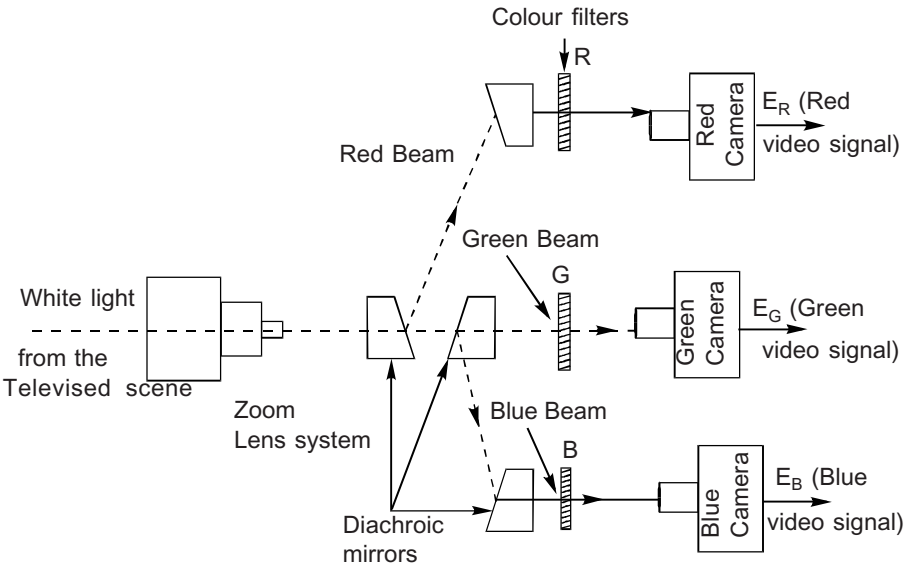


Fig. 17.8 Optical System of a 3-way Beamsplitter

The image formed by the lens system is split into three images by means of glass prisms designed as diachroic mirrors. A dichroic mirror is a device that reflects one particular colour of light and allows the light of all other colours to pass through. Thus red, blue and green voltages are developed into the respective camera tubes. Suitable matrices are then used to combine these R, B and G signals to form the 'Y' signal as already discussed.

A 4-way beam splitter is used in a 4-tube camera. One part of the spilt beam is focussed on to the face plate of the fourth camera tube, in the luminance channel to provide the Y-voltage for the monochrome part of the video signal. The other part of the beam is focussed on to a system of dichroic mirrors to produce the R, B and G signals as described above.

17.7 TV PICTURE TUBES

A TV picture tube is the end product in a TV receiver and provides the necessary screen for viewing the picture received by the TV receiver.

It is a specialized form of cathode ray tube which consists of evacuated glass envelope or bulb inside which is rigidly built an electrode structure consisting of an electron gun and other electrodes (grids) for producing and focussing a thin electron beam on to a fluorescent screen which glows when struck by the electron beam. The electron beam can be deflected by coils which are mounted on the neck of the picture tube. This produces the raster which gets converted into a picture when the video signal is applied to the picture tube.

In a solid state TV receiver, that most modern TV receivers are, picture tube is the only component that is based on vacuum tube technology and occupies bulk of the space in a TV receiver and is also perhaps, responsible for large part of electrical power consumption. Solid state display devices are now available which will soon replace the vacuum tube picture tubes and the modern TV receiver will become solid-state TV receiver in the real sense of the word.

There are two types of picture tubes, monochrome picture tube and a colour picture tube.

A monochrome 'picture tube has only one electron gun and produces a single electron beam that strikes a continuous phosphor coating to produce a black and white picture. A colour picture tube on the other hand, produces three different electron beams which strike a screen coated with three different types of phosphors which emit red, green and blue light when struck by their respective beams. The three primary colours Red, Blue and Green, so produced combine to produce a colour picture on the screen. This needs an elaborate process of deflection and alignment of the different beams which will be described in detail under the descripton for colour picture tubes.

17.8 MONOCHROME PICTURE TUBE

The picture tube converts the video signals from the video amplifier into a picture of the televised scene with the help of other auxiliary circuits in the sweep section of the TV receiver. The constructional details of a modern monochrome picture tube are shown in Fig. 17.9. The electron gun constituted

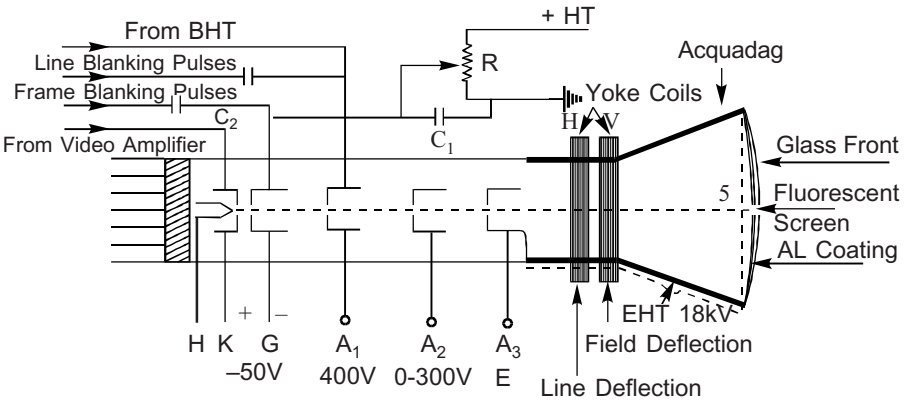


Fig. 17.9 A Monocrome Picture Tube

by the heater H, cathode K, control grid G and the focussing anodes A_1 and A_2, produces an electron beam which, when accelerated by the high voltage on the plate A_3, strikes the fluorescent screen to produce light at the point of impact. The control grid is at a negative potential of about 50 V with respect to the cathode whereas the succeeding electrodes A_1 and A_2 are at a positive potential. The accelerating voltage of about 18 kV is applied to the anode A_3 through a conducting coating called the *aquadag* on the inner walls of the tube and which is connected internally to A_3. This high voltage called EHT (Extra High Tension) is applied to a connection outside on the wall of the picture tube. A fine aluminium coating inside the fluorescent screen reflects light to illuminate the screen. Two pairs of deflection coils called *yoke coils* mounted externally round the neck of the picture tube produce the horizontal and vertical deflection of the beam to produce a pattern of 625 illuminated horizontal lines called the *raster*. This raster gets converted into a picture when the video signal from the video amplifier is applied to the cathode of the picture tube. The raster exists even when there is no video signal applied to the picture tube. The negative bias on the grid can be varied by a potentiometer R to control the brightness of the screen and this is known as the *brightness control*. The frame blanking pulses and the line blanking pulses are also given to the grid G and anode A_1 respectively to suppress the electron beam during flyback periods.

The actual circuit of a typical monochrome picture tube stage in a TV receiver is given in Fig. 17.10.

The video signals form the video amplifier are given to the cathode of the picture tube through 0.1 μF capacitor which has a diode OA79 across it to provide DC restoration to maintain average illumination on the screen. This is called cathode modulation. The grid voltage is maintained at about 25 to 50 V negative with respect to the cathode for proper brightness on the screen. This is done by adjusting the HT voltage on G through the potentionmeter R_2 which is the brightness control. Frame blanking pulses are also applied to the grid via a 0.05 μF capacitor. The voltages for the A_1 and A_2 electrodes are produced by

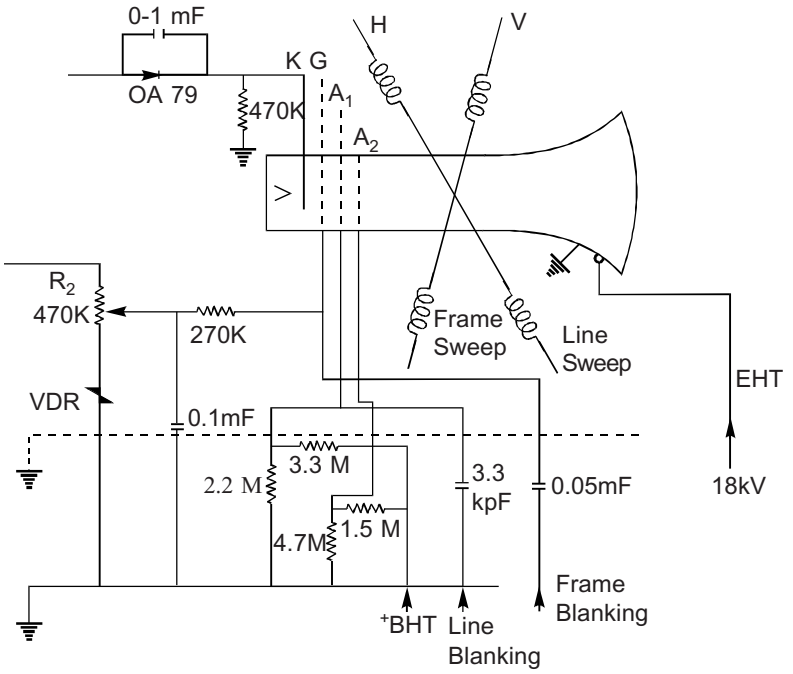


Fig. 17.10 Circuit of a Picture Monochrome Picture tube Stage Monochrome Tube Stage

the Boosted HT (BHT) and line blanking pulses are applied to A_1 through a 3.3 kpF capacitor. EHT of about 18 kV is applied to a connection provided for the purposes on the body of the tube. Sweep voltages for the deflection coils are obtained from the respective sweep output stages.

The two main circuits which distinguish a colour TV from a monochrome TV are the colour picture tube and the chroma section containing the colour circuits. These two sections will now be described in detail.

17.9 COLOUR PICTURE TUBES

The picture tube in a colour TV receiver is of special construction. It differs in many ways from the conventional picture tube used in a monochrome TV receiver. Various types of picture tubes have been developed for the colour receiver but the one most commonly used is shown in Fig. 17.11. It is called a Delta-gun picture tube. It has three separate electron guns to produce three electron beams. These electron beams are intensity modulated by the red, blue and green voltage signals produced by the demodulation circuits of a colour TV receiver. Colours are produced when the intensity modulated electron beams strike their respective targets on a phosphor coated screen coated with phosphors corresponding to red, blue and green primary colours. The three phosphors are arranged on the screen in triangular groups called *triads*.

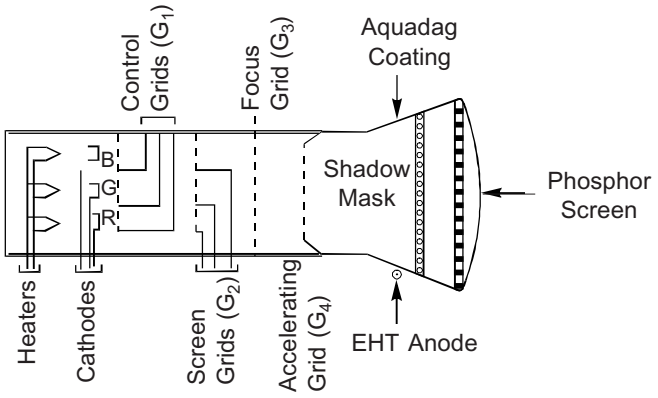


Fig. 17.11 Colour Picture Tube Showing Details of Electron Gun Assembly and Other Elect.

Each triad contains a red-emitting dot, a blue-emitting dot and a green-emitting dot as shown in Fig. 17.12. The alignment of the electron beams is such that each electron beam always strikes its own phosphor dot to excite that particular colour. When deflected, the three electron beams produce red, green and blue rasters on the screen. Viewed from a normal distance, the three primary colours are integrated by the human eye to reproduce as closely as possible all the natural colours of the transmitted scene.

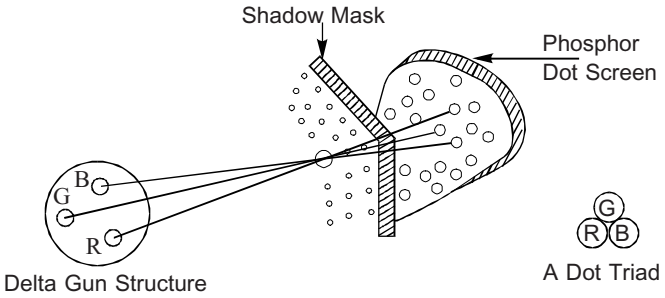


Fig. 17.12 Delta Gun Picture Tube Showing Shadow Mask, Phosphor Screen and a Dot

Shadow Mask To ensure that a particular electron beam strikes only its associated phosphor dot in the triad, the three electron beams are made to converge at a point a small distance behind the phosphor screen. At this point is placed a thin perforated metal sheet called a *shadow mask* as shown in Fig. 17.12.

The shadow mask contains several hundred thousand very fine holes so that one hole is available for every dot triad on the phosphor screen. When properly aligned, the red beam can illuminate only the red phosphor dot while the blue and green dots remain hidden under the shadow of the shadow mask. Similarly, the electrons from the green gun can illuminate the green phosphor dots while the blue electron beam will strike only those dots that emit blue light.

A shadow mask picture tube is capable of producing good quality colour pictures but the shadow mask blocks a considerable amount of electron beam current. The phosphor screen gets only about 20% of the total energy of the electron beam. As a result, the colour picture produced on the screen will not be sharp unless higher accelerating voltages are used. This is the reason why colour picture tubes use 25 kV for their accelerating anodes compared to only 18 kV used in a monochrome picture tube.

17.9.1 Components Mounted on the Neck of the Colour Picture Tube

Besides the internal structure of the tri-gun picture tube described earlier, a number of components are also mounted externally on the neck of the picture tube. These components are the deflection-coil assembly, the convergence-yoke assembly, the purity magnets and the blue lateral-convergence assembly. All these components are shown in Fig. 17.13. These components play a very important part in the formation of a colour picture on the screen. The final quality of the colour picture depends to a great extent on the proper adjustment of these components. The functions of these components will now be briefly described.

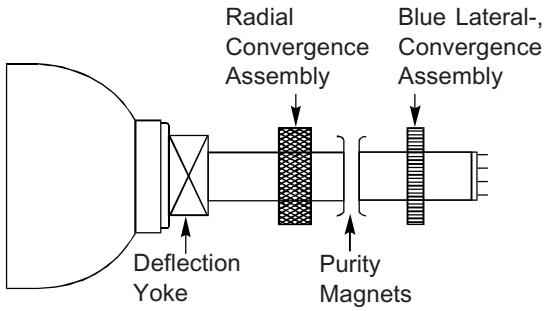


Fig. 17.13 Components Mounted on the Neck of the Colour Picture Tube

Deflection Yoke Assembly This contains the horizontal and vertical deflection coils for all three electron beams. The deflection coils help in producing a raster for each colour on the phosphor screen. The function of the deflection assembly is the same as in a monochrome TV.

Convergence Yoke Assembly This assembly enables a positional adjustment of the individual beams to be made so that they converge through the same opening in the shadow mask. The three assemblies for the red, blue and green beams have their individual coils and permanent magnets. The coils and the magnets are placed symmetrically in a yoke assembly, around the neck of the picture tube as shown in Fig. 17.14(a). The internal structure includes the pole pieces and the magnetic shield which allow each beam to be shifted in position without affecting the other beams.

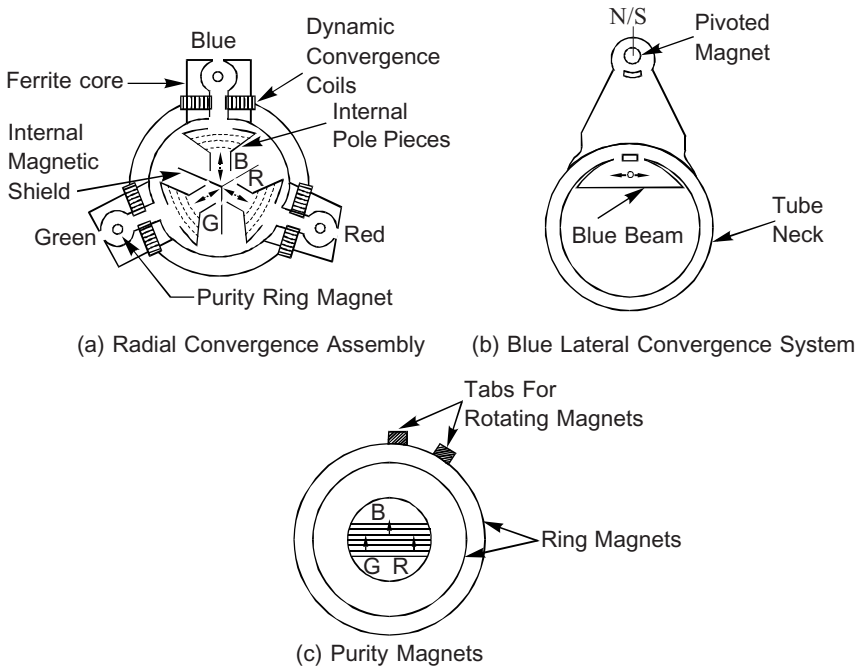


Fig. 17.14 Convergence Assembly and Associated Parts (a) Radial Convergence Assembly (b) Blue Lateral Convergence System (c) Purity Magnets

The permanent magnets are meant for *static convergence* in the central area of the screen. The convergence coils on *U*-shaped ferrite cores keep the beams in convergence over the rest of the screen by continuously varying electromagnetic fields. This is called *dynamic convergence*.

Blue Lateral Magnet The radial adjustment of all three beams alone is not enough to ensure their convergence to one point. Lateral adjustment of at least one beam to left or right is necessary for this purpose. To achieve this, a blue lateral magnet is employed which shifts only the blue beam to right or left for proper convergence. The construction of the blue lateral magnet is shown in Fig. 17.14(b).

Purity Magnet Purity means that each of the three electron beams strike only their respective phosphor dots. For establishing purity, each individual beam should form its own raster when the other two beams are off. For example, with the green and blue beams off, the raster should be pure red and so on for the other two colours.

The purity magnet, shown in Fig. 17.14(c), consists of two rings similar to the centring rings. These can be rotated by means of tabs till the beams have the correct angular positions to strike their own phosphor dots.

Proper adjustment for convergence and purity are necessary for a good reproduction of the coloured picture. The test for good convergence is a black-

and-white picture without any signs of colours on the fringes of the objects. Correct purity adjustment is indicated by the production of red, green or blue raster with only one beam operating at a time. Red raster is normally preferred for this adjustment. The final test for purity is the production of a solid white raster with all the three guns operating.

17.10 DEGAUSSING

Certain steel fittings inside the colour picture tube, particularly the shadow mask, get magnetised either due to Earth's magnetic field or due to some other stray magnetic fields in the vicinity of the TV receiver. Such magnetism adversely affects the purity adjustment. Degaussing is the process by which the iron and steel parts in the picture tube are demagnetised.

Degaussing can either be manual or automatic. In manual degaussing, a degaussing coil connected to the AC mains is moved in circular motions in front of the screen and then slowly moved away. The process can be repeated two or three times. In automatic degaussing (ADG), the degaussing coil is fitted around the front rim of the picture tube. When the set is switched on, a strong mains current passes through the degaussing coil. The magnetic field so produced gradually dies away, thereby completing the degaussing process every time the set is turned on. In certain receivers, the automatic degaussing operates when the receiver is switched off.

ADG—A number of ADG circuits have been devised but the simplest is the p-type thermistor ADG circuit shown in Fig. 17.15.

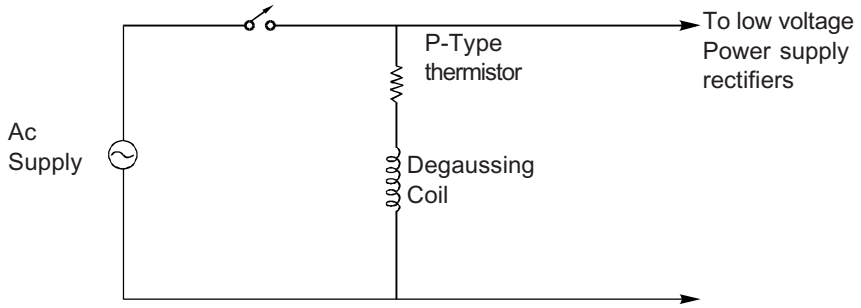


Fig. 17.15 P-type Thermel ADG Circuit

In the circuit, a positive temperature coefficient resistor or *P* type thermister is placed in series with the degaussing coil and the combination is placed across the AC power supply source as shown in Fig. 17.15.

When the P-type resistor ADG circuit is first energized, the resistance is very small (25 Ω) and high current flows though to the ADG circuit. When the (*P*) resistor gets hot its resistance is high (1 MΩ or more) and AC current flowing through ADG coil drops to almost zero, thereby completing the degaussing action.

17.11 CONVERGENCE

This is a technique of bringing all the three beams in colour picture tube together so that they hit the same part of the screen at the same time to produce three coincident rasters. Convergence falls into two categories:–

(a) *Static Convergence* This process involves the deflection of the three beams by permanent magnetic fields so that they converge into the center area of the screen.

(b) *Dynamic Convergence* This type of convergence is achieved on the rest of the screen by continuously varying the dynamic magnetic fields, the instantaneous strengths of which depend on the portion of the spot on the screen. These fields are set up by electromagnets which carry currents at horizontal and vertical frequencies.

17.12 PROBLEMS WITH DELTA-GUN PICTURE TUBE

Delta gun picture tube is one the first and oldest type of colour picture tube used in earlier colour TV sets. However, this picture tube has the following drawbacks.

1. Because of the electric gun structure the convergence adjustments require a complicated and complex service adjustment for obtaining static and dynamic convergence.
2. The focussing of the three beams does not remain sharp over the entire screen because the three electron guns are located in three different planes.
3. The shadow mask intercepts most of the electron beam, thereby necessitating the use of much higher EHT voltages of about 25kV.

The shadow mask or Delta-gun type of picture tube has been superceded by other types of colour picture tubes which are described below.

17.13 OTHER TYPES OF COLOUR PICTURE TUBES

17.13.1 Sony Trinitron Picture Tube

This picture tube developed by Sony Corporation of Japan, uses in line beam structure with vertical phosphor stripes (Fig. 17.16(b)). It employs a single gun with three in-line cathodes as shown in Fig 17.16(a). The mask used has vertical slots through which the beam hits respective stripes of red, blue and green phosphors. Convergence is achieved electrostatically by four convergence plates. This tube has a much greater brightness capability and overcomes all the convergence problems of the conventional delta-gun shadow mask picture tube.

Figure 17.16(c) gives full details of the elements used in a Trinitron colour picture tube.

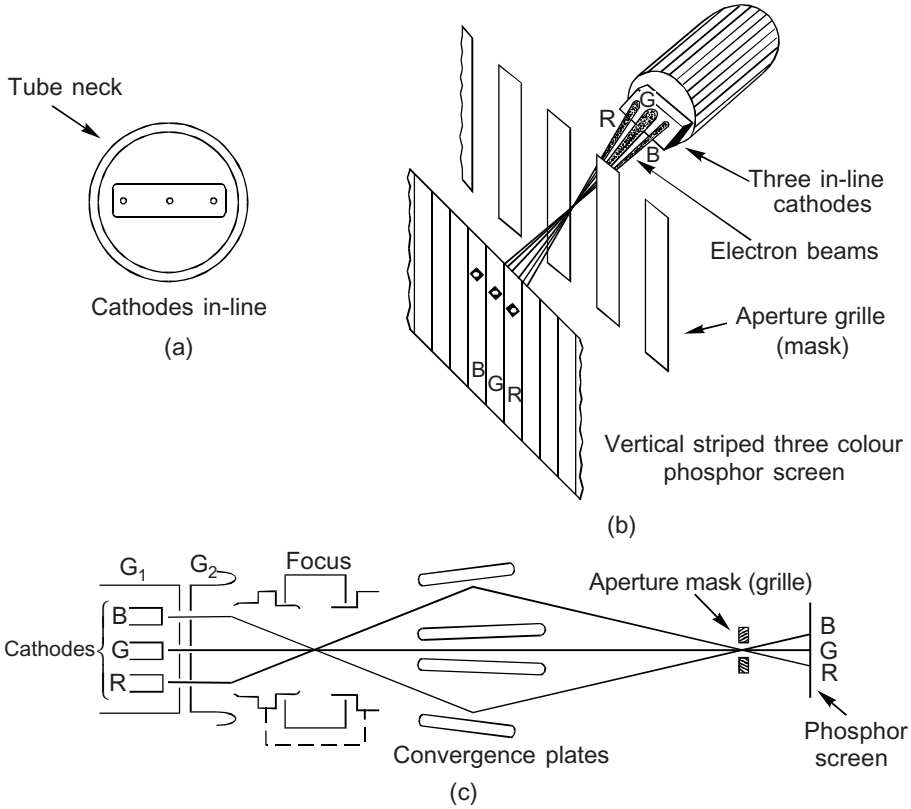


Fig. 17.16 Trinitron (cathodes in-line) colour picture tube (a) gun structure (b) electron beams, vertical-striped three colour phosphor screen (c) constructional, focus and convergence details.

17.13.2 RCA in Line Picture Tube

Another type of in line colour picture tube developed by RCA has an in-line three-gun structure instead of the in-line single structure of the trinitron tube. The phosphor screen has vertical stripes and the mask contains vertical slots. The components mounted on the neck of the tube include the deflection yoke and magnet assembly for purity and static convergence. All these components are factory adjusted cemented on to the neck of the picture tube. This picture tube does not require any adjustment for dynamic convergence.

17.13.3 Flat Picture Tubes

Research work has been going on for sometime to develop a flat pannel display device to replace the existing electron beam picture tubes. The major advantage of such a picture display panel is the reduction in television cabinet size due to the elimination of the picture tube neck.

Flat panel picture display will permit television receivers to be hung on the wall like an oil painting. In addition portable receivers of very small size like that of a wrist watch will become possible.

Various techniques are being used for the development of a flat picture tube. Besides the use of solid state image scanner discussed earlier, the use of various types of LCD (liquid crystal display) devices is being made for both the B/W and colour display flat panels for picture display devices.

17.14 MAINTENANCE AND CARE OF PICTURE TUBES

17.14.1 Troubleshooting

A picture tube is the most delicate and expensive component in both the B/W and colour TV receivers. Replacement of a defective picture tube can mean a lot of expense on the part of the owner and a means of income for the service technician. The troubles in a picture tube should be carefully investigated and steps taken to either repair or replace the picture tube.

Both black and white and colour picture tube troubles fall into the following four categories.

1. Open heaters 2. Low emission or gassy tubes 3. Short and open between tube electrodes 4. Poor colour tube gun tracking.

Most of these troubles can be either checked or repaired by a colour technician with the help of a picture tube tester.

1. Open Heaters In a B/W picture tube, the symptoms produced by an open heater are no raster, no picture but sound O.K. Replacement of the picture tube is the only remedy for this defect. The open heater should, however, be confirmed with a multimeter before the tube is condemned and replaced. Colour picture tube generally has three heaters which are connected in parallel. If one of the heaters is open, the colour associated with that heater will be missing.

2. Low Emission The symptoms are low brightness and weak picture. In some cases low brightness picture is also accompanied by silvery high lights. These symptoms normally appear as the picture tube ages and the cathode emission becomes weak.

Low emission can be remedied either by picture *tube boosters* or by *rejuvenation.*

Picture Tube Boosters

A booster is only a step-up transformer that steps up the heater voltage from 6.3 V to 8.1 V. This is attached to the picture tube base on one end (secondary) and the picture tube socket (primary) on the other. All other pin connection are transferred from the receiver socket to the picture tube base by means of wires that are a part of the booster. A suitable booster must be selected for a particular type of picture tube. The increased heater voltage increases cathode emission and restores picture quality.

Picture tubes in solid-state receivers often obtain their heater voltage from windings on the horizontal output transformer. Such picture tubes are supplied with special boosters that are designed to be placed between the fly back and the picture tube heater, increasing the heater voltage and restoring emission to normal.

Rejuvenation

In this process the picture tube is adjusted so that CRT heater voltage is increased by 50% and high voltage (700 V) is applied to the control grid for an instant. This causes a sudden increase in cathode current that strips away the cathode emission. This is a process that need special equipment (picture tester) and expert technique.

3. Shorts Shorts between heater and cathode and between grid and cathode will result in an increased raster brightness and loss of picture information. In colour tubes short will result in loss of only one colour.

Shorts between screen grid and the first anode or screen and focus electrode or short to the second anode will all result in loss of raster.

The remedy for a short is to apply about 600 V between the shorted electrodes from the tester. If enough high current is passed, it might burn up the flake and remove the short.

4. Opens Open picture tube electrodes usually result in no raster. In some cases it may be possible to weld the open electrodes by application of high voltage (1000 V) between the electrode.

Tracking Colour picture tube electron guns are designed to have similar electron emission control characteristics. There characteristics sometimes change and have to be matched. Some picture tube testers are equipped to measure these tracking characteristics of the three guns. In certain cases rejuvenation of each gun may be used to remove tracking defects and restore tube operation to normal.

17.15 SPECIFICATIONS OF TV PICTURE TUBES

Two types of TV picture tubes are used with TV receivers manufactured in India. These are (i) Monochrome TV picture tubes and (ii) Colour TV picture tubes.

17.15.1 Monochrome TV Picture Tubes

Monochrome TV picture tubes are mainly manufactured and marketed by Bharat Electronics Ltd (BEL).

The commonly used picture tubes are 310 CIP_4, 470 CIP_4, 590 CIP_4 and 610 CIP_4. The 47 cm and 50 cm tubes have the same electrical characteristics and 54 cm and 61 tubes have identical performance data. These picture tubes employ electrostatic focussing and electromagnetic deflection. These tubes make use of P_4 types of aluminized phosphor for screen coating which has only medium short persistence characteristics. Brief specification of three important

picture tube types manufactured by BEL (310 CIP_4, 410 CIP_4 and 610 CIP_4) are given in Table 17.1. Full details are available in BEL manual on TV picture tubes.

Table 17.1 Specifications of Monochrome Picture Tubes

Type Number	*Size Face diagonal*	*Deflection angle*	*Accelerating anode (FHT)*	*Heater voltage /current*	*Typical G_2 voltage*	*Tube length*	*Remarks*
310 CIP_4	310 mm = 12	100°	10 kV	12V Ac/Dc/ 75 ma	300 V	24 cm	
470 CIP_4	470 mm = 19	114°	16kV	6.3v Ac/Dc/ 300 ma	400 V	20.5 cm	
601 CIP_4	601 cm = 24	110°	18 kV	6.3v Ac/Dc/ 300 ma	400 V	36.2 cm	

Base Connections

310 CIP_4	470 CIP_4	610 CIP_4
1 — G_1	1 — H	1 — H
2 — K	2 — G_1	2 — G_1
3 — H	3 — G_2	3 — G_2
4 — H	4 — G_4	4 — G_4
5 — G_2	5 — Nc	5 — NC
6 — G_3	6 — G_3	6 — G_3
7 — G_4	7 — K	7 — K
8 — NC	8 — H	8 — H

Specifications of Colour Picture Tubes Most modern colour TV receivers make use of PIL tubes. The deflection unit of a modern P.I.L colour picture tube is inherently self. Converging and only minor corrections are necessary to compensate for tolerance and assymetries. Table 17.2 below gives the brief specifications of only two colour picture tubes viz A51 – 210 X (ITT) and A—56 – 500 X (Philips)

Table 17.2 Specifications of Colour Picture Tubes

Type. No.	*Size-Face diagonal*	*Deflection angle*	*Accerlating Anode voltage (EHT)*	*Heater voltage current*	*1st Anode G_2 voltage*	*Tube length*
A51–210X (T.T.)	51 cm	(i) diagonal—90° (ii) horizontal—78° (iii) vertical—60°	25 kV	6.3V/ 680 mA	350 V 750 V	42cm
A 56–500X (philips)	56 cm	(i) diagonal— 110° (ii) horizontal—90° (iii) vertical—77°	25 kV	6.3 V/ 730 mA	465 to 705 V	37cm

Note. 1. First anode voltage (VG_2) depends on the spot cut off voltage.
2. External conducting coating and implosion protection hardware must be grounded.

Summary

A TV camera is a device for conversion of optical image signals from the scene to be televised into electrical signals for the formation of the picture on the picture tube. A camera tube is the eye of the TV system and as such it should possess characteristics similar to the human eye.

Basic principle of a camera tube is the photoelectric 'effect' which allows the conversion of light energy into electrical energy with the help of photosensitive materials like potassium, selenium, lead and their oxides. Two types of photoelectric effects are photoemission and photoconduction. Iconoscope and image orthicon are two camera tubes based on photoemission whereas vidicon and plumbicon are based on the principle of photoconduction.

Besides the vaccum type camera tubes mentioned above, solid state image scanners based on the functioning of charge coupled devices (CCD) are also now available. This new concept reduces the size of a TV receiver.

A colour TV camera actually connects three camera tubes one each for Red, Blue and Green colours.

A picture tube provides the necessary screen for the display of picture received by the TV receiver. A monochrome picture tube is similar to CRT based an electromagnetic deflection. A colour picture tube is of special construction and differs from a monochrome picture tube in many ways. Most commonly used colour picture tube is Delta-gun picture tube which has three separate electron guns to produce three electron beams whose intensities are modulated by red, blue and green voltage signals to produce corresponding colours. These beams strike a screen coated with phosphors corresponding to red, green and blue primary colours arranged in triangular groups called triads.

Besides the internal structure of a trigun picture tube, a number of components like deflection coil assembly, the convergence assembly, yoke-assembly, the purity magnets and blue-lateral convergence assembly are mounted on the neck of the picture tube.

To overcome certain draw backs in the Delta-gun picture tube a number of other colour picture tubes have been devised. Some of these are Sony Trinitron picture tube and the RCA in Lime Picture Tube.

Review Questions

1. What is a TV Camera?
2. Explain why TV camera is described as the 'eye' of the TV system.
3. What is photoemission? Describe a TV Camera based on the principle of photoemission.
4. What is the basic principle on which the Vidicon Camera Tube works? What are the drawbacks of this tube and how are these overcome?
5. Describe the construction and working of a Plumbicon Camera Tube. What are its advantages?

6. What are solid state image scanners? What are the advantages of these scanners over the other types of scanners?
7. How does a colour TV camera differ from a monochrome TV camera? Describe the construction and working of a colour TV camera.
8. Describe a Delta gun picture tube. What are its drawbacks and how are they overcome?
9. Define the terms (i) purity (ii) convergence and (iii) degaussing as applied to a colour picture tube.
10. Explain the difference between a TV camera and a TV picture tube.
11. Discuss briefly the important steps and precautions to be observed in the maintenance and care of TV picture tubes.

chapter 18

TV Broadcast Techniques– TV Studios and Control Room

18.1 TV BROADCASTING-GENERAL REQUIREMENTS

Basic principles of TV Broadcasting or Telecasting have been briefly described in Chapter 15 with the help of a block diagram. It has been explained there how a video signal obtained from the TV camera at the TV studio is converted into a composite video signal with the help of a synchronizing, blanking and other control signals to make it suitable for amplitude-modulation of a video carrier to be transmitted by the TV transmitting antenna.

Besides the formation of the composite video signal the sound picked up by the microphone is also converted and amplified by the audio amplifiers and suitably processed to be able to frequency modulate a sound carrier whose frequency is 5.5 MHz higher than the frequency of the picture carrier. The frequency modulated (FM) sound carrier is also radiated by the same transmitting antenna as used for the transmission of the video or picture carrier. Thus at the TV transmitting station, two separate RF carriers one for the transmission of picture signals and the other for sound signals are radiated by a common transmitting antenna.

The equipment and the methods used for the processing of the video and audio signals in the TV studio will now be studied in this chapter.

18.2 TELEVISION STUDIOS

A television studio is an acoustically treated hall suitably lined with acoustic material to provide a control over the reverberation time of the TV studio. The main difference between a TV studio and a sound broadcasting studio is that

the TV studio besides being acoustically treated is also provided with flood lights to illuminate any part of the studio in order to obtain the desired illumination level to suite the scene to be televised. The illumination level is controlled with the help of dimmerstats or Silicon Controlled Rectifiers (SCRs). The following other points regarding TV studios shall be considered.

18.2.1 Lighting

A number of light fittings of various types are provided and suitably distributed all over the TV studio to obtain the desired art form for the reproduced picture. The light fittings consist of incandescent lamps and quartz iodine lamps operated at suitable colour temperatures-the light fittings consist of spot lights of 0.5 kW and 1 kW and broads of 1 kW, 2 kW and 5 kW ratings.

A number of these fittings are suspended from the ceiling so that they can be adjusted unseen. The adjustment of lowering and raising of these light fittings is manual in the case of small studios but winch motor operated controls are used for bigger installations carrying a large number of light fittings of batten suspension. Catwalks (passages close to the ceiling) are also provided for ease of changing the location of unipole suspensions carrying wells of light fittings. As many as 100 to 120 fittings are employed in bigger studios suitably arranged on battens each carrying 4 light fittings. The lighting is controlled by varying the effective current flow by means of dimmerstats or Silicon Controlled Rectifiers (SCRs). The power to all the lines is fed through automatic voltage stabilizers in order to maintain a steady voltage supply. The main distribution boards and switches are located in a separate room close to the studio. The lighting is controlled by switches and faders on the dimmer console in the Programme Control Room (PCR) on the technical presentation panel.

The lighting is so arranged as to prevent shadows and produce desired contrast effects.

18.2.2 TV Cameras

A number of TV cameras are required for commanding different views of the scene to be televised. Bigger studios require three to four cameras which are either the Image Orthicon or Plumbicon type cameras. Two of these cameras are generally mounted on ground operated Trollies and the third one is mounted on a crane for dramatic shots.

Two basic types of cameras used in TV studios are the heavy professional type and the light portable type. The former type requires the use of a tripod and is wheeled about in a TV studio with the camera man seated on the trolley. The latter type allows greater mobility and can either rest on the shoulder for outside work or can be mounted on a tripod in small studios like the news studio. This is a self contained unit and has all the elements to view a scene and generates a video signal which can be recorded on a VCR or passed on to PCR for further use as required.

The trolley mounted camera used on the studio floor is a remote controlled camera. Head unit usually contains only the photosensitive pick-up tube, its associated deflection circuitry, video amplifier and a video monitor. However, bulk of the circuitry is contained in the camera control unit (CCU), which is connected to the camera head with multicore cable.

The remote camera control unit (CCU) contains most of the electrical operating and set up controls. All camera controls are available on a panel in the production control room.

A TV camera looks like a cine camera so far as its external appearance is concerned but it contains a number of electronic circuits and other controls which are different from a cine camera. Besides, the camera tube, which in most cases is a vidicon, the other important parts of a TV camera are the lens, the view finder and the electronic circuits. These are briefly described below:

18.2.3 Lens

The type of lens used determines the sharpness and contrast in the picture. The light gathering property of a lens is represented by its F number. The smaller the F number, the faster is the lens and greater its light gathering property at max. aperture which is controlled by a diaphragm. The angle of view depends on the focal length of the lens. The longer the focal length, the narrower the angle of view. The focal length of the lens must be varied to present different angles of view. This is done with a zoom lens.

18.2.4 Zoom Lens

A zoom lens has a variable focal length and is designed to provide choice of fields of view. Zoom lens is a standard equipment on most video cameras. It gives a choice of different angles of view, without any need for changing lenses. The range of angles of view presented by a zoom lens is determined by its zooming ratio. If the focal length of a lens is expressed in millimeters, a zoom lens with a focal length variation of 12.5 to 50 mm has a zooming ratio of 4 : 1. The zooming ratio also depends on the size of the camera tube. For example, a 17-mm vidicon tube will use 11 to 60 mm range giving a zooming ratio of 6 : 1 and the angle of view can be varied from 8° to 48°.

Zoom lenses are composed of a number of optical elements. The change of focal length can be brought about by moving only the central portion of the elements. The zoom control can either be manual or auto. In the auto control, which is normally used with professional cameras, several speeds are available, which are selected according to the requirements of the shot. *Zooming-in* and *zooming-out* is a technique which a clever camera-man learns by experience.

18.2.5 View-Finder

A view-finder is a mini-monitor attached to TV camera. The view-finder is either optical or electronic in nature. The optical view-finder is the least expensive of the three types available, but it does not represent the exact state of

affairs in recording. The other optical-type is the TTL (through the lens) type which diverts a certain amount of light to the eyepiece and gives a good idea of framing and focus, but deprives the lens of a part of the light. The electronic view-finder is expensive, but it is most useful because it produces a true assessment of the picture quality, framing, focus and contrast during recording. The final product is before your eyes as you proceed with the recording. An electronic view-finder also enables a replay of the video recording during a location shooting.

View finders vary in size from a one-inch viewer for a portable camera to a three-inch monitor for a professional studio camera which can be viewed with both eyes.

18.2.6 Camera Controls

The two principal controls in a camera are the iris control and the white balance. Both these controls are partly mechanical and partly electronic. The iris control regulates the amount of light reaching the vidicon tube. This is adjusted either manually or automatically by an aperture in the diaphragm, so that the tube is provided with an adequate amount of light. The iris control is governed by the video gain of the camera. It is an automatic control like the AGC and controls the amount of light reaching the tube. The white balance enables the colour balance to be adjusted on a domestic video colour camera. It is a combination of crude mechanical filtering and an electronic adjustment between red and blue. Manufacturer's instructions regarding white balance must be followed. Generally the camera is pointed at a white card in the existing light conditions or a white translucent cap is fixed over the lens. An indicator in the view-finder is then adjusted manually with a knob on the side of the camera. In certain cases this setting is automatic.

To avoid damage to the camera tube, the vidicon tube should never be pointed towards a strong source of light for a long duration. Under no circumstances should the camera be pointed directly towards the sun.

18.2.7 Electronic Circuits

The electronic circuitry that the video camera, including the view-finder has to share with the VCR is shown in Fig. 18.1 as a block diagram. During recording, the video signal from the vidicon tube is fed to both, the VCR and the view-finder. The sync and scanning pulses are also fed by the VCR to the camera tube and the view-finder. For monitoring during playback, the video signal from the tape is fed to the view-finder which serves as a miniature TV monitor on location shooting. The sync pulses, clipped from the video tape, provide horizontal and vertical deflection drives for the view-finder. The latest cameras supplied as accessories with modern VCRs used a 2 : 1 interlace, both when recording in colour or in black and white. The view-finder scans only frequency locked and not phase locked in playback. This slightly affects the framing of the picture in the view-finder during playback due to lack of phase-locking.

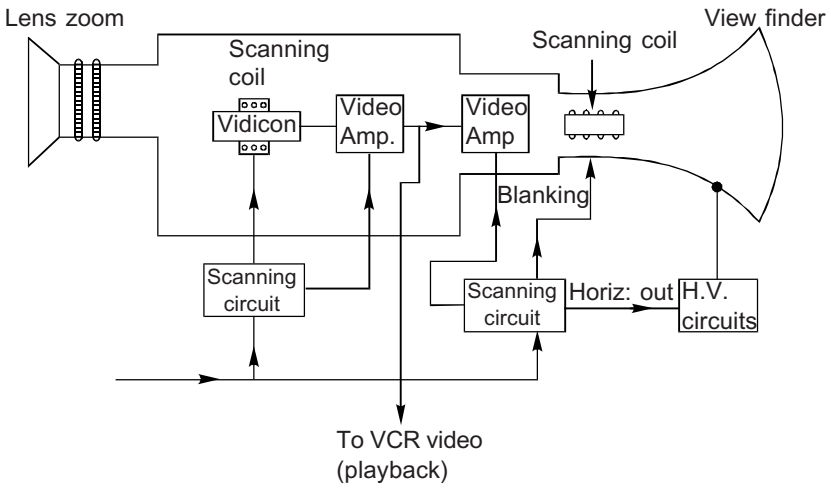


Fig. 18.1 Block Diagram of a Video Camera

Some of the less expensive black and white cameras, used in closed circuit TVs in surveilance operation use deflection systems that operate independently and the TV fields produced by such cameras are said to be randomly interlaced. The playback of recordings from such random interlace cameras contain heavy and coarse beat patterns.

Portable cameras use a 12-V supply from the VCR deck for the camera circuits and the 3-kV supply for the view-finder is derived from the incoming horizontal sync pulses. Thus, the view-finder loses its raster and the red light inside the eyepiece goes off if the machine stops or the tape runs out.

18.2.8 Camera Types

The two basic camera types are the heavy professional type and the light portable type. The former type requires the use of a tripod and is wheeled about in the TV studio with the cameraman seated on a trolley. The latter type allows greater mobility and can rest on the shoulder or can be carried in hand. Domestic VCRs require the use of portable cameras. Of all the various types of portable cameras in use, the ENG camera seems to be the most useful and popular.

ENG Cameras ENG camera or the Electronic-News-Gathering camera is a colour camera which is equally useful for studio work and for location shooting. ENG cameras are now marketed by all well known manufacturers such as Akai, Hitachi, Philips, RCA, Sony and JVC. This electronic device can provide instantaneous video pictures and is a very good replacement for a 16 mm movie camera.

Salient Features *Mechanical*: This is a light-weight camera which can either be held in hand or can rest on the shoulder. The weight of the camera varies from 0.5 to 5 kg. All optical and electronic components are generally fitted in a

light-weight housing which provides protection against dust, sun and moisture.

Optical: All ENG cameras are provided with a Zoom lens which can either be operated manually or by motorised action with variable speed. The iris control is also either manual or automatic. Dichroic filters or prisms are used to split the image into three primary colours, viz. red, blue and green. Colour temperature can be controlled by optical filters of 3200 K, 4500 K and 6000 K with 25% neutral density.

Colours are generally represented by their colour temperature, which is the temperature attained by a heated black-body, while emitting that particular colour. For example, the red light of an incandescent bulb is 2000 K, bright sunlight is 6000 K and white light has a colour temperature of 6500 K. It is at this level that TV monitors are adjusted: K stands for degrees kelvin, which is 273 +°C. Thus, 1000 °C is equivalent to 1000 + 273 = 1273 °K. Blue light corresponds to 9000 °K.

Most ENG cameras are three-tube colour cameras with 2/3'' Plumbicon tubes. Single tube cameras have also been manufactured by Hitachi and Sony which have the advantage of reduced size, weight and power consumption, but are slightly inferior to the 3-tube model insofaras excellence or colour registration is concerned. Circuits are provided for flare compensation, shading compensation, contour correction etc. to get a clear and sharp picture. Automatic controls are provided for sensitivity, iris control, white and black balance control.

View-Finder: The electronic view-finder monitors the output picture and provides immediate playback facilities after recording. The closing and opening of the iris and the white balance are also indicated by the view-finder. The condition of the battery is also indicated by the view-finder. The picture being slightly out of the focus with a run-down battery. Additional facilities are: a built-in colour bar generator, intercom facilities and a connection for an external microphone.

Power supply: A rechargeable nickel-cadmium battery pack for the power supply is either clipped on to the camera body or is worn in a battery belt. Ac adapters are available for mains operation. The consumption is low and varies from 25 to 30 Watts at 12V dc.

18.2.9 Precautions for the Use and Operation of a Video Camera

1. Never expose the camera to high humidity, dust, excessive vibrations or strong magnetic field.
2. Do not remove the lens cap, except during actual shooting and replace it immediately thereafter.
3. Do not allow direct sunlight to enter the camera and do not expose the camera to sunlight for long durations.

4. Do not touch the front surface of the lens with ordinary cloth or with fingers. If the lens is dirty, wipe it with soft cloth or a soft brush, moistened with a lens cleaner solution. Avoid scratching the lens.
5. If the deflection circuit stops accidentally, switch off the camera and cap the lens to prevent damage to the camera tube.
6. Avoid using the camera near a radio or TV transmitting antenna or motors with strong magnetic fields.
7. Do not allow any moisture to condense on the lens. Wipe it off with alcohol.
8. White balance adjustment should be carried out each time the scene changes. For this, press the auto-balance button for two or three seconds.
9. Choose the correct filter positions corresponding to the prevailing conditions of shooting.
10. Keep an eye on the battery voltage by observing the red LED indicator on the view finder. Flickering of the indicator represents low battery voltage.

18.3 TV BOOMS AND FISHPOLES

When the microphone is required to follow TV or film action, or move over an audience to pick up individual voices, an advanced form of boom is needed. A popular type has a long telescopic arm with the microphone suspended in a complete cradle at one end and a counterweight at the other (Fig. 18.2). The arm is supported on a pivoting pillar which is built on to a three-heeled trolley or pram.

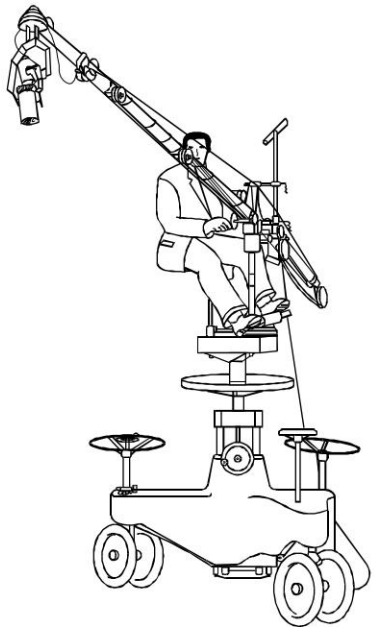


Fig. 18.2 Typical Trolly-mounted Microphone Boom

The operator stands on a platform on board the trolley and an assistant propels the whole boom assembly around the floor as necessary. The operator turns a handle held in the right hand to extend or retract the boom arm, and this action simultaneously shifts the counterweight through the smaller distance needed to maintain balance. Control of the microphone position in all three planes is carried out by the left hand. The operator can swing the arm round for horizontal position, tilt the arm on its pivot for height, turn a lever to rotate the microphone cradle for direction and squeeze the two grips of the lever to change the microphones angle of tilt.

Simpler mobile booms are available and also more elaborate designs with a seat for the operator, boom extension up to 5 m and a wheel-lock arrangement allowing sideways or 'crab' tracking as well as the normal back-wheel steering.

A skilled operator can control all this manoeuverability to keep the microphone continually at just the right position and angle to pick up the artists with the required sound balance in relation to each other and to any accompanying instruments or effects. In TV drama this is no mean feat, and is complicated by the need to keep the microphone and its shadow out of the picture. Operators need to know all about lenses and camera angles so that they can raise or retract the boom for long shots and move in for close-ups—a necessary procedure if visual and aural perspectives are to be kept in synchronism. Headphone monitoring may not give them an ideally clear indication of the sound balance, and so playback to and from the sound mixer is a further requirement.

Where less manoeuverability is needed, a lightweight microphone can be mounted on a fishpole. This is a single telescopic arm with rubber hand grips and three or four extension sections reaching up to, say, 4.5 m.

18.4 WIND SHIELDS

Air movement can lead to turbulence around the diaphragm and the phase-shift vent-holes (in pressure gradient microphones). This can have a devastating effect on a microphone's performance or, at the very least, introduce predictable bursts of overload distortion (blasting). Much of the energy in wind noise is at low frequencies or even subsonic, so that a high-pass filter designed to remove all frequencies below, say, 60 Hz can be quite effective. In addition, for close speaking or singing and outdoor work, some form of pop shield or windshield is often needed.

As with clamps, many microphones can be fitted with specially designed 'dedicated' wind shields, but there are some 'universal' models which will fit a range of microphone types.

18.4.1 Sound Pick-up

In TV broadcasting the sound signal and the picture signals are broadcast simultaneously but by modulation of separate carrier frequencies. The audio signal is also generated simultaneously alongwith the video signal in the same TV studio. For good quality sound, the placement and location of microphones

in the TV studio is as important as the placement of TV cameras for good quality video signal. The microphone placement technique depends on the type of program. The microphone can be visible to the viewers in programs like discussion, news items and musical programmes and so it can be fixed on a desk stand or a floor stand but for dramas, plays, serials and other such programmes the microphone has to be kept out of view. For this type of programmes either the microphone is kept hidden or mounted on a boom-stand and operated from trolly with the boom operator constantly adjusting the position of the microphone pickup so that it is close to the area of sound pick up and also kept high enough to be out of the camera frame. A trolley mounted boom can be wheeled round the studio and a skilfull operator can turn the microphone in all the three planes and adjust its height and position so as to pickup all the desired sounds but still keeping the microphone and its shadow out of the picture frame.

The type of microphone used in TV studio again depends on the type of programme to be recorded or telecast.

Directional Microphones are generally used to keep the noise level low because of the high level of ambient noise in a TV studio. For this reason the acoustic treatment of TV studios is such as to make them as dead as possible. Artificial reverberation methods are then employed to obtain proper audio quality for the sound. Lavalier or tie-clip type of microphones which can be pinned on to the tie or the lapel of the performer are used to avoid obstructing or hiding the face of the person being picturised as in the case of a news-reader. Radio microphone or cordless microphones are becoming quite popular in all free movement programmes with a hand held microphone. Sound is mono in most TV programmes but the use of stereo-microphones is becoming necessary where genuine stereo sound tracks on video recording of TV programmes are beginning to grow in importance. Full details of various types of microphones used in TV studio are available in Chapter 21.

18.5 PROGRAMME CONTROL ROOM (PCR)

The video and audio outputs from various sources like the studio and VTR are all routed through a camera control room called the programme control room (PCR). The entire programme is controlled and monitored by the programme director (producer) with the help of his assistant, a camera control unit (CCU) engineer, a video mixer expert, an audio engineer and a light director. They have in front of them a panel containing the camera control unit, a video mixer and a stack of monitors for previewing and editing all incoming and outgoing programmes. Another panel houses the microphone controls and switch-in controls for other auxiliary equipment. A talkback system enables the producer to give instructions to the camera man, the boom operator and the floor manager in the studio. The studio lighting can also be controlled from the PCR with the help of switches, and dimmerstats located in the PCR itself.

A PCR provides the following monitoring and control facilities for effective control and transmission of TV programmes.

18.6 CAMERA CONTROL UNIT (CCU)

The CCU controls the various parameters of the camera like the lens aperture, Zooming action, beam focus and the brightness control of the camera tube. The video gain camera sensistivity, blanking level and video polarity have to be carefully adjusted by the CCU engineer to conform to standard levels to suite different lighting conditions. The outside broadcast (OB) programmes like sports commentaries received on the microwave link are also demodulated and processed by the CCU to form a part of the programme broadcast on a particular channel. The inter connection between the various units, the sources and various areas are made by 75 ohm multicore coaxial cables.

The lighting control in the technical presentation panel is also located by the side of the CCU which is followed by a vision mixer or a video switcher.

18.7 VISION MIXER

A Vision Mixer or a Video Switcher is a cross bar type of switch which allows the selection of any one of a large number of inputs available and switching them on to any of the available output lines or outgoing circuits. The inputs available may consist of a number of cameras, VTRs, outputs from telecine machines besides test signals and outputs from special effects generators.

A vision mixer also enables the producer to produce special effects like fades, wipes, dissolves, superimposition of two signals and so on. By suitably mixing the output from different cameras at suitable angles, a two dimensional TV picture can be made to look like a three dimensional one. The ultimate destination of the output from the vision mixer may be a transmitter, a VTR or a number of monitors in a closed circuit television system. A monitor is a television receiver which does not contain any HF circuit and produces a picture and sound with the help of video and audio inputs only.

18.8 TYPES OF VIDEO SWITCHERS

There are three main types of video switchers

(i) *Mechanical Push button Switcher*: In this type of switcher the signals are terminated on the actual switch contacts. The bank of switches is interlocked to avoid any simultaneous operation of switches. This type of switch is mainly used in portable field units or in CCTV where any monetary disturbances in the picture during switching can be tolerated.

(ii) *Relay Switcher*: This system makes use of rack mounted reed relays for cross point switching. These reed relays have fast operating time of less than 1 ms allowing switching action to take place within the vertical blanking interval, preventing loss of signal during switching.

(iii) *Electronic Switcher*: This is also called a vertical interval switcher because it switches electronically one source to another during the vertical blanking interval following the vertical sync in a matter of a

few microseconds only. After the cut button is pressed a memory holds the information until the next field blanking period when the switching takes place. Electronic switches use solid state technology. The small size, fast action and inherent reliability make these switches eminently suited for all TV broadcasting systems. Figure 18.3 gives the scheme of a typical electronic video switcher. In this system each vertical row of switching buttons on a video switching panel corresponds to a single picture source and each horizontal row correspond to an output bus. The buttons operate electronic switches at the cross points of the input and output lines during the vertical blanking period and hence the name vertical interval switcher. The programme bus feeds the output to the transmitter line and the other bus provides signals from special effects equipment, A-B mixer and previous monitors. The output buses are interlocked, so that at one time only one picture source can appear, although a given output signal may be connected simultaneously to several output lines. The output of the adder o/p amplifier goes to a Master Control switcher where the entire system output is controlled.

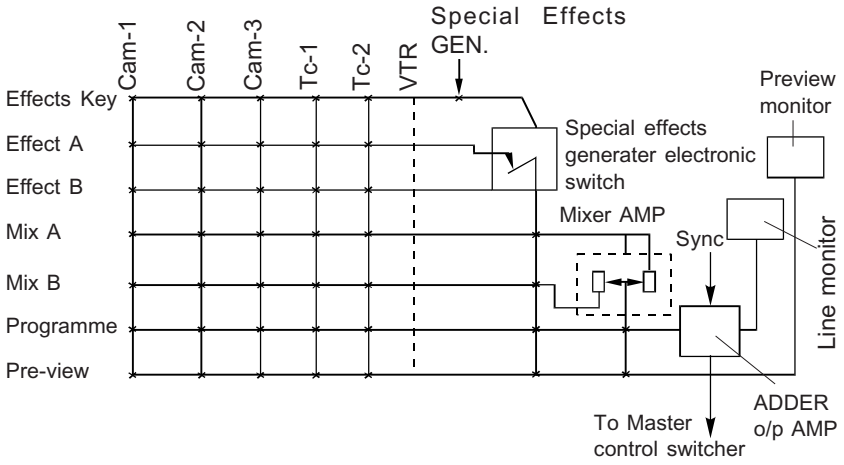


Fig. 18.3 Block Diagram of an Electronic Video Switcher

18.9 A-B MIXER

This type of mixer is formed by combining the outputs of two fading currents as shown in Fig. 18.4. This mixer can be used to superimpose two video input signals to get a camera output.

Each fading circuit employs a different amplifier with the input to the common emitter and output at the collector of one transistor which is turned on and off by its base bias resulting in signal fade out. Mixing is done by having a common collector load so that the output is an addition of the two input signals.

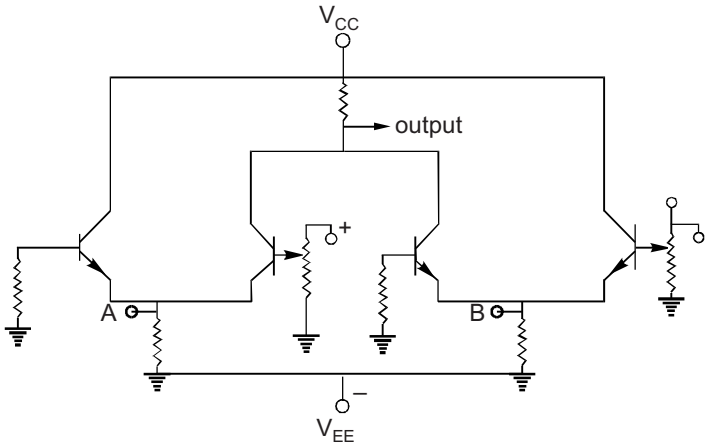


Fig. 18.4 A-B Mixer

The two mix buses A and B, help carry out the process of lap dissolve or fade-in and fade out process of mixing. Lap-dissolve switching is done by two potentiometers connected to the two signals that are to be switched. The signal amplifier from camera 1 is slowly reduced while that from camera 2 is slowly increased at the same rate as shown in Fig. 18.5(a).

The fade-out fade-in method is illustrated in Fig. 18.5(b). This can also be done by the potentiometer employed in the lap-dissolve method. However, in this method, by separate operation of the two potentiometers, Signal no. 1 is slowly reduced in amplitude until zero level is obtained, and the Signal No 2 is slowly raised from 0 to 100% by the second potentiometer. The amplifiers used are commonly called mixers or faders.

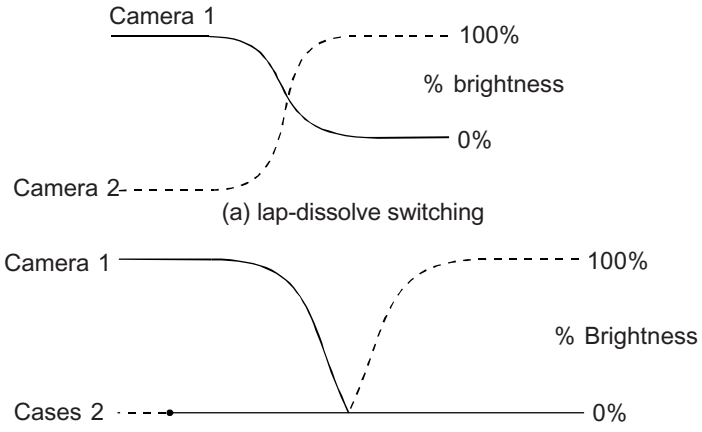


Fig. 18.5 (b) Fade-out Fore-in Transition

18.10 SPECIAL EFFECTS GENERATOR

As is clear from Fig. 18.3 the vision mixer provides signals for special effects on two rows of the vision switching panel. This is done with the help of a

waveform module that generates keying waveforms to produce certain special effects. These special effects include means for blanking out one or more parts of the areas of a signal picture of inserting other signals into the area. Certain effects both horizontal and vertical can be introduced. These signals are inserted while changing from one scene to another. In fact many patterns are available for interposing between any two programmes. Any required type of modules can be purchased with the equipment.

18.11 SYNC PULSE GENERATOR

In big TV Broadcast studios a number of cameras and other video equipment are used to produce a sequence of series for a meaningful picture. In order that change over from one camera to another does not produce any noticeable change in the picture on the TV receiver or the monitoring system, it is necessary that a common source of deflection pulses such as sync and banking pulses are used to keep the deflection circuits in phase with each other all the time. Not only that scanning in the various cameras and monitors in the studio is in phase but the scannig in the TV receivers should also be in perfect synchronism with the camera and other sources of video signals to the TV studio. This is done by providing a common "Station Sync" which is generally a crystal controlled sync pulse generator (SPG) as in Fig. 18.6.

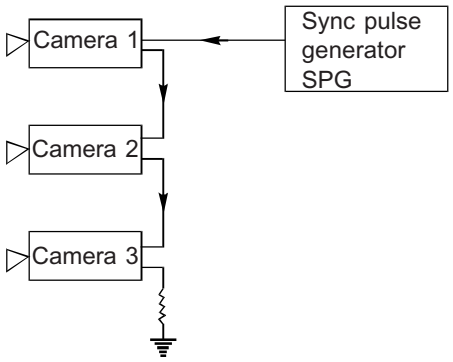


Fig. 18.6 Common Sync Pulse Generator (SPG) for a Multicamera System

The sync pulse generator provides the following controlling and deflection pulses:–

(i) Line drive or H-pulses at the rate of 15625 Hz
(ii) Field drive or V-pulses at the rate of 50-Hz
(ii) System blanking pulses at H and V rate
(iv) Sync pulses at H and V rate
(v) Camera blanking pulses at the H and V rate

An SPG produces these pulses which are passed through pulse shapers and then distributed to various units through distribution amplifiers.

18.12 AUDIO CONSOLE

In order to solve the complex problems of TV studio the Programme Control Room (PCR) is also provided with an audio console with a number of inputs and a cross bar switching arrangement for facing a number of output channels, each of which can select any mike and input line. These channels are also equipped with equalizing facilities for correction of frequency response, if required. Any of these channels can also be connected to an echo chamber for providing echo facilities. The other complement of the PCR equipment include two tape recorders and a playback turn-table for playback of audio effects.

18.13 MASTER CONTROL ROOM (MCR)

In a large TV studio complex consisting of a number of TV studios and production control rooms, the outputs from various sources are routed through the Master Control Room (MCR) which houses a master switcher and video equipment associated with each studio is available on separate racks in the MCR. This room also houses centralized video equipment like sync pulse generators, special effects generator, test equipment, video and audio monitors.

To ensure good quality and linearity before the signal goes to the transmitter, special video monitors employing 90° deflection picture tubes rather than the common 110° picture tubes are used.

In the master control room the composite video signal is raised to about 1 volt p-p level before feeding it to the cable that connects the control room to the transmitter, when the transmitter is located in the same building. However, matching network are provided at both ends of the connecting cable to avoid unnecessary attenuation and frequency distortion.

Microwave links are used as Studio to Transmitter Links (STL) when the studio is not very close to the transmitter installation. Microwave links are also used when relaying live programmes from Outside Broadcast (O.B) point like sports stadium and other special functions.

Direct Coupled (DC) amplifiers are used to preserve dc component of the signal at the camera level output but the use of AC coupling is unavoidable further and any loss of dc component is made up at the transmitter before modulation by the use of a dc restorer circuit often called a blanking level clamp.

18.14 TELECINE MACHINES

Many TV programmes originate from the use of 35 mm or 16 mm photographic camera film and slides are also projected in the TV programmes. Telecine film camera chain which couples the film projectors or the slide projectors forms a very important programme source for a TV studio. Special type of vidicon cameras are used for the projection of cine film and these cameras are different from vidicon cameras used in TV studios in having a much reduced lag factor.

For high utilization of the film camera chain, it is usually convenient to use a multiplex type of system where a single television camera is used for three or four film sources viz 35 mm or, 16 mm film and a film projector. For the accompanying sound track pick up, the visual, optical or magnetic track playback facility is incorporated in the multiplexer set up.

One important problem that is faced in the use of telecine machines is the difference in the frame rate used in motion pictures and the frame rate used in television scanning. The Cinema pictures use a frame rate of 24 frames/sec which is increased to 48 frames/sec by projecting each frame twice to reduce flicker. In TV scanning a frame rate of 25 frame/sec is increased to 50 fields/sec in interlaced scanning. The difference of frame rate or frame/sec must be corrected to avoid a rolling bar on raster besides loss of some signal output. To overcome this difficulty, the film in the telecine equipment is pulled down by the shutter mechanism at the rate of 25 frames/sec. The distortion caused by this small change in speed is hardly noticeable on the screen. The slight change in the pitch of the sound will also go undetected.

Similarly in the 525/30 TV system, it is necessary to convert 24 frames per sec into 30 frames/sec rate to avoid presentation of incomplete information during some scanning periods. This is generally done by projecting one frame three times and the next frame two times and repeat alternately by means of an intermittent mechanism of 3 : 2 pull down cycle for the film. The pull down is covered by 60 Hz shutter. The first set of alternate 12 frames are thus projected three times and the remaining set of alternate 12 frames are projected two times giving $(12 \times 3 + 12 \times 2) = 60$ fields/sec from the 24 frames/sec.

The speed alternation and sequencing explained above makes direct scanning of motion picture possible through TV network. While slides are used for stills, small advertisement films are recorded beforehand on video tapes for telecasting when required.

18.15 TELEVISION PICTURE RECORDING (VIDEO RECORDING)

Television picture recording, or video recording as it is popularly known, plays the same important part in a TV Broadcast system as does the sound recording system in audio telecast system. The video recording system was originally demonstrated in 1926 by J.L. Baird, whose pioneering work in the development of modern TV system is of utmost importance. Baird recorded pictures on a wax cylinder of the type used by Eddison for sound recording but video recording has since made very great strides and the present day video tape recording (VTR) and the video disc recording (CD) systems form an integral part of any high quality TV system. Of the various methods used for the recording of TV pictures like the photographic method of recording pictures from a high quality TV monitoring screen (Konoscope), the modern method of recording TV picture on a magnetic tape has almost superceded all other TV picture recording systems. The use of telecasting tape recorded pictures was

made by BBC in 1955 when the VERA (Vision Electronic Recording Apparatus) developed by BBC and DECCA in Britain was used for the purpose. Video machines devised by RCA and AMPEX were used in America for TV broadcasting. All these machines which used longitudinal system of recording were cumbersome, difficult to operate and maintain, have since given place to the most modern and sophisticated forms of VTR and VCR machines using transverse system of tape recording and many other electronic technologies. Digital recording in place of analogue recording has introduced a new dimension of high quality recording both in audio and video systems of tape recording.

Video tape recording which has the original advantage of erase and re-record, has also the capability of editing and duplicating without delay. Programmes can be recorded weeks and months in advance to be played back at the time of telecast. Further, VTR has added advantages of (i) immediate playback (ii) convenience of repeating the recorded programme as many times as necessary and (iii) making as many duplicate copies for 'distribution to other users.

Full details of the various methods used for TV picture recording (Video recording) have been discussed in Chapter 22 of this book under the heading ''Sound and Video Recording System.''

18.16 TV TRANSMITTERS

A TV transmitter is a combination of a picture (Video) transmitter and a sound (audio) transmitter. The outputs from both the transmitters are fed to a common transmitting antenna through a common feeder line connected to a combining unit called the diplexer. Whereas the video (picture) transmitter is amplitude modulated, (AM), the audio (sound) transmitter is frequency modulated (FM). Both the transmitters operate in the VHF or UHF range where all the TV channels are located and the techniques used for the design of the various stages is the technique applicable to such HF transmitters.

The power of a picture TV transmitter is generally given in terms of Effective Isotropic Radiated Power (EIRP) of the peak carrier corresponding to the sync peaks modulated at 100%. The sound transmitter is rated in terms of the RMS value of the carrier and its value is 10 or 20% of the video power. Even the lower power rating of the aural transmitter is considered satisfactory from the point of view of having noise free reception over the same range as the visual transmitter using amplitude modulation. Greater powers for audio are not only uneconomical but also cause complication in colour TV transmission due to the sound carrier beating with the colour sub carrier. Thus for a 10 kW picture carrier a 2 kW sound carrier is adequate for satisfactory TV transmission and reception.

A block diagram of a 10 kW-VHF TV transmitter is given in Fig. 18.7.

The 10 kW final power amplifier is grid modulated by the video signal of 500 VA supplied to the video amplifier. The Video signal is processed through

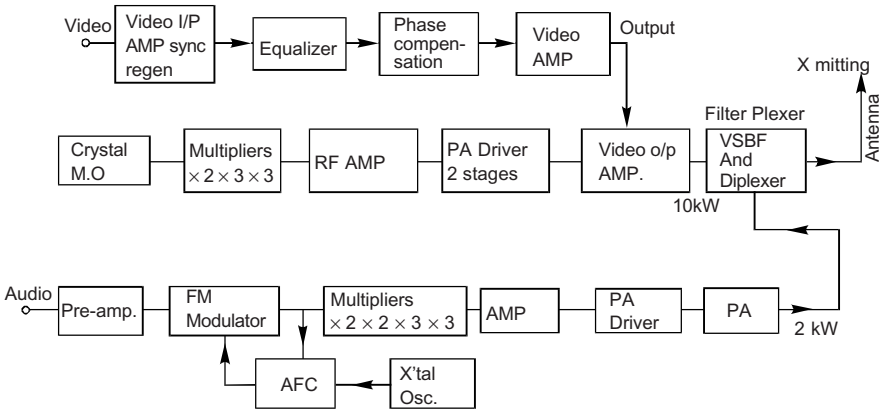


Fig. 18.7 Block Diagram of a 10 kW VHF Transmitter

the *visual exerter unit* consisting the video I/P amplifier and sync regenerator, the equalizer and the phase compensator. The incoming audio signal is processed in the *aural exciter unit* when the incoming audio frequency modulates a crystal controlled oscillator which drives a chain of multipliers to raise the FM carrier frequency to the final sound carrier frequency for the channel used. The power amplifier in the FM chain produces the 2 kW output power which combines in the diplexer unit with the visual power after the latter has been passed through the VSB filter. In this case the VSB filter and the diplexer unit are combined to form a single compact unit called a *Filter blixer*.

18.17 TV TRANSMITTING ANTENNAS

A TV antenna has already been described in detail in Chapter 13 of Part III of the book. It is an omnidirectional antenna which is kept as high as practical to increase the line-of-sight distance and hence the range and coverage of the TV transmitter. Doordarshan TV antenna at Pitampura, New Delhi is one of the highest and contains a number of architecture and electronic features. A photograph of the Doordarshan TV Tower is shown on next page.

18.18 TV RELAY SYSTEMS

TV signals are very high frequency signals and the distance TV waves can travel in free space is limited to only line of sight distance. When the distances over which the TV signals are to be transmitted are larger than the line-of-sight distances, use is made of TV relay systems. One of the most common relay systems that is used for the relay of TV signals is the MICROWAVE COMMUNICATION LINK SYSTEM.

Microwaves are electromagnetic waves whose wavelength varies from 1 metre down to 1 mm with corresponding frequency of 300 MHz to 300 GHz (1 GHz = 1000 MHz). At microwave frequencies, it is possible to obtain very

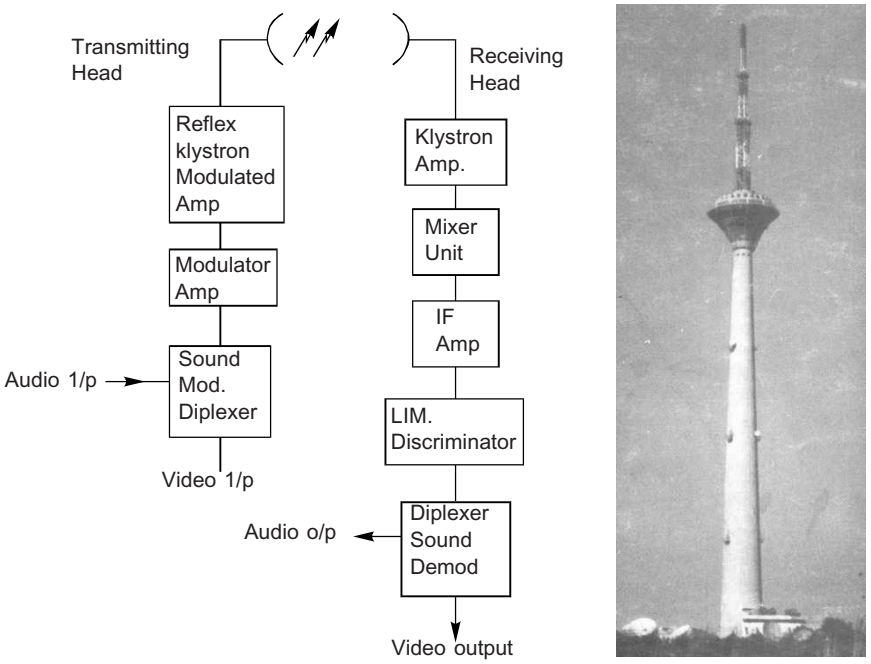


Fig. 18.8 Block Diagram of Microwave TV Relay System

high power gain by concentrating the radiated power from the antenna into a narrow beam by employing parabolic reflectors to replace dipole antennas. The frequencies normally employed for the Microwave relay links are 2 GHz, 7 GHz and 13 GHz bands-7 GHz being the most common for the TV broadcast relay links.

In TV broadcasts, Microwave links are commonly used for the following purposes.

18.18.1 Studio-To-Transmitter Link (STL)

Microwave links are used for the relay of programmes from the studio to the transmitter when the transmitter installation is not very close to the studio. When the studio and the transmitting antenna are in the same premises coaxial cables are used for feeding the signal from the studio to the transmitter.

18.18.2 Outside Broadcasts (O Bs)

When live programmes required to be relayed from places remote from the TV studio like the sports commentaries from sports stadiums, commentaries for Republic Day Parade, Independence Day programmes etc., Microwave links are established between the TV studios and the O B Spot. Use is also made of O B. Vans, which are fitted with all necessary equipment for processing the weak signals from the OB spot and transmitting the same to TV studio. The OB van is stationed at a convenient point near the OB spot so that a Micro-

wave Communication link can be easily established between the transmitting antenna on the OB van and the receiving antenna at the TV studio.

18.18.3 Inter-Station TV Relay Systems

For interchange of TV programmes between TV stations, microwave links are commonly employed at 50 to 70 kms apart depending on the terrain condition between the two stations. The very high gain parabolic antennas used are such that even a fraction of a volt of transmitted power is sufficient to establish an efficient relay link between the two relay points. The system is a broad band system which enables a number of TV channels to be handled with a single microwave link.

A simplified block diagram of a Microwave TV relay system is given in Fig. 18.8.

The relay system shown in Fig. 18.8 makes use of a klystron microwave tube both at the transmitter and the receiver. A klystron is a special type of vacuum tube which can develop and amplify much higher powers at microwave frequencies than is possible with orthodox vacuum tube, which failed to develop suitable powers at these frequencies due to limitations of transit time.

The klystron amplifier at the transmitter is frequency modulated by the video signal alongwith audio signal. At the receiving end, the microwave signal is received and amplified with a klystron amplifier. It is then converted to IF in the mixer unit amplifier and the demodulated in a limiter discriminator to get back the video and audio signals.

In a TV relay system one receiving and transmitting antenna faces one direction and another similar pair faces the other side.

18.19 SATELLITE RELAY SYSTEM

A communication satellite is one of the best TV relay systems. It is essentially a microwave link repeater. It receives the energy beamed at it by an earth station, amplifies and returns it to earth at a different frequency to prevent interference between up link and down link.

A detailed description of satellite communication is given in Chapter 20 under satellite TV.

Summary

The video and audio signals broadcast by a TV antenna are processed in the TV studios using special technique and equipment.

A TV studio is different from a sound studio in the sense that it is acoustically treated, it is also properly illuminated by flood lights where the illumination level is controlled by dimmerstats and light fitting on the cerling to obtain desired effects without casting any shades.

A variety of cameras are used to obtain different views of the scene to be televised. These cameras are professional cameras which can be wheeled about with the cameraman seated on the trolly. The cameras are filled with special

lenses including the zoom lens and view finders. Different types of microphones including lapel microphones, tie-clip microphones and radio microphones are used for sound pickup.

The programmes from various studios are routed through the Programme Control Room where various type of switches for camera control, switching and monitoring are located. Equipments like camera control unit, vision mixer, A-B mixer, special effects generator, sync pulse generators, special effects generators form a part of the Master Control Room, Arrangements are available for telecasting of films, slides and also for making radio recordings.

The TV transmitter is linked to the TV studio either through a microwave link or a coaxial cable. The transmitter consists of a HF transmitter which is a combination of vision transmitter and a low power sound transmitter. The TV antenna is an omnidirectional antenna.

OB Vans fitted with microwave links are used for outside broadcasts. Microwave links are used for interstation relays. Satellite relay system is also becoming popular for interstation relays.

Review Questions

1. Explain the difference between a sound studio and a TV studio.
2. Give details of the lighting arrangements in a TV studio. How is lighting arranged to prevent shadows and produce the desired contrast effect?
3. What type of cameras are used in a TV studio? Explain in detail how a TV camera is different from a photographic camera.
4. Describe the salient features of an ENG camera and give examples of its practical applications.
5. What type of microphones are used in a TV studio? Explain the type of stand and forms that are used for sound pick up without exposing the microphones to viewers.
6. Explain the various monitoring and control facilities provided by a PCR
7. Explain with a block diagram how the program from a TV studio is routed to the TV transmitter when the transmitter is located at some distance from the TV studios.
8. Give the block diagram of a TV transmitter and briefly explain its functioning.
9. Explain the main difference between a TV transmiting antenna and a TV receiving antenna.
10. Describe a 3-element TV receiving antenna giving details of the relative lengths and distances between the various elements.
11. Describe the rise of microwave links for OB and interstation relay of TV programmes.

chapter 19

Television Receivers

19.1 RADIO RECEIVER VS TELEVISION RECEIVER

The difference between a radio receiver and a television receiver lies in the fact that whereas a radio receiver is required to convert only a modulated sound carrier wave into sound through the loudspeaker, a TV receiver, on the other hand, is required to convert not only a modulated sound carrier wave into sound through its loudspeaker, but also a modulated picture carrier wave into a picture through its picture tube. A TV receiver has, therefore, to perform more functions and as such uses more stages and more components and the technical processes involved are more complicated than in a radio receiver. The block diagram of a receiver given in Fig. 19.1 will be used to explain the functioning of the various stages in a TV receiver. It may be seen that this block diagram pertains to a black-and-white or monochrome TV receiver. However, it may be noted that a colour TV receiver is basically a monochrome receiver with a few additional circuits added to introduce colour into the picture information. A colour TV receiver will be described in detail in a later part of the chapter.

In effect, a TV receiver is a combination of an AM heterodyne receiver for the picture and an FM receiver for the associated sound signal with a few additional circuits for scanning and synchronization of the picture on the TV picture tube.

19.2 USE OF TRANSISTORS IN TELEVISION RECEIVERS

A transistor can perform more or less the same functions as a vacuum tube. Besides, a transistor has many advantages over a vacuum tube like small size, rigid construction, longer life, absence of heater and operation at low voltages. Moreover, a transistor is better suited for sweep circuits which form an important part of a TV receiver. In view of these advantages, transistors are rapidly replacing vacuum tubes or valves in TV receiver circuits.

19.2.1 Commonly Used TV Receivers can be Divided into the Following Three Categories

(*a*) *All-valve type TV receivers*
These TV receivers use only vacuum tubes in their circuits. All the functions are performed by tubes, some of which are multipurpose tubes.

(*b*) *Hybrid TV receivers*
These TV receivers use both valves and transistors in their circuits. Signal-circuits are generally transistors and ICs, whereas deflection circuits mostly use power tubes.

(*c*) *Solid state TV receivers*
These TV receivers use only transistors, ICs and semiconductor devices in all stages except the picture tube. In this type of receiver, only the picture tube requires warm up time. The picture takes sometime to appear while the sound comes through instantly. The warm-up time for the picture tube can, however, be considerably reduced by keeping the tube filaments warm with about one-third of the filament current on by a special standby mode switch.

Basically, all TV receivers whether all-valve type, hybrid or solid state, have the same stages which more or less perform the same functions. As such, the block diagram shown in Fig. 19.1 applies to all the types of TV receivers mentioned above. Functions of various stages represented in the block diagram will now be described in detail.

The block schematic in Fig. 19.1 can be divided into the following main sections for a detailed discussion of the various stages in a representative black-and-white television receiver.

1. Tuner or RF section
2. Picture (video) section
3. Receiver sweep section
4. Sound section
5. Power supply section.

19.3 TUNER OR RF SECTION

Tuner or RF section actually consists of two stages—RF amplifier and frequency changer. The RF amplifier generally contains one or two stages of RF amplification meant to provide the necessary amplification for the signals passed down to it from the TV antenna through the feeder line. These stages help in achieving a good noise figure: they produce image rejection and prevent radiation of local oscillator energy.

The amplified signals consisting of the picture and sound signal are passed on to the frequency changer stage where these are converted into their respective IF signals by heterodyning with a higher frequency produced by the local oscillator. In the case of TV receivers designed for use in India, the IF for the

picture carrier is 38.90 MHz and the IF for the sound carrier is 33.40 MHz. The actual frequency generated by the local oscillator will depend on the signal carrier frequency so that the difference of the two frequencies is always equal to the IF frequency.

For example, the Delhi station operates on Channel IV of Band I with a picture carrier of 62.25 MHz and sound carrier of 67.75 MHz. Accordingly, the local oscillator will generate a frequency of 101.15 MHz so that the picture IF produced is 101.15 – 62.25 = 38.90 MHz and the sound IF is 101.15 – 67.75 = 33.40 MHz.

The intermediate frequencies produced by the frequency changer are then passed on to the IF amplifier stages for necessary IF amplification.

The tuner circuit described above is for single channel but multichannel tuners are used when the TV receiver is required to operate on a number of channels.

Multichannel Tuners In a multichannel tuner the coils and other associated components required for tuning to different channels in different bands or in the same band are brought into the circuit by a suitable switching arrangement. Two systems commonly used are: (a) turret tuning, and (b) incremental tuning.

In turret tuning, the components for each channel in the form of a printed circuit are mounted on a turret which can be rotated with the selector switch to bring into the circuit the desired set of components. A view of a turret type multichannel tuner is shown in Fig. 19.2.

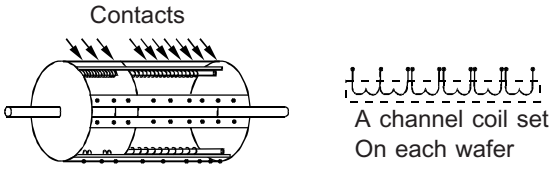


Fig. 19.2 A Turret Tuner

In the case of incremental tuning, tapped indictors are used together with a rotary switch as shown in Fig. 19.3. In any position of the selector switch a portion of the inductance is shorted and only the amount of inductance required for that particular channel remains in the circuit. The arrangement shown in Fig. 19.3 is for a six-channel tuner.

Most modern incremental tuner are of the wafer type. The coils for the three tuner stages viz RF stage, the mixed stage and the oscillator stage are mounted on separate wafers which are ganged together and can be rotated by a common shaft as shown in Fig. 19.4.

The number of turns used on the coils progressively decrease for the higher channels. Resonance is obtained with the help of the inductance of the coil and the stray capacitance, distributed capacitance and small trimmers used wilth each coil. For fine tuning, the oscillator coil is provided with a brass or aluminium core which varies the inductance when moved because of the production of eddy currents in the core.

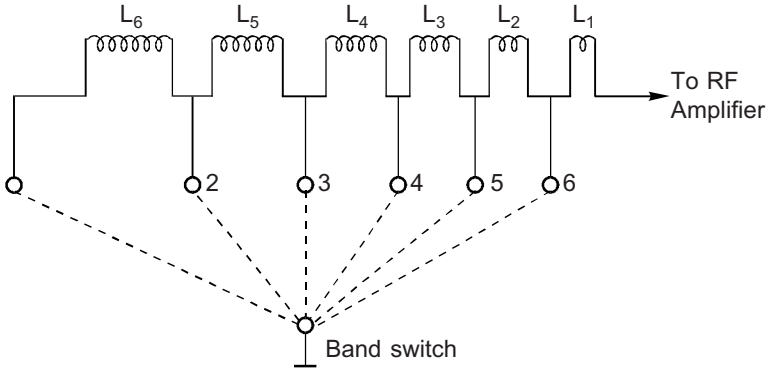


Fig. 19.3 Incremental Inductance Tuner

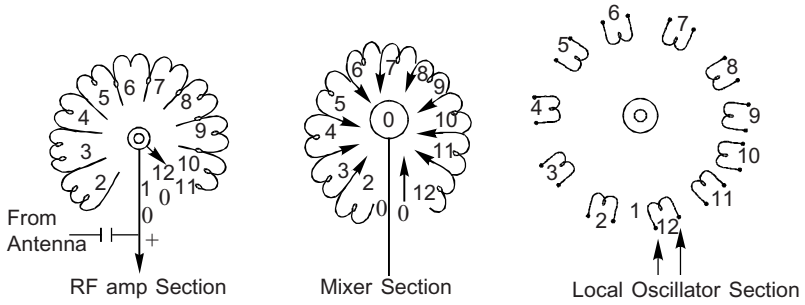


Fig. 19.4 Wafer Type of Tuner

Automatic Frequency Tuning is also used particularly in Colour TV receivers. This is done with the help of an AFC or automatic frequency control circuit provided with the local oscillator.

With the increase of number of channels used in the TV broadcast these days, the receivers are generally provided with VHF tuners (41-230 MHz) and UHF band tuning (470 to 890 MHz). Usually the UHF channels are converted to VHF channels No. 5 or 6, and these are further processed by the VHF tuner in the normal way.

19.3.1 Electronic Tuning

In the mechanical type of tuner described above, the entire set of components for the resonant circuit for one channel are changed to another set of components for another channel by changing the position of the tuning switch. In electronic tuning, the tuning is done with the help of a varactor diode whose capacitance varies inversely with the amount of reverse bias applied across the diode. Figure 19.5 shows the basic circuit of an electronic tuner.

Without any other voltage, the tuned circuit is resonant to any particular frequency. However, the DC tuning voltage supplied through R_2 varies the reverse bias on the varactor diode $D1$ and its capacitance changes to vary the

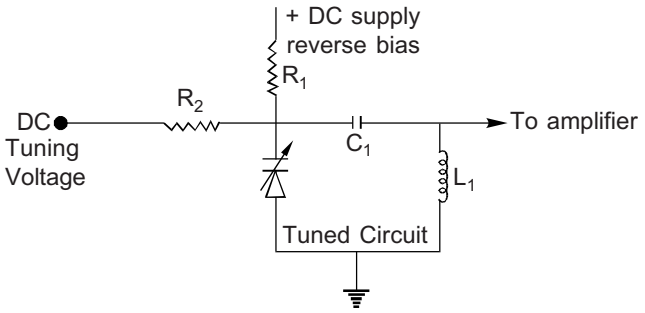


Fig. 19.5 Basic Circuit for Electronic Tuning

resonant frequency of the tuned circuit. The capacitance C decreases with more reverse bias.

The coupling capacitor C_1 has little reactance at resonant frequency but it prevents the dc supply voltage from shorting to ground through L_1. The DC voltage through R_1 provides the reverse bias for D_1 being positive at the N-cathode of the diode. Separate varactor diodes provide electronic tuning for the oscillator, mixer circuit and the RF amplifier. All the tuned circuits of a channel are adjusted simultaneously to ensure proper tracking.

19.4 IF AMPLIFIER

The intermediate frequency amplifier stage in a TV receiver is meant to provide necessary amplification for the picture IF signal (38.9 MHz) and the sound IF signal (33.4 MHz) received from the frequency mixer stage and to reject certain unwanted frequencies from the adjacent channels. This stage is required to perform the following functions:

(a) The IF amplifier is required to provide necessary IF amplification with appropriate bandwidth to accommodate the highest video modulation frequencies which can extend up to 5 MHz.

For the required IF amplification, three or four stages of amplification are used so that even the weakest signals can be received without much interference. A number of methods are available for obtaining the desired low Q tuned circuits, overcoupled tuned circuits and *stagger-tuned* circuits. Low Q circuits are obtained by putting resistors across the tuned circuits and the use of overcoupled tuned circuits for increasing the bandwidth has already been explained in the case of radio receiver IFTs. In stagger tuning the desired overall response is obtained by using three single-tuned circuits, each circuit tuned to a frequency slightly different from the frequencies of other circuits as shown in Fig. 19.6. One stage is tuned to the centre frequency which may be near the picture IF frequency, while the other two are tuned to frequencies slightly above and below the centre frequency. This produces a wideband frequency response as shown in Fig. 19.6.

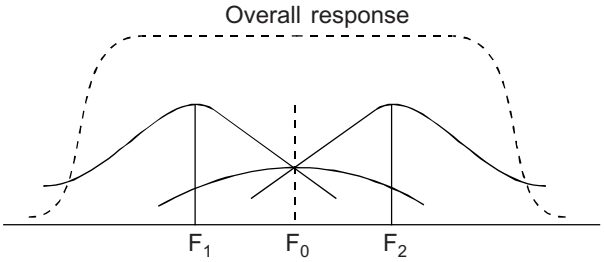


Fig. 19.6 Stagger-tuning

(b) The IF amplifier is also required to reject certain interfering signals from adjacent channels.

The production of interfering signals from adjacent channels is explained in Fig. 19.7. When the receiver is tuned to channel 3 band 1 (54-61 MHz) and the lower adjacent channel is channel 2 (47-54 MHz) and the upper adjacent channel is channel 4 (61-68 MHz).

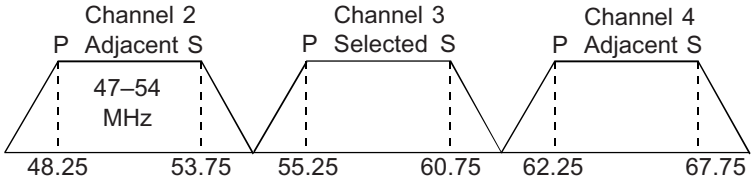


Fig. 19.7 Adjacent Channel Interference

Local oscillator frequency for channel 3 will be 55.25 + 38.9 = 94.15 MHz. This local oscillator frequency will beat with the sound carrier of lower adjacent channel 2 to give a beat frequency of 94.15 – 53.75

= 40.4 MHz.

Similarly, the picture carrier of upper adjacent channel 4 will beat with Local Oscillator (LO) frequency to produce another interfering signal frequency 94.15 – 62.25 = 31.9 MHz. Trap circuits have to be introduced to give an attenuation of about 40 dB at these interfering frequencies.

Video IF response of a TV receiver is shown in Fig. 19.8. Besides indicating the position of the trap circuits, it fulfills the following other requirements of the TV receiver:-

1 The picture IF (38.9 MHz) lies 6 db (50%) below the max. response of 100% between 35 to 38 MHz.
2. The bandwidth available should be 5 MHz between points where the response is 6 db down.
3. The sound IF (33.4 MHz) should be 20 db below or 10% of the level of picture IF.
4. The response should be flat within ± 2dB between 35 to 38 MHz

The IF response also provides necessary correction for the vestigial sideband modulation used in the transmission of TV signals (see Section 15.5).

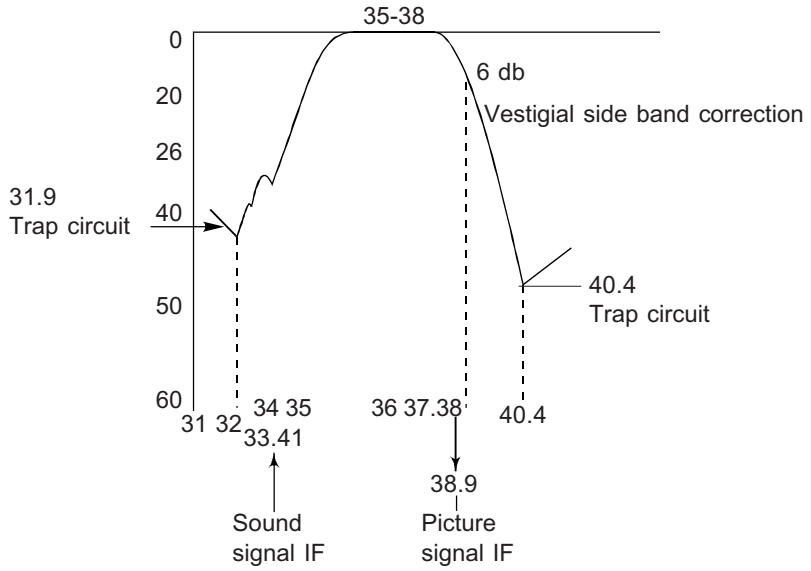


Fig. 19.8 Video IF Response Indicating the Position of Trap Circuits.

Detailed circuit of a video IF Amplifier: Full details of the video IF amplifier of a solid state TV are given in Fig. 19.9. Sound IF signal of 5.5 MHz is separated out by a special filter circuit consisting of L_3 and C_3 and this is passed on to the sound IF amplifier in the sound section of the TV receiver.

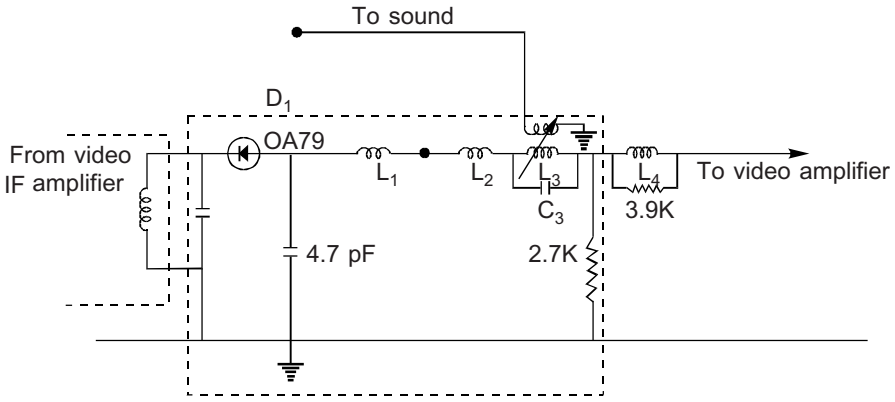


Fig. 19.9 Video Detector

19.5 VIDEO DETECTOR

The video detector in a receiver is required to perform the following two functions:

(a) To detect the amplitude modulated picture carrier IF and produce the video signal for modulating the electron beam in the picture tube.

(b) To produce and separate out the inter-carrier sound IF of 5.5 MHz by the beating together of the picture (38.9 MHz) and sound (33.4 MHz)

IF carriers. The sum frequency of 72.3 MHz produced by the rectifying action of the diode is filtered out and the difference frequency of 5.5 MHz is separated from the detector circuit and fed on to the sound section for further processing.

The circuit of a video detector used in most modern TV receivers is Fig. 19.9. A semiconductor diode type OA79 is used as the detector with a diode load of 2.7 kΩ connected across it. This small value of load resistance is necessary to obtain wideband response from the detector circuit. Various IFs and other frequencies produced during detection are filtered out by the 4.7 pF capacitor and the inductances L_1 and L_2 except the video signals which are passed on to the video amplifier stage. The inter-carrier sound IF signal of 5.5 MHz is separated out by a special filter circuit consisting of L_3 and C_3 and this is passed on to the sound IF amplifier in the sound section of the TV receiver.

19.6 VIDEO AMPLIFIER

The video amplifier stage provides sufficient amplification for the video signal from the detector to enable it to properly modulate the electron beam in the picture tube. This stage also provides the required voltage for the sync separator and AGC stages of the TV receiver. As in the case of the detector stage, the main requirement of the video amplifier is to produce a sufficiently strong video signal with uniform response for the entire range of picture signal frequencies from 0 to 5 MHz. This wideband frequency response is obtained by the use of an RC-coupled amplifier stage employing such frequency compensating devices as the use of series and shunt peaking coils, selective negative feedback and the use of an emitter follower stage after the video amplifier.

The output from the video amplifier stage is fed to the cathode of the picture tube via a 0.1 μF capacitor with the diode type OA79 connected in parallel with the capacitor. The diode provides the DC component for DC restoration so that the average brightness of the picture scene corresponds to the actual brightness of the scene being televised.

19.7 DETAILED CIRCUITS

Detailed circuits of Tuner, Video IF amplifier and transistorised keyed AGC stages of a transistorised TV receiver are given below:

19.7.1 Tuner

The tuner stage of a transistorised TV receiver consists of an RF amplifier, a mixer and a local oscillator as shown in Fig. 19.10.

The signal from the antenna is applied to the base of the RF amplifier transistor Q_1 (BF167) through the Balun transformer T_1 which matches the balanced impedance of the transmission line to the unbalanced input impedance of the transistor Q_1. Capacitors C_3 and C_4 of the resonant circuit $L_1C_3C_4$ provide necessary matching conditions of the base input impedance of the RF

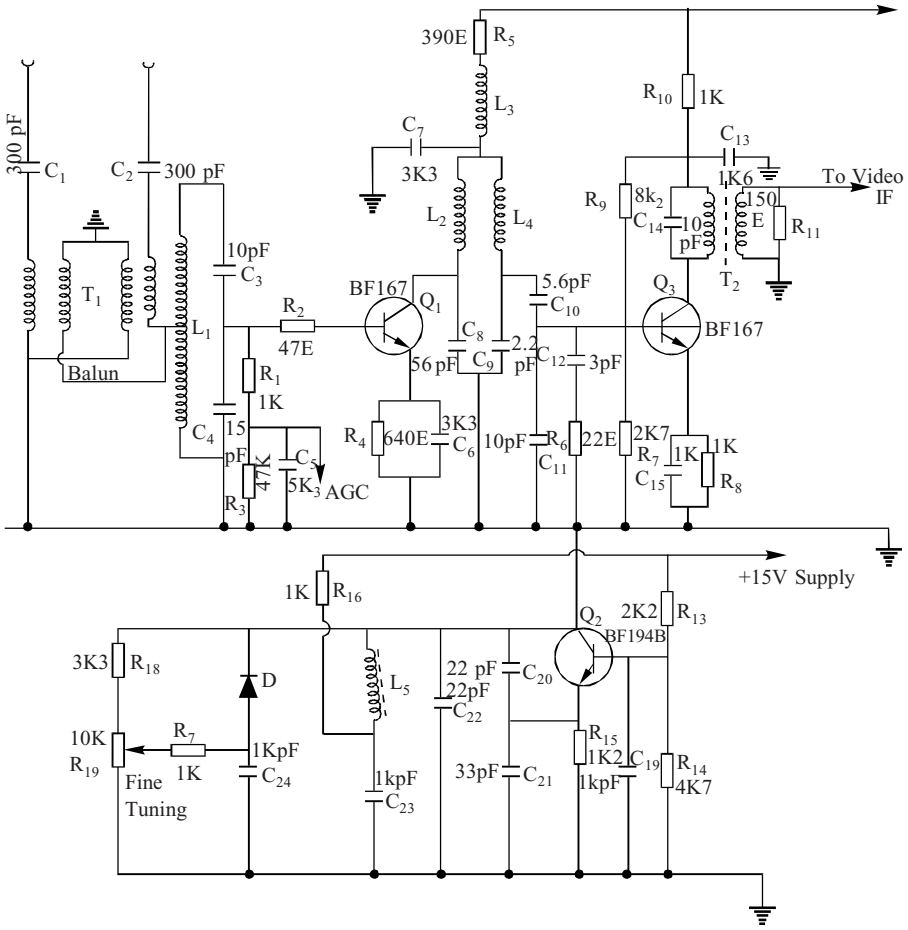


Fig. 19.10 Tuner Stage

transistor Q_1. AGC is applied to the base of Q_1 through resistance R_1 with capacitor C_5 acting as the decoupling capacitor. The RF signal after amplification is transferred from the collector Q_1 of the base of the mixer transistor Q_2 through the bandpass filters formed by the tuned circuits L_2—C_8 and L_4—C_9. Capacitors C_{10} and C_{11} provide the matching conditions.

Local Oscillator

Transistor Q_2 (BF194) acts as a modified form of Colpitts oscillator to produce the local oscillations required for the mixer stage. The tuned circuit is formed by L_5, C_{22}, C_{20} and C_{21}. The emitter feedback voltage is provided by C_{21}. The frequency of the resonant circuit can be varied by the variable capacitance of the Varactor diode D which is only the base-collector junction of the transistor BF194B and is connected across the resonant circuit. The capacitance provided by a Varactor depends on the DC voltage applied across it. This DC voltage can be varied by means of the potentiometer R_{19}, which serves as the fine tuning control normally provided on the front panel of TV receivers.

Mixer

The oscillations produced by the transistor Q_2 are applied to the base of the mixer transistor Q_3 (BF167) through R_6 and C_{12}. Here the local oscillator frequency and the amplified signal frequencies from the RF transistor Q_1 are produced by heterodyning action of the picture and sound IFs of 38.9 MHz and 33.4 MHz respectively. Both the IFs appear in the collector circuit of Q_3 and are selected by the wideband IF transformer T_2. The two IFs are then passed on to the video IF amplifier stage.

19.7.2 Video IF Amplifier

As in a superheterodyne radio receiver, the IF amplifier stage in a TV receiver also provides the major part of the RF amplification required to raise the strength of Rf signals to a sufficiently high level before being detected by the detector stage of the receiver. Besides providing the required RF gain, the video IF stage in a TV receiver is also required to provide rejection for the interfering signals from adjacent channels and to produce waveshaping of the signals, as already explained in the case of an all-valve TV receiver. Availability of HF transistors like BF 196, BF 173 and BF 197 and the use of CASCODE configuration has made it possible to design a stable IF stage without neutralisation.

In the transistorised IF stage of a TV receiver shown in Fig. 19.11, the required IF amplification is obtained by three stages of amplification using four transistors. The output from the tuner stage is applied as input to the base of the first IF amplifier consisting of transistor Q_1 (BF196/BF167). As in the case of a valve type TV receiver, three wavetraps C_1L_1, $C_2L_2C_3$ and C_4L_3 tuned to 31.9 MHz, 33.4 MHz and 40.4 MHz respectively are connected between the tuner stage and the input to Q_1. These wavetraps are meant to reject the interfering signals from the higher adjacent channel picture carrier (31.9 MHz), lower adjacent channel sound carrier (40.4 MHz) and to attenuate the sound IF signals of the channel being received to make the same suitable for inter-carrier sound system. AGC bias is applied to the base of Q_1 through L_5 and R_2 with capacitors C_8 are C_5 providing the decoupling for the AGC line. A single tuned circuit T_1 acts as load in the collector of Q_1 (BF196/167) through the coupling capacitor C_{11}.

The second stage of IF amplification is provided by the two transistors Q_2 (BF196/167) and Q_3 (BF196/167) connected in a cascode configuration. This arrangement is a combination of common emitter (CE) and common base (CB) stages and has the advantages of both configurations ensuring higher gain with good stability. The use of RF chokes in the collector supply leads provides more effective decoupling at higher frequencies. IF transformer T_2 in the collector of Q_3 selects the amplified video IF signal and applies the same to the base of Q_4 (BF197/173) through the coupling capacitor C_{20}. Q_4 provides the third and the last stage of IF amplification after which the IF signal appearing in the secondary of IF transformer T_3 is applied to the detector diode D_1 (OA79). It is at this detector stage that the picture IF (38.9 MHz) and the sound

IF (33.4 MHz) recover by a process of detection, the inter-carrier sound frequency of 5.5 MHz which is trapped by the transformer T_4 and passed on to the secondary of this transformer for amplification by the IF stages of the sound section of the primary of T_4 also acts as a filter for the 5.5 MHz FM sound signal which is not allowed to pass on to the video amplifier stages.

The video IF signal from T_3 is also detected by the diode D_1 and appears across the diode load resistor R_{18} from where it is coupled on to the video amplifier stage.

The video signals from the video amplifier are given to the cathode of the picture tube through a 0.1 μF capacitor which has a diode OA 79 across it to provide DC restoration to maintain average illumination on the screen. This is called cathode modulation. The grid voltage is maintained about 25 to 50 V negative with respect to the cathode for proper brightness on the screen. This is done by adjusting the HT voltage on G through the potentiometer R_2 which is the brightness control. Frame blanking pulses are also applied to the grid via a 0.05 μF capacitor. The voltages for the A_1 and A_2 electrodes are produced by the boost HT (BHT) and line blanking pulses are applied to A_1 through a 3.3 kpF capacitor. EHT of about 18 kV is applied to a connection provided for the purposes on the body of the tube. Sweep voltages for the deflection coils are obtained from the respective sweep output stages. (See Fig. 19.11(a)).

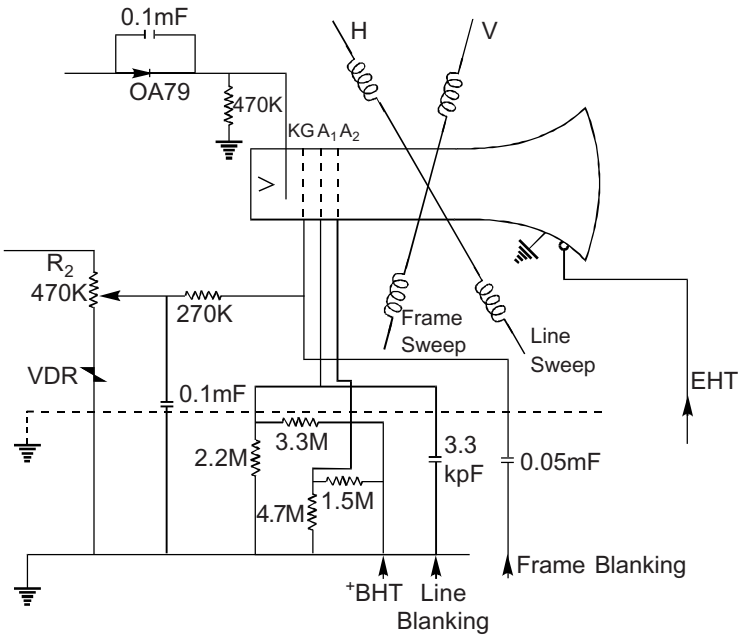


Fig. 19.11(a) Circuit of a Picture Tube Stage

19.7.3 Automatic Gain Control (AGC)

AGC in a TV receiver serves the same purpose as the AGC or AVC in AM radio receivers. AGC in a TV receiver controls the gain of the RF and IF stages

by means of a negative bias voltage which is proportional to the strength of the received signals so that variations in the strength of the signals at the antenna due to various causes do not affect the picture quality which remains more or less constant. With the AGC circuit working properly, the contrast control need not be reset everytime a new channel is tuned in. However, in TV receivers a special type of AGC called *keyed* AGC is used

19.7.4 Transistorised Keyed AGC System

In transistorised keyed AGC system, a transistor functions as the pulsed or gated-in stage. The transistor conducts when the sync signals from the video stage are received at the base at the same time as the flyback pulses from the horizontal output transformer are received at the collector. The conduction current will charge a capacitor which will provide the required AGC bias for the RF and IF stages of the TV receiver.

Figure 19.12 shows the circuit of a typical transistorised keyed AGC system used in TV receiver.

Transistor Q_1 is the keyed transistor and transistor Q_2 is the AGC amplifier. Both transistors are of the same type BC147B(NPN). Q_1 is normally not conducting due to cut off emitter bias applied through the adjustable potential divider formed by R_{10}, R_9 and the potentiometer R_8 (1 K) which also functions as the AGC control. The transistor conducts when positive going flyback pulses from the line output transformer (LOT) are applied to its collector. The collector current that flows, charges the capacitor C_3 (2.5 μF) negatively, while C_3 already has a small positive voltage provided by the potential divider formed by the resistors R_6 and R_7. Although the voltage developed across C_3 can provide the AGC bias, this voltage is amplified and also inverted in polarity to obtain AGC bias required for the forward AGC biasing of the commonly used NPN transistors in the RF and video IF stages of the TV receivers.

Transistor Q_2 is normally conducting in the absence of the video signal because of the forward bias provided by the voltage on C_3. As the video signal amplitude increases, the negative voltage across the capacitor increases thereby reducing the forward bias on Q_2. The resulting decrease in collector current raises the collector voltage which is applied to the RF (tuner) stages and the video IF stages after being suitably filtered by the RF filters provided for the purpose.

The diode D_1 (OA81) is connected in the collector circuit of Q_1 to prevent the capacitor C_3 (2.5 μF) from discharging through the transistor Q_1 by providing forward bias to its collector junction.

The AGC given to the RF amplifier is a *delayed* AGC so that it acts only when the received signals are above a minimum strength and AGC will not act on weak signals. This delay is provided by a diode (OA81) which is connected across the AGC line and is given a positive voltage to its anode through a high resistance. (Not shown). The diode conducts and shorts the AGC line till the negative AGC voltage developed exceeds the positive bias on the diode and the diode gets reverse biased and stops conducting. The non-conducting diode with its high resistance no longer acts as a shunt across the AGC line which is applied to the RF amplifier through a filter network.

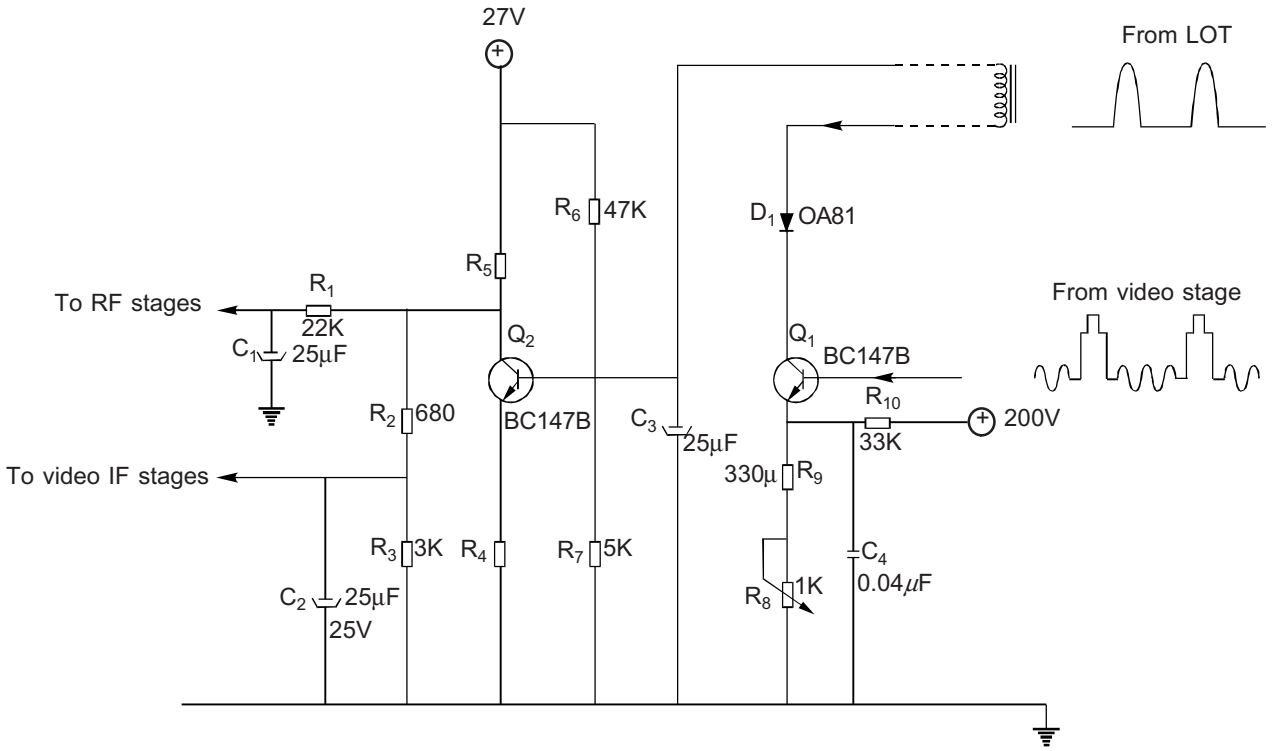


Fig. 19.12 Transistorised Keyed AGC Circuit

19.8 SOUND SECTION

The sound section of a TV receiver consists of the sound IF amplifier, the FM detector and audio amplifier.

The inter-carrier sound IF frequency of 5.5 MHz obtained as a result of the beating of the picture IF (38.9 MHz) and the sound IF (33.4 MHz) is picked up from the video detector stage and fed as input to the sound IF amplifier stage. The IF amplifier generally consists of one or two stages of wideband amplification allowing a bandwidth of ±50 kHz at 5.5 MHz. One of these stages also acts as the *limiter* to limit the amplitude of output signals beyond a certain limit to avoid amplitude modulation of the carrier by noise voltages and thereby improving sound output quality.

The sound carrier is frequency modulated. In frequency modulation the centre frequency, 5.5 MHz in this case, swings above and below the centre frequency by amounts proportional to the amplitude of the modulating frequencies. In FM detection or demodulation, these frequency swings have to be converted into corresponding amplitude swings which can be amplified by normal audio amplifiers to produce the required sound. FM detection is a little more complicated than AM detection and therefore needs special circuits.

A large number of circuits are available for FM detection and these have already been described under Section 14.11.

19.9 AUDIO AMPLIFIER

The detector output is fed into a two-stage audio amplifier consisting of a voltage amplifier and a power output stage.

In earlier TV receivers vacuum tubes or transistors were used as audio amplifiers, but these have now been replaced by ICs which beside serving as audio pre-amplifiers also serve as sound IF amplifiers, limiters and sound FM detectors.

ICs most commonly used in the sound section of modern TV receivers are CA810, TBA 120s and CA3065.

CA810 This is generally used as an output amplifier in the sound section of a TV receiver. Full details of the use of CA810 as an audio output stage have been given in Chapter 7.2.

IC-TBA 120s This is a linear IC chip, specially designed for use in the sound section of a monochrome TV receiver. Besides performing the functions of sound IF amplifier, and sound FM detector. It also acts as the pre-amplifier for the audio amplifier stage.

IC-CA 3065 This is an IC manufactured by BEL and is capable of performing the functions of a multistage IF amplifier, a limiter, FM detector and AF pre-amplifier. Beside, it has a built-in regulated power supply system and an electronic attenuator to act as the volume control. Figure 19.13 indicates two ways of connecting IC BEL CA 3065 in the audio stage in a hybrid TV receiver.

19.10 SWEEP SECTION AND SYNCHRONISATION

The process of interlaced scanning requires that the electron beam in the picture tube should be swept across the screen from left to right by means of an

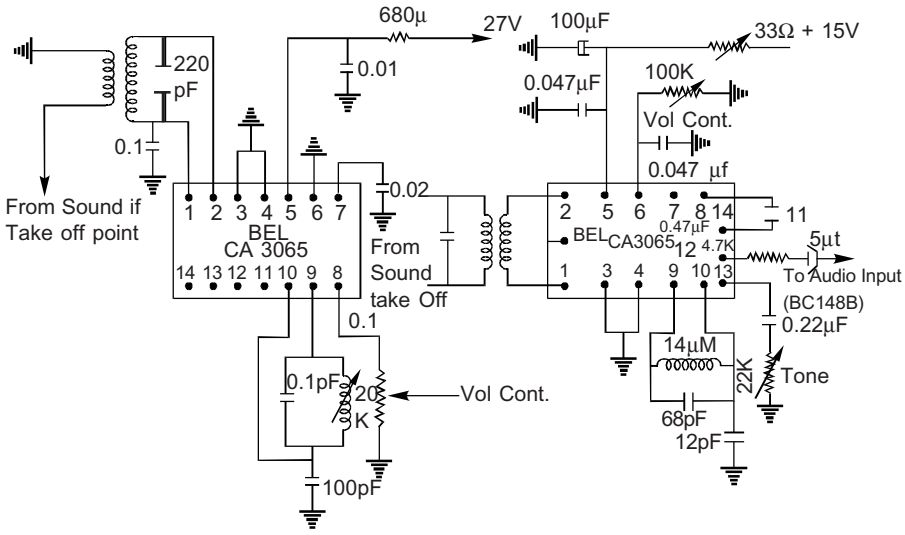


Fig. 19.13 Two Ways of Connecting IC BEL CA3065

oscillator called the line or horizontal oscillator. Similarly, the electron beam has to be deflected down at a slower rate and then brought back to its original position by another oscillator called the frame or vertical oscillator. Moreover, for the formation of a steady and correct picture, it is necessary that the two processes of horizontal and vertical deflection at the receiver should be in complete step or in synchronisation with similar scanning processes at the transmitter. This synchronisation is achieved with the help of horizontal and vertical sync pulses which are sent from the transmitter as a part of the composite video signal. At the receiver the sync pulses are clipped and separated from the video signal by a circuit called the *sync separator*. The horizontal sync pulses are then used to control the horizontal oscillator frequency with the help of an automatic frequency control (AFC) system to keep it locked-into the horizontal scanning at the transmitter. The vertical sync pulses are used to trigger the vertical oscillator to keep the picture frames locked-in vertically. These processes will now be described in detail with the help of the circuits involved.

19.10.1 Sync Separator

The composite video signal containing the sync pulses riding on top of the blanking pulses is applied to the input of the transistor which is biased beyond cut off. The tips of the positive sync pulses raise the bias of the transistor above the cut off and the collector current flows only for the duration of the sync pulses as shown in Fig. 19.14. The horizontal, vertical and equalising pulses are all clipped from the peak amplitudes of the video signal. The capacitor C_1 and resistor R_1 produce cut off bias for the transistor. During the peaks of the input video signal the bias current charges cap-C_1 and the negative voltage developed on C_1 causes the transistor to conduct. Sync output is taken at the collector as shown in Fig. 19.14.

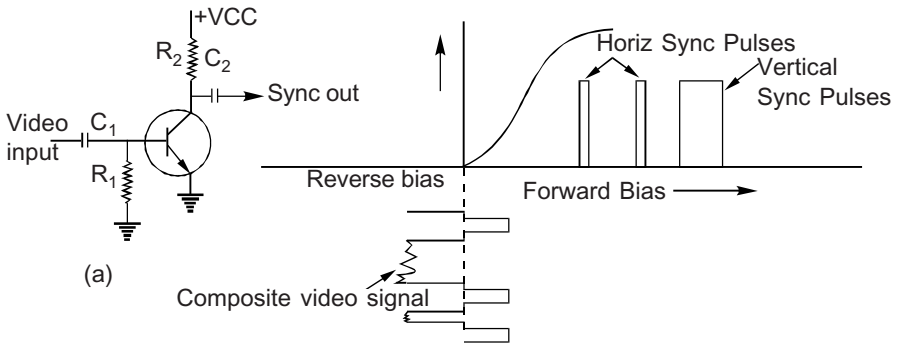


Fig. 19.14 Sync Separator

19.10.2 Separation of Line and Field Sync Pulses

The output of a sync separator contains the horizontal and vertical sync pulses, along with equalising pulses. The horizontal and vertical sync pulses must be separated from each other to enable them to control their respective oscillators. Although these pulses have the same amplitude, their width or duration is different. Whereas the line sync pulses have a duration of about 5 μs only, the duration of the field sync pulses is about 160 μs. This fact enables the line and field sync pulses to be separated from each other by the use of special circuits known as differentiating and integrating circuits.

A differentiating circuit consists of a capacitor C_1 and a resistor R_1 connected as in Fig. 19.15. The time constant of this circuit (R_1C_1) is very small compared to the duration of each pulse, so that it charges and discharges

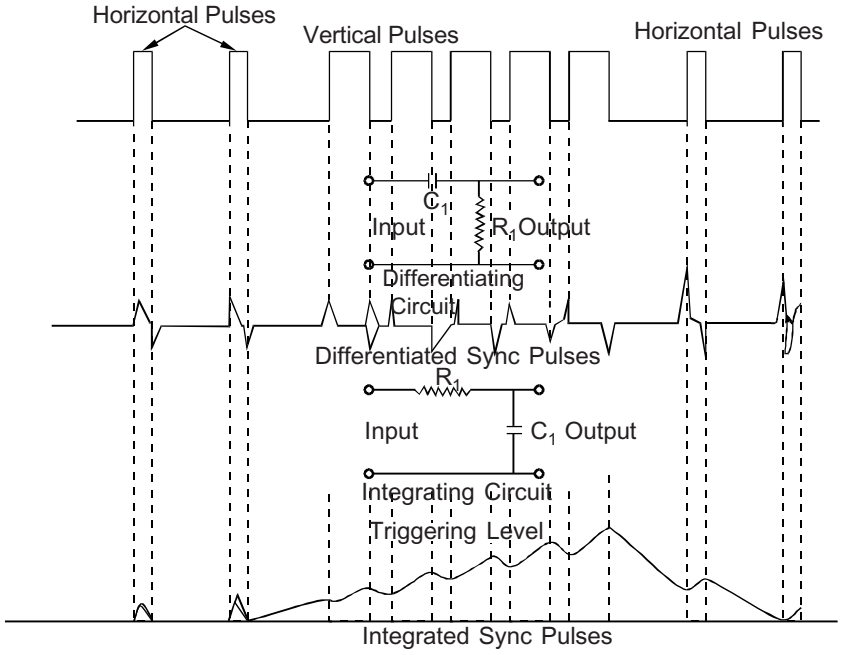


Fig. 19.15 Output from Differentiating and Integrating Circuits

quickly whenever the input to the circuit changes, producing very sharp spike-like pulses corresponding to each of the sync pulses, whether they are line, field or equalizing pulses. The circuit does not, however, separate the line and field sync pulses but any of the voltage spikes produced by the line sync pulses and alternate field and equalising pulses can be utilised to synchronise the line oscillator. This is possible because the field sync pulses (serration) and equalising pulses have double the frequency of the line sync pulses and, as such, the leading edge of these alternate pulses will be at the exact time required for the line sweep synchronisation. The circuit is basically a high pass filter.

For the separation of line sync pulses from field sync pulses, an integrating circuit consisting of C_1 and R_1 connected as in Fig. 19.15 is used. The time constant (R_1C_1) of this circuit is kept large compared to the duration of the field sync pulse. Hence the capacitor gets charged to a high voltage which does not discharge to any appreciable extent before the arrival of the next field sync pulse because the interval between successive pulses is small compared to their width. This voltage builds up with the arrival of each group of vertical sync pulses and the return period of the vertical sweep oscillator is then triggered into action when the amplitude of the integrator output pulse reaches some particular level as shown in Fig. 19.15. Equalising pulses help to reach this triggering level at a time in the vertical synchronising period that is the same for successive fields, without which proper interlacing is not possible. Line sync pulses are not able to build up any appreciable voltages on this circuit due to their small duration. This circuit acts like a lowpass filter.

A typical circuit arrangement for a sync separator used in solid-state TV receiver is given in Fig. 19.16. Transistor circuits require a much smaller input voltage than valve circuits. An input signal of the order of 5 to 10 volts can give rise to output sync pulses of about 50 volts peak-to-peak. Noise suppression circuits are necessary to avoid the synchronisation, being upset due to noise pulses. In some receivers ICs are used to perform the function of a sync separator. A single module can combine the functions of a noise invertor, sync separator AGC.

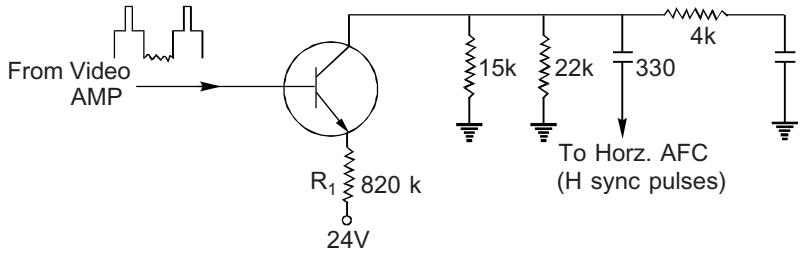


Fig. 19.16 A typical Sync Separator stage

19.11 SWEEP CIRCUITS AND THEIR SYNCHRONISATION

For the formation of a raster on the picture tube, the electron beam has to be deflected horizontally as well as vertically. The deflection of the cathode ray beam is obtained by passing sawtooth current waves of the correct frequency through a set of deflector coils placed round the neck of the picture tube. This

type of deflection called magnetic deflection enables the electron beam to be deflected over a wide angle with a minimum of deflection defoccusing. Linear sawtooth current waves for horizontal as well as vertical deflections are obtained from sawtooth oscillators which may be block oscillators, multivibrators or sine wave oscillators. Circuits for these oscillators and their properties have already been described in Chapter 10. Coupling transformers in the output stages of the amplifiers driven by sawtooth oscillators supply linear sawtooth wave currents to the respective deflection coils under proper matching conditions. Sawtooth wave currents are supplied to the horizontal and vertical deflection coils at the rate of 15625 and 50 Hz respectively. For satisfactory reproduction of the picture, these frequencies should not only remain constant within specified limits but should also remain in step with the transmitter. Sync pulses sent out by the transmitter with the video signals and separated by the sync separator are used as references to control the frequency of the respective oscillators to keep these frequencies always locked to the transmitter sync signals. Any deviations from the reference sync frequencies are corrected automatically.

19.11.1 Vertical Deflection Stage

The vertical deflection system is a power amplifier which supplies sawtooth currents to the vertical deflection coil at field frequency of 50 Hz. The stage generally consists of an oscillator and an output stage with an output transformer (VOT) to match the low impedance of the deflection coil to the high impedance of the output stage. The entire stage operates as a multivibrator which is synchronised by sync pulses obtained from the integrating circuit of the sync separator. The use of feedback circuits and other wave shaping techniques is necessary to obtain linear deflection from the deflection coils.

The type of circuit employed for the vertical deflection stage depends on the type of the receiver, i.e. whether the receiver is a hybrid type or completely solid-state receiver. Most modern solid state receivers use either a completely transisto-rised vertical deflection stage or make use of a suitable IC like the IC TDA 1044.

19.11.2 Transistorised Vertical Deflection Stage

Figure 19.17 shows the vertical deflection stage using transistors only. The vertical sync pulses from the output of the sync separator are applied as input to the base of the transistor T_1 (BC 147 B), which works as an oscillator and is synchronised by the 50 Hz vertical sync pulses from the sync separator. Feedback voltage from the collector of Transistor T_2 is applied to the base of transistor T_1. The transistor T_3 (SK 100) pair the output amplifier stage consisting of a complementary symmetry pair comprising transistors T 2N5294 (npn) and transistor 2N6107 (pnp) the output impedance provided by the complementary pair is quite low and does not need any output transformer. This transformerless output stage provides higher efficiency and improved performance. The output deflection signal for the vertical deflection coils is taken from the junction of the emitters of T_4 and T_5 the 200 μF capacitor blocks any DC voltage to the deflection coil to avoid shifting of centre portion. T_4 and T_5

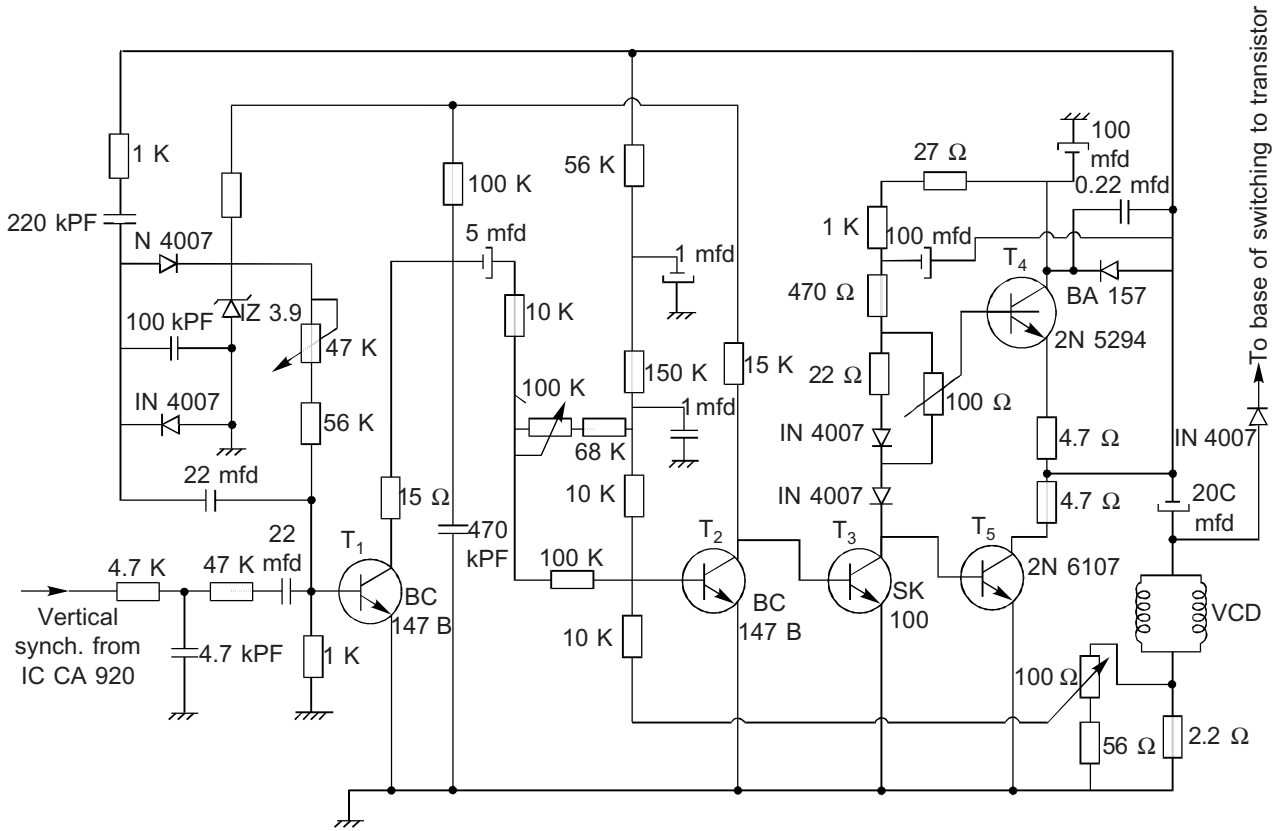


Fig. 19.17 Vertical Stage using Transistors.

operate as Class B push-pull amplifier, each transistor supplying alternately one half of the AC sawtooth current for the deflection coil. The 100 Ω variable resistor serves as the height control.

19.12 HORIZONTAL DEFLECTION STAGE

Horizontal deflection system is one of the most critical stages in a TV circuit and plays an important part in determining the quality of the picture. It performs the following functions in the working of a TV receiver:

(a) It supplies the sawtooth wave currents at the line frequency of 15625 Hz to the horizontal deflection coils fitted in the neck of the picture tube.
(b) It helps in keeping the line frequency correctly synchronised to the frequency of the line sync pulses from the transmitter by means of a built-in automatic frequency control (AFC) system.
(c) It produces the extra high tension (EHT) voltage for the final anode of the picture tube and also the boosted high tension (BHT) voltage required for the first anode (A_1) of the picture tube and the plates of the tubes in the vertical and horizontal deflection stages.
(d) It provides pulses for the functioning of the keyed AGC system.
(e) It also supplies pulses for its own AFC system for synchronising the horizontal oscillator with the horizontal sync signals.
(f) Negative flyback pulses, as horizontal blanking pulses, are also supplied from the output of the stage for suppression of the horizontal retrace. A complete horizontal deflection system actually consists of two stages—the horizontal oscillator and the horizontal output stage. These two stages are described below separately for clarity and proper understanding of the complicated operation of the horizontal sweep system.

19.12.1 Horizontal Oscillator

The horizontal oscillator generates sawtooth waves of line frequency (15625 Hz) which is kept synchronised to the line sync signals by means of the AFC system.
AFC System.

The AFC circuit consists of two diodes D_1 and D_2 connected in series so that when sync pulses of equal amplitude but opposite potentials are supplied to the two diodes, the output currents balance each other and no voltage appears at the centre tap of the AFC control. When sawtooth pulses from the line output transformer are applied at the junction of two diodes the balance will not be disturbed if the sawtooth waves are in phase with the sync pulses. If, however, there is a

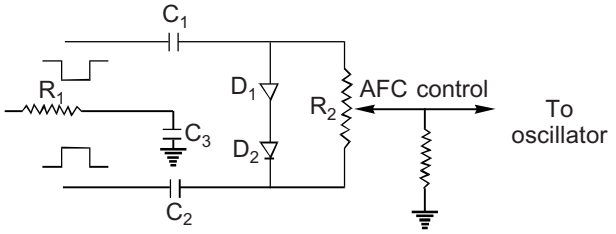


Fig. 19.18 AFC System

phase difference between the sawtooth waves and the sync pulses, one of the diodes will conduct more than the other and a DC voltage will appear at the AFC control which when applied to the oscillator through a filter network will pull the frequency of the oscillator back into synchronisation with the sync pulses.

The output from the horizontal oscillator is applied to the line output stage.

19.12.2 Horizontal Output Stage

Basic circuit of the horizontal output stage of a solid stage TV is shown in Fig. 19.19. The output transformer T_1 is lined to the horizontal frequency 15625 Hz. T_2 is the EHT transformer and D_1 is the damper diode. The diode D_2 is the EHT rectify for the rectifier of high voltage pulses. Silicon diodes are normally used for EHT rectification.

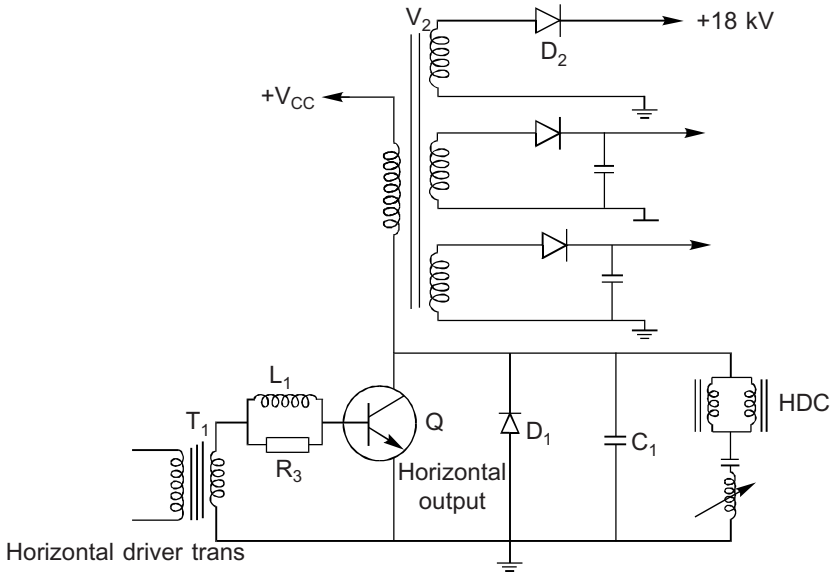


Fig. 19.19 Basic Circuit Horizontal Output

19.12.3 Horizontal Output Section

A complete circuit of the horizontal output section is shown in Fig. 19.20.

19.13 POWER SUPPLY SECTION

The power supply system in a TV receiver is not as simple as the power system of a radio receiver. This is mainly due to a large number of stages and the complicated nature of the circuitry employed in a TV receiver. Two types of power supplies are used in a TV receiver. One is the low-voltage, high-current power supply and the other is a high-voltage, low-current power supply. Whereas the low-voltage, high current power supply is used for the operation of vacuum tubes employed in various amplifiers, oscillators and similar other circuits including the heater circuits of these tubes, the high-voltage low-current power supply is exclusively meant for the operation of the picture tube.

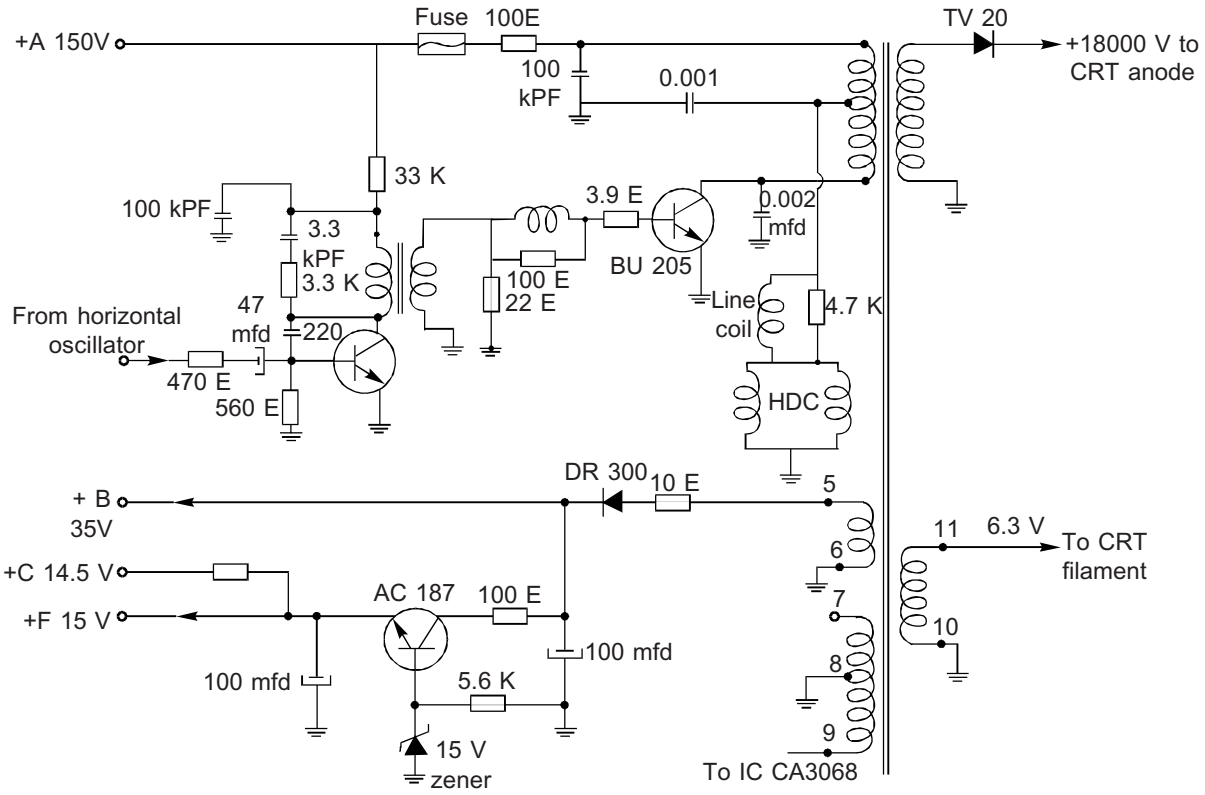


Fig. 19.20 Circuit Diagram of Horizontal Output Section

Low-Voltage Power Supply

The low-voltage power supply system in a TV receiver is similar to the power supply system of a radio receiver. 230 V 50 Hz mains supply is stepped up or stepped down by means of a power transformer and the AC voltage so obtained is converted into DC with a rectifier. A filter circuit is used to filter out the pulsating variations from the DC output voltage of the rectifier. One or two step-down windings provided on the power transformer supply 6.3 V AC for the heaters of the various tubes and the picture tube.

Due to the simplicity of their construction and easy operation, silicon diodes are generally used as rectifiers even in tube type TV receivers. The DC obtained after rectification is divided into a number of branches for use in separate sections and stages of the receiver. Each branch has its own resistors and capacitors for filtering and dropping the voltage to the particular value required for the stage concerned.

In the case of hybrid TV receivers which make use of both tubes and transistors in their circuits, two branches for low DC voltages (15 V and 27 V) have to be produced for the operation of transistors and ICs, etc. Zener diodes of suitable ratings are required for stablising the DC voltage for satisfactory operation of transistors.

High-Voltage Supply (EHT)

In the high-voltage supply system of a modern TV receiver, very high voltage of the order of 18 to 20 kV (25 kV for colour receivers) are produced, but the current requirements are extremely low ranging about 300 μA. This high voltage which is commonly called EHT is induced in the horizontal output transformer (HOT) when the magnetic field produced in the horizontal deflection coil collapses during the flyback period of the horizontal deflection pulses. The high voltage developed in the primary of the HOT is stepped up by an EHT secondary winding, rectified by a diode (DY802), filtered and applied to the picture tube through a connection provided on the outer wall of the picture tube. The problem of a suitable filter capacitor of such high-voltage rating (18 to 25kV) is solved by the capacitance formed by the aquadag coating on the inside of the picture tube and the outer graphite coating which is earthed. The glass wall acts as the dielectric. A capacitance of above 1500-2000 pF is automatically connected across the EHT and no additional filtering capacitors are required.

As the voltage developed in this section of the receiver is extremely high, extra care and caution is called for while testing and checking the operation of the section of the TV receiver. It is to avoid any risk of sparking and accidental contact that all components for the EHT circuit are mounted in a separate and properly screened compartment on the receiver chassis.

The HOT or LOT which develops such high voltages and operates at very high frequencies is wound on a special ferrite core material to keep the eddy current and other core losses to the minimum. The primary, secondary and auxiliary windings are wound on one support, impregnated with special resins and placed on one limb of the core. The EHT winding, also well-insulated and

encapsulated is placed on the other limb of the core. All the winding terminals are brought out to a terminal plate and the EHT terminal is connected to a well-insulated tube cap.

Another voltage that is produced in the horizontal output stage as a result of the flyback action is the boosted B+ voltage or the BHT. The damper diode (DY88) which conducts immediately after the flyback is a half-wave rectifier for the AC deflection voltage produced during the trace time in the horizontal output circuit. The DC output voltage (600-700 V) produced by the damper diode is in series with the normal HT voltage (250-300 V) on the plate of the damper and the two voltages add together to produce a DC voltage of about 1000 V known as the BHT.

The damper diode is also known as the *efficiency diode* because it provides horizontal scanning for the left side of the raster when the horizontal output tube is not conducting. Use of regulated power supply is essential for modern TV receiver using transistors and ICs, particularly the colour TV receiver. The details of regulated power supplies including SMPT have already been given in Chapter 11.

19.14 SOLID STATE TV RECEIVER

Bharat Electronics Limited (BEL) will be progressively discontinuing the manufacture of electronic valves now used in hybrid receivers of Indian manufacture. As such, the manufacture of TV receivers using solid state devices like transistors and ICs, will progressively increase. Moreover, solid state TVs using specially designed transistors and ICs possess many advantages such as rugged and compact construction, easy serviceability, high reliability, perfect stability and high quality audio output with low power consumption. A block diagram of such a receiver based on a BEL design is given in Fig. 19.21.

19.15 MAINTENANCE AND SERVICING OF TV RECEIVERS

The maintenance and servicing of a TV receiver requires not only a clear understanding of its complex circuitry but also the use of certain special types of measuring and test equipments. The chief indicators of the satisfactory functioning or otherwise of a TV receiver are the quality of its raster, picture and sound. Any abnormalities in the picture or sound or the complete absence of one or both are definite pointers towards the failure or malfunctioning of one or more stages in the receiver. This may be the result of the failure of any components like the resistors, capacitors, tubes, transistors and other semiconductor devices or due to certain adjustments and alignments getting disturbed. A systematic analysis of the indications provided by the screen and the speaker together with certain tests and measurements help in localising the faults and isolating the defective components for necessary replacement or repairs. Methods for conducting such tests and measurements and the use of certain special types of equipments required for the purpose will be described in this chapter.

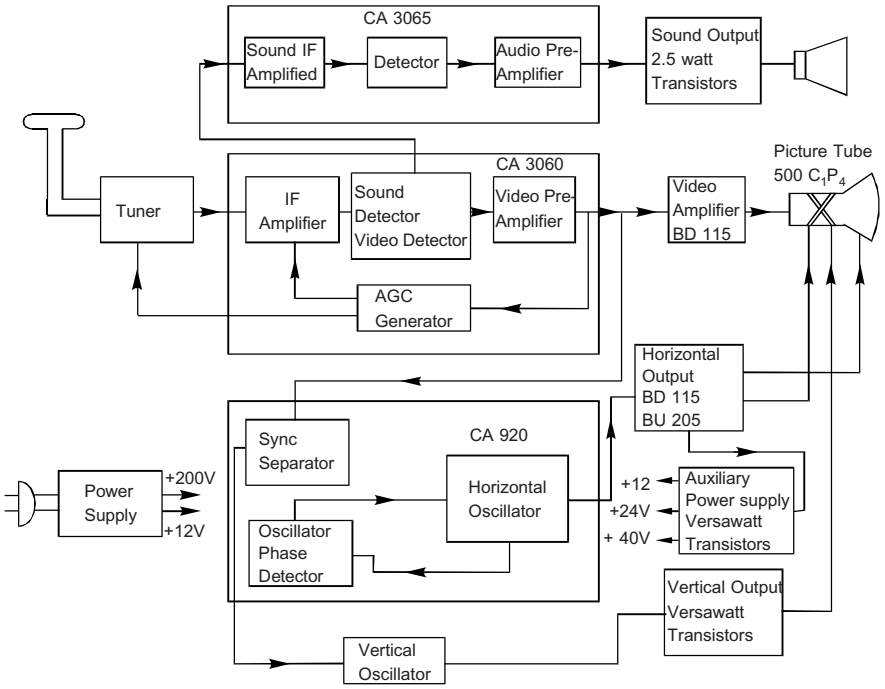


Fig. 19.21 Solid State TV Receiver—Block Diagram of a Solid State TV Receiver (*Courtesy*: BEL)

Troubleshooting in TV Receiver

The first step in the process of fault finding in TV receivers is to localise the fault to a particular stage or stages with the help of indications or symptoms provided by the raster, picture and sound quality of the receiver. The second step is to pinpoint the trouble to a defective component like a tube, capacitor, resistor, coil or a semiconductor device. The simplest check for defective tubes and other plug-in components is to replace them by new ones. Voltage and resistance measurement with a multimeter or VTVM comparing these values with the normal values indicated on the manufacturers circuit diagrams help in locating their defective components. The last step is to replace or repair the defective component and to carry out such adjustments and alignments as will restore the receiver to normal working condition. While carrying out the DC voltage tests with a multimeter, it is always useful to remember that:

(a) A DC voltage higher than normal indicates an open circuit which may be due to an open resistor or a coil. It can also be due to a transistor or a tube not functioning properly.

(b) A voltage lower than normal indicates a short circuit which may be due to a shorted capacitor or a tube transistor drawing abnormal current.

(c) Resistors do not short but they get burnt out or develop high resistance with age. Burning out of resistance is often due to shorting of bypass capacitors, which should be carefully checked up. For suspected open capacitors, a good capacitor should be tried in parallel with the suspected open capacitor.

Main indicators of faults in a TV receiver are the picture tube screen and the loudspeaker. Faults in the raster and picture are indicated by the picture tube screen and the faults in sound are indicated through the loudspeaker. A knowledge of the faults in the raster, picture and sound of a TV receiver together with likely defective stages responsible for these faults is necessary for quick and effective repairing. It is also important to remember that a raster exists even when there is no picture or video signal being applied to the receiver. A good raster is necessary for the formation of a good picture on the picture tube screen.

Faults in Raster, Picture and Sound

Some common troubles in raster, picture and sound of a TV receiver are given below. Likely defective stages or sections which can cause these troubles are also indicated to help locate these faults and rectify them.

Raster Troubles

Common raster troubles are: no brightness, insufficient height, insufficient width, horizontal line, vertical line, foldover, trapezoidal, non-linear, tilted or off-centre raster, no raster. Raster troubles are mainly due to scanning stages. When the sound is normal, no brightness on the screen should result from the EHT not being applied to picture tube. This could be due to horizontal oscillator, horizontal output stage or the damper diode developing no booster HT. No brightness can result even with normal EHT, if the voltage on the screen grid and other electrodes of the picture tube are not normal. Insufficient height is the result of a defective or weak vertical stage whereas loss of width is generally associated with the horizontal output stage and damper diode. The horizontal line on the screen is due to a defective vertical stage and only a vertical line is due to a defective horizontal output stage or deflection coil. A trapezoidal raster, tilted raster or off-centre raster can result from a defective deflection yoke or yoke coil. It should be remembered that the picture is built on the raster itself. A good raster is necessary for a good picture.

Picture Troubles

When the raster is satisfactory, the picture can still develop faults, some of which are given below.

1. No picture, snow in picture, vertical and horizontal rolling, drift in picture, excessive contrast, blooming, hum bars and sound bars.
2. Loss of sync., torn picture, jittery picture, negative picture, picture bending vertically.
3. Ghosts, venetian blind effects, streaks, black and white dots and dashes in the picture.

In localising picture faults, it is useful to remember that the picture (video) and sound signals pass through a number of common stages (RF, IF and detector) before the two are separated. If both the picture and sound are absent the fault could be attributed to the RF, IF, detector or AGC stages, whereas if only the picture is absent with normal sound, the fault lies in the video amplifier or the picture stage.

Vertical rolling of the picture can be the result of a defective vertical oscillator or lack of sync. voltages for the vertical oscillator. Horizontal drift or instability in horizontal direction, torn or jittery picture results when the line oscillator frequency is not correct or the discriminator circuit is not functioning properly due to lack of sync. pulses or flyback pulses. When the picture is unsteady both in horizontal and vertical directions the fault generally lies in the sync. separator stage.

A weak picture with snow, an excessively contrasted picture lacking synchronisation or a negative picture can be the result of a defective AGC circuit, otherwise the RF and IF stages need alignment.

50 Hz mains pick-up or ripple from the rectified voltage entering the AFC circuits, horizontal oscillator or sync. separator circuits can make the picture bend vertically.

Horizontal and vertical bars are produced when a frequency less than 15625 Hz or more than 15625 Hz respectively is applied to the grid of the picture tube. Sound bars appear on the picture when the FM sound signal also produces some audio frequency due to slope detection in the AM detector stage. This audio combines with the video at the grid-cathode circuit to produce horizontal sound bars which vary with the amplitude of sound.

Blooming is the defocussing of the picture when brightness is increased. It results from a weak EHT rectifier tube (DY802). Ghosts are duplicate pictures produced by reflection of the signal from buildings and other objects which appear by the side of the main picture produced by direct reception by the TV antenna. Ghosts can be avoided by use of directional antennas.

Sound Troubles

Common troubles with sound in a TV receiver may be weak or distorted sound, sound with a hum or buz, or no sound at all.

Picture and sound in a TV receiver have a number of common stages and hence some of the sound troubles are inter-related to picture troubles, whereas other sound troubles are due to the sound section of the receiver. If all the three indications namely raster, picture and sound are absent, the trouble is due to the power supply circuit. The absence of sound alone with raster and picture normal is due entirely to the sound section which will include sound IF amplifier and limiter, sound discriminator, AF amplifier or loudspeaker. Sound IC if used in the circuit should also be checked up. If both sound and picture are defective or are absent with normal raster, the trouble is due to either RF (tuner), IF, video detector or AGC stages. A weak or distorted sound (picture normal) can be due to an improper discriminator adjustment, defective output stage or a defective loudspeaker.

Troubleshooting Charts

There are no thumb rules or cut and dry methods for the repairing of TV receivers except a thorough knowledge of the function of the various stages and a careful observation of the indications provided by the picture tube screen and the loudspeaker. However, certain troubleshooting charts or tables based on major fault indications provided by the raster, picture or sound together with

suspected stages and the likely defects sometimes prove helpful in quick location of the faulty stage or component. One such chart is given in Table 19.1.

Table 19.1 Troubleshooting Chart

Trouble symptoms	*Suspected stage (section)*	*Likely defect*
1. No raster, no picture (no light) and no sound	Power supply	(i) Mains voltage not being applied either due to a blown off fuse or some defective lead. (ii) Heater circuit open. (iii) Defective DC rectifier circuit.
2. No raster, no picture (no light), but sound O.K.	EHT circuit, line output stage, picture tube, its bias and socket connections, video output circuit	Defective EHT transformer (LOT) or line output valve or transistor, EHT rectifier or booster condenser, line oscillator defective picture tube or improper voltages on its pins, defective brightness control, picture tube socket, video amplifier tube or transistor.
3. Raster and picture normal but no sound	Sound section	Sound IF, FM detector, AF stage, loudspeaker, sound IC.
4. Raster and sound normal but no picture (or weak picture)	Video amplifier	Video amplifier tube or transistor, contrast control or coupling capacitor between video amplifier plate and picture tube cathode
5. Raster normal but no picture and no sound	Antenna, feeder line, tuner, video, IF amplifier, video detector, AGC	Broken antenna or feeder line, defective tuner stage (valve or transistor), defective video IF amplifier stage (valve transistor or IFT) defective detector diode, AGC adjustment.
6. No raster, no picture but only a bright horizontal line on the screen, sound normal	Vertical sweep, vertical deflection coil	Defective vertical stage, open vertical deflection coil.
7. No raster, no picture but only a bright vertical line on screen (less picture width), sound normal	Horizontal deflection coil	Line deflection coils or circuit components between LOT and horizontal deflection coils, weak line output stage.

(*Contd.*)

(*Contd.*)

Trouble symptoms	*Suspected stage (section)*	*Likely defect*
8. Raster and sound normal but picture height is less even with maximum position of height control	Vertical oscillator, vertical output or vertical deflection coil	Defective oscillator valve or transistor, reduced voltage to vertical oscillator or output stage, defective vertical output transformer, defective vertical deflection coil, defective VDRs.
9. Raster and sound normal but picture width is less	Horizontal deflection stage, horizontal deflection coils	Line oscillator, line output amp., defective booster diode, reduced HT voltage, defective horizontal deflection coil.
10. Picture rolling vertically, raster and sound normal	Vertical oscillator	Vertical oscillator stage, defective vertical hold, vertical sync separator.
11. Picture torn, diagonal bars and slanting streaks	Horizontal oscillator, Horizontal sync., AFC	Horizontal oscillator stage, oscillator coil adjustment, coil core defective, discriminator circuit, power supply filter circuit.
12. Picture unstable both in horizontal and vertical directions	Sync. separator, AGC, signal section	Sync. separator tube or IC, defective AGC stage, tuner or video amplifier gain adjustment,
13. (a) Vertical non-linearity	(a) Vertical oscillator or vertical output	(a) Vertical output valve or transistor, Vertical linearity control.
(b) Horizontal non-linearity	(b) Horizontal deflection drive	(b) Line oscillator or line output stage, booster diode or booster capacitor, horizontal linearity coil.
14. Blooming in picture	EHT rectifier stage	Defective EHT rectifier.
15. Ghosts	Antenna	Antenna direction not correct, director or reflector missing.

19.16 TV RECEIVER CONTROLS

In a TV receiving set, controls are necessary for proper reproduction of both sound and picture compared to a radio receiving set where only the reproduction of sound is to be controlled. A TV receiver has, therefore, many more controls than a radio receiver. There are two types of controls in a TV receiver—the operator's controls and the service controls for service adjustments. The operator's controls have to be operated frequently for proper

reproduction of picture and sound and are located in an easily accessible position, generally the front panel of the TV receiver. The service controls are preset controls which have to be adjusted once in a while. To avoid accidental handling, these controls are located at relatively inaccessible positions, generally the rear or side of the receive cabinet. The operator controls are operated by knobs to make the operation smooth and easy, but the service controls are generally adjusted by a screw driver or some other tool designed for the purpose. The operator controls are operated by the viewers to obtain the picture and sound quality to suit their tastes but the service controls should be adjusted only by a trained service technician who really understands the purpose and functioning of these controls. Actual location, the type and design of each control and the number of controls used vary from manufacturer to manufacturer. The terms used to identify the controls are also different with different manufacturers. A list to typical controls used in TV receivers of Indian manufacture is given in Table 19.2.

Table 19.2

Typical Operators Controls	*Typical Service Controls*
1. On-off switch	1. Vertical linearity control
2. Volume control	2. Horizontal linearity control
3. Tone control	3. Height control
4. Channel selector	4. Width control
5. Brightness control	5. AGC control
6. Contrast control	6. Picture position and centering controls
7. Fine tuning control	
8. Vertical hold control	
9. Horizontal hold control	

Controls Nos. 8 and 9 in the list of operators controls are treated as service controls by some manufacturers.

Since the quality of reproduction of picture as well as sound in a TV receiver depends to a great extent on the suitable manipulation of the controls even with a normal TV receiver, it is useful to understand the functions of these controls which are described below.

19.16.1 Operator's Controls

On-off Switch

This switch is meant to connect or disconnect the power supply mains to the receiver. It is located at a convenient position on the front panel and may be in the form of a toggle switch, a push-button switch, a piano-key type or a car-ignition type switch operated with a key. In most modern receivers, this switch is a part of the volume control.

Volume Control

The level of sound output from the speaker can be controlled by the volume control which generally controls the audio voltage applied to the input of the preamplifier stage of the audio amplifier.

Tone Control

This control is similar to the control in a radio receiver and it controls the proportion of high and low frequencies in the audio output. Separate bass and treble controls are used in certain cases.

A dual control with concentric knobs is sometimes used to combine the volume and tone control.

Channel Selector

This control is used in multichannel TV receivers. Its function is to select the coils and other components for the desired channel and connect these to the circuit in a proper manner.

Brightness Control

This control adjusts the illumination or brilliance on the screen by varying the DC bias of the grid-cathode circuit of the picture tube. The brightness control and the contrast controls are adjusted together to get a well-defined clear picture on the screen.

Contrast Control

This control is located in the video amplifier circuit and controls the amplitude of the video signal applied to the picture tube and works like the volume control for the audio signal. This control adjusts the sharpness of the picture on the screen and has to be operated in conjunction with the brightness control to get a proper contrast of white and black portions of the picture.

Dual controls with concentric knobs are also used in certain receivers for brightness and contrast.

Fine Tuning Control

This control varies slightly the frequency of the local oscillator to produce the correct IF in the frequency changer. It is in the form of either a variable capacitor, a variable inductance or a potentiometer that adjusts the voltage across a Varactor diode. This control is operated, after selection of the desired channel, till a sharp and crisp picture with clear and undistorted sound is obtained.

Vertical Hold Control

This control adjusts the frequency of the vertical oscillator to bring it close enough to 50 Hz so that it synchronises with the sync. signals from the transmitter. If the picture rolls up or down, the vertical hold control should be adjusted till the picture is steady.

Horizontal Hold Control

This control adjusts the frequency of the horizontal oscillator to bring in synchronisation with horizontal sync. signals. When the picture shifts horizontally or tears apart into diagonal segments, this control is adjusted to provide horizontal synchronisation till the picture is again complete and steady.

19.16.2 Service Controls

Vertical Linearity Control

Vertical linearity control is in the vertical output stage and adjusts the operating characteristics so as to make the scanning lines equally spaced from top to bottom for good linearity. If the picture is not uniform in the vertical direction it is the result of vertical non-linearity which can be corrected by the vertical linearity control. This control either varies the cathode bias to the vertical amplifier or it controls the amount of feedback in the vertical circuit. The vertical linearity control corrects the overall vertical non-linearity. In certain cases, separate linearity controls are provided for the top and bottom of the picture. Vertical linearity control affects the height of the picture also. It is, therefore, necessary that the vertical linearity control and the height control should be adjusted simultaneously to obtain proper picture height and linearity. Vertical linearity can be tested either with a test pattern from the TV transmitter or with the help of a pattern generator.

Horizontal Linearity Control

Lack of horizontal linearity results in a non-uniform picture in the horizontal direction. If there are people in the picture, they appear too broad at the left or too thin at the right. For correcting horizontal non-linearity, a linearity coil is provided in many TV receivers. The core in the linearity coil can be adjusted to provide a uniform picture. A pattern generator or a test pattern can be used to check up horizontal linearity.

Height Control

This control is located in the anode circuit of the vertical oscillator. It is a potentiometer that controls the voltage to the anode of the oscillator and the sweep voltage applied to the vertical deflection coil. This control is adjusted whenever the picture height is less than normal. Adjusting the vertical linearity control also affects the picture height which can be adjusted by the height control.

Width Control

This is a control in the horizontal output stage. It is a variable resistor in the screen-grid voltage for the horizontal output tube (or transistor).

AGC Control

The AGC control is for the adjustment of the AGC voltage applied to the RF and IF stages to control their gain. Different settings of the AGC control are required for strong and weak signals. Variation of the AGC control cuts off the picture and sound at one end or produces an overloaded picture at the other end. For proper adjustment of the AGC control, an overloaded picture is produced by turning the control on one side and then backing off a little to produce a clear picture with good contrast. Different settings of the AGC control may be needed on different channels.

Picture Position and Centering Control

Picture tubes are fitted with a pair of deflection coils for horizontal and vertical deflection of the electron beam. The two deflection coils are arranged in a yoke housing fitted round the neck of the picture tube. The entire yoke can be turned round in its housing thereby shifting the raster.

For positioning the picture, the yoke is made free to rotate by loosening a wing nut on the yoke coil. The entire yoke is then rotated till the raster is parallel to the edges of the screen. After this the yoke is pushed in fully and the wing nut tightened.

For centering of the picture, the yoke coil is fitted with two magnetic discs or rings which can either be rotated together or with respect to each other so that the beam can be moved horizontally, vertically or at any other angle. The magnetic rings are first moved together till the best position of the picture in the screen is obtained. These are then rotated with respect to each other till the raster or the picture is properly centered.

While making service adjustments by means of controls at the rear of the TV receiver it is necessary to watch the raster picture carefully. This can best be done by placing a mirror of suitable size in front of the TV screen. If it becomes necessary to make certain adjustments with a pattern generator, these should be finally checked up with the test pattern transmitted by the TV station before the start of a regular telecast.

19.17 CLOSED CIRCUIT TELEVISION (CCTV)

In the closed circuit television (CCTV) system the video output from a TV camera is fed directly by coaxial cable or a low power wireless link to a special type of TV receiver called a monitor which is installed at a remote position so that the picture is produced on the screen of the monitor. A television monitor or a video monitor is an ordinary TV receiver without the RF-IF stages. It reproduces the picture directly from the composite video signal supplied by the TV camera. When a number of cameras are used for monitoring at different locations, a camera selector switch is used to select the signals from different cameras.

CCTV is a very useful system which finds many applications in education, industry, business, medicine and traffic control. In education, one teacher can serve a number of classrooms or one experiment can be shown to many classes. In industry a night watchman can keep a watch on stores and other important locations in a factory. In hospitals, important surgical operations can be shown to students outside the surgical theater and also a watch can be kept on patients in bed. Police can use CCTV for traffic control and control of crime and in banks and business houses, the CCTV is helpful in observing customers and sales people from remote positions.

Another system called the cable television system (CATV) provides television service to customers over a network of coaxial cables. The difficulty of ghosts and other interference is not there in this system.

19.18 COLOUR TV RECEIVER

All modern colour TV receivers make use of solid state circuitry with a number of specially designed ICs. Electronic RF tuners and regulated power supplies is a common feature of these receivers. Remote control operation is a special feature of all these receivers. A colour TV receiver contains all the essential circuits of a monochrome receiver plus some additional circuits required for the reproduction of a colour picture on the TV receiver. Basically a colour TV receiver is a black and white receiver with a decoder for the colour signals. Figure 19.22 is a functional block diagram of a colour TV receiver.

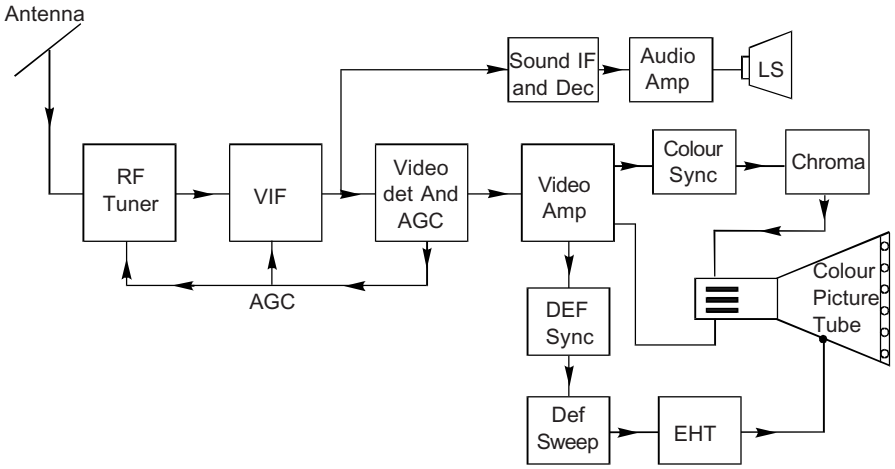


Fig. 19.22 Functional Block Diagram of a Colour TV Receiver

The block diagram shows that circuits like the RF tuner, VIF amplifier, the video detector, the video amplifier, the deflection sync, the sweep circuits and the EHT sections are virtually the same as in a black-and-white receiver. There are, however, some minor differences in design and detail. For example, the RF response in the case of a colour TV is kept more uniform than in a monochrome receiver. This is to avoid any attenuation of the colour sub-carrier. The tuning of a colour TV is rather critical. To avoid any mistuning of the receiver, an arrangement called AFT (Automatic Fine Tuning) is used in most cases. This arrangement is similar to AFC and can be switched off whenever manual tuning is required.

The Colour TV also uses the intercarrier sound system with one difference. The sound take-off point is at the last VIF stage immediately before the video detector. This is done to avoid interference between the sound IF and the chroma signal. A separate diode detector is used to produce the sound IF but the rest of the audio circuits are the same as in a monochrome receiver.

The two main circuits which distinguish a colour TV from a monochrome TV are the colour picture tube and the chroma section containing the colour circuits.

19.19 CHROMA SECTION

The chroma section of a colour TV receiver contains all the circuits necessary for the amplification and detection of the modulated chroma (*C*) signal from the output of the video detector stage. The colour difference signals obtained as a result of demodulation of the *C*-signal are then mixed with the *Y*-signal from the video detector in a special adding matrix. This process produces again the red, blue and green video signals required for the modulation of the three electron beams of the tri-colour picture tube. The chroma section also contains the local oscillator for regenerating the sub-carrier which was suppressed at the transmitter. The sub-carrier is required at the receiving end for the demodulation of the *C*-signal. For the local-oscillator to have a correct phase relationship with the suppressed sub-carrier, it is synchronised with the colour-burst signal received on the back porch of the horizontal blanking pulses. Circuits for the colour burst separator and amplifier together with the phase discriminator are also contained in the chroma section. Besides the circuits mentioned above, the chroma section also contains many automatic circuits to produce a good and stable colour picture. In fact, the chroma section serves as a decoder for the modulated colour sub-carrier *C*-signal.

Two types of colour TV receivers commonly used are the NTSC type and the PAL type. The colour circuits for these two types are more or less the same except for some minor differences explained below:

NTSC TV Receiver

Figure 19.23 shows the chroma section of a NTSC type of colour TV receiver.

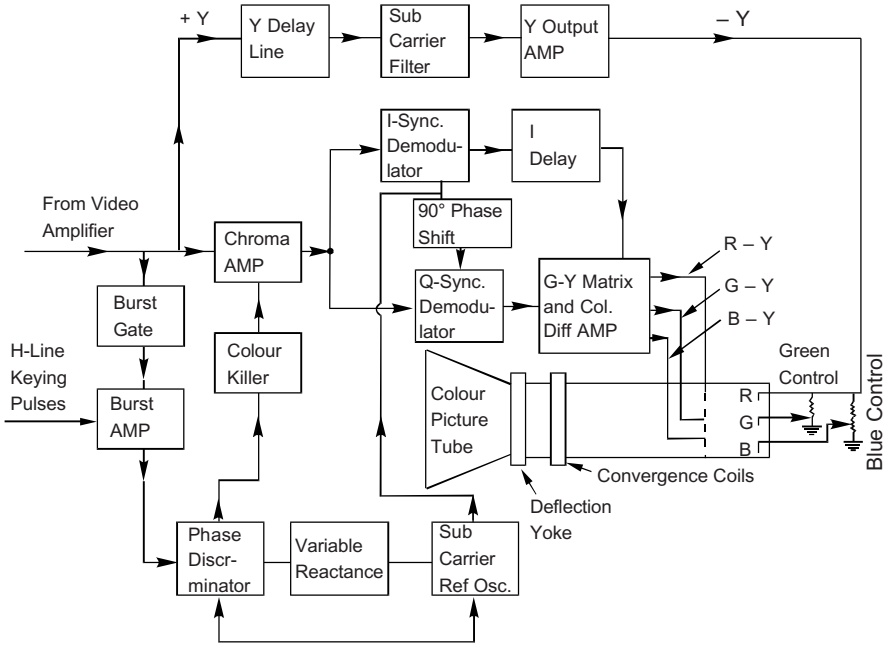


Fig. 19.23 Basic NTSC Receiver

The chroma amplifier is a band-pass amplifier which allows only the *C*-signal with its sidebands to pass through. The *I*-and *Q*-signals are obtained as a result of demodulation by the synchronous demodulators. A synchronous demodulator is a detector circuit that develops an output which is synchronous with the reinserted sub-carrier. Therefore, the reinserted sub-carrier has to be given a 90° phase shift for the detection of *I*-or *Q*-signal. The two colour difference signals *R–Y* and *B–Y* produced by the synchronous demodulator are further added in another matrix to produce the *G–Y* colour difference signal. All the three colour difference signals are then applied to their respective grids in the picture tube which itself acts as a matrix to produce the *R*, *B* and *G* voltages. The matrix action of the picture tube is based on the fact that the phase inverted luminance signal (*–Y*) applied to the cathodes is equivalent to a + *Y* signal applied to the grid. The following algebraic process is performed by the picture tube.

$$R - Y + Y = R$$
$$B - Y + Y = B$$
$$G - Y + Y = G$$

The three electron beams of the colour picture tube are then intensity modulated by the *R*, *B* and *G* video voltages.

Two other circuits that form an important part of the Chroma section are the *sub-carrier reference oscillator* and the *colour killer* stage. The sub-carrier oscillator is generally a crystal oscillator (3.58 MHz) which is kept synchronised to the colour-burst frequency by the discriminator circuit. The phase of the reinserted sub-carrier determines the hue or tint of the picture. The colour-killer cuts off the chroma amplifier when monochrome transmissions are to be received. The colour killer acts like an electronic switch which is closed during colour transmissions and open during monochrome transmissions. The switch is controlled by a voltage developed in the discriminator only when the colour burst signals are received. The device prevents the appearance of colour noise called *confetti* in the picture during the reception of monochrome programmes.

19.20 PAL TV RECEIVER

The block diagram of a PAL TV receiver is given in Fig. 19.24. This is basically the same as the NTSC receiver except that a 0.64 μs delay line has been added before the synchronous demodulators. The output of the delay line is added to the direct signal to detect the modulation of the *U*-signal but subtracted from the direct signal to obtain the *V*-signal. The sub-carrier frequency of 4.43 MHz is supplied direct to the *U*-demodulator but the sub-carrier frequency is inverted in phase on alternate lines before it is applied to the *V*-demodulator. This electronic switch is operated by a 7.8 kHz (half the line frequency of 15625 Hz) square wave received with the colour-burst signal.

The colour difference signals can be combined in a *RGB* matrix to produce the *R*, *B* and *G* signals for the picture tube. Alternatively, the picture tube can act as a matrix for the *R*, *B* and *G* signals as explained in the case of the NTSC receiver.

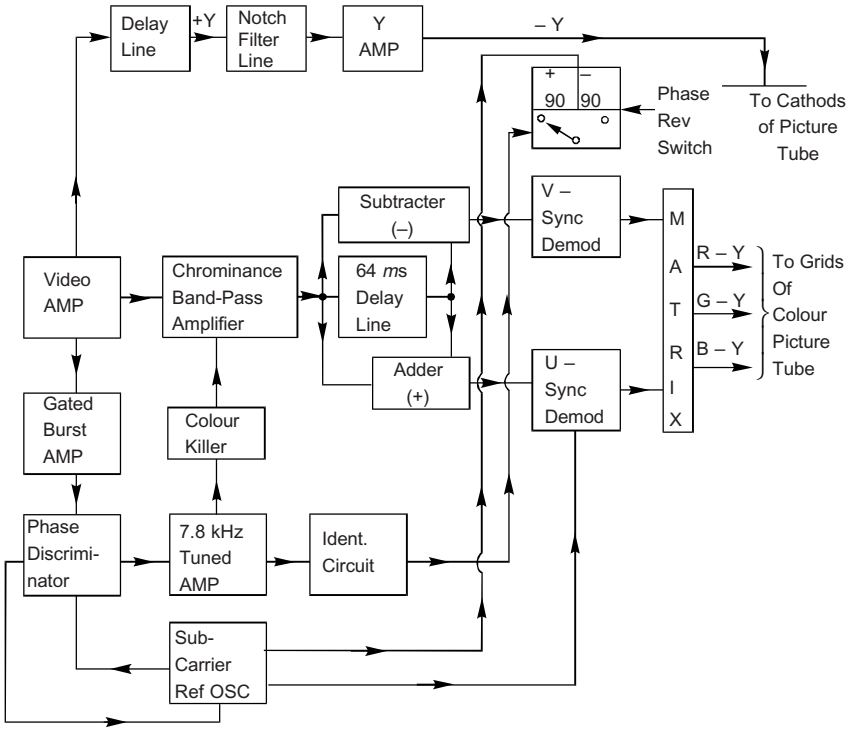


Fig. 19.24 Basic PAL Receiver

19.21 COLOUR OPERATING CONTROLS

Besides the normal operating controls common to a monochrome TV receiver, the colour receiver generally has two additional controls mounted on the front panel of the receiver. These operating controls are the colour control and the tint control.

Colour Control This control is actually a gain control for the chroma band-pass amplifier. As the control is rotated clockwise, the intensity or the shade of the colours in the picture varies from delicate light shades to strong pastel colours. This control is also called the chroma or the saturation control.

Tint Control Tint control varies the phase of the sub-carrier oscillator with respect to the colour-burst signal. Rotating this control clockwise or counter clockwise from its centre position changes the flesh tones in the picture from magenta to red, green, etc. The tint control is also known as the hue control.

Automatic Colour Controls The operating controls mentioned above are manually operated controls. In addition to these controls, there are a few controls which automatically keep the working of certain circuits within specified limits for best results on the TV screen. These controls or circuits should be

adjusted, if necessary, only by experienced technicians. The functions of these circuits and their abbreviations are given below.

Automatic Fine Tuning (AFT) This circuit automatically adjusts the frequency of the local oscillator in the RF tuner for best colour. The AFT can be switched off when necessary for manual fine tuning.

Automatic Frequency and Phase Control (AFPC) This circuit controls the frequency and the phase of the colour sub-carrier reference oscillator with the help of a discriminator and a variable reactance circuit. The function of this circuit is to keep the hue of the picture adjusted to the correct value.

Automatic Tint Control (ATC) This circuit takes over the functions of the manual tint control whenever switched on by an on-off switch.

Automatic Colour Control (ACC) This circuit functions like the AGC for the chroma band-pass amplifier to maintain a constant colour level in the picture. The DC bias obtained by rectifying the colour-burst signal controls the gain of the chroma band-pass amplifier.

Automatic Brightness Limiter (ABL) This circuit controls the bias on the picture tube to keep the brightness of picture constant. The changes in brightness can occur either due to variations in the AC line voltage or due to large variations in brightness of the televised scene.

Automatic Degaussing Control (ADG) Functions of this circuit have already been explained. This circuit demagnetises the steel components of the colour picture tube automatically when the receiver is switched on or off.

19.22 TROUBLESHOOTING IN COLOUR TV RECEIVER

A colour TV receiver is basically a black-and-white receiver with some additional colour circuits. It will, therefore, develop all the trouble symptoms of a normal monochrome receiver plus trouble symptoms that are peculiar to the colour circuits alone. A number of stages like the RF tuner, video IF, video detector and video amplifier are common to both the monochrome and colour TV receivers. Any defect in these common circuits will affect the reproduction of the colour picture because the luminance signals as well as the chroma signals pass through these stages. A weak or completely washed out colour picture can result if the chroma signal is attenuated due to poor alignment of the RF tuner stages and the IF amplifier stages. It is, therefore, necessary that before rectifying colour troubles, it must be made sure that the receiver is able to produce a good black-and-white picture of normal quality. Troubleshooting methods already explained in the case of monochrome TV receivers are equally applicable to colour TV receivers.

The first step in troubleshooting in the case of a colour TV receiver is the proper adjustment and setting of the operating controls including the black-and-white controls. Many troubles in a colour TV are due to improper adjustment of these manual controls. It is to reduce these troubles to a minimum that some of these operating controls have been replaced by automatic controls.

The brightness and contrast controls should be adjusted to get the correct ratio of light and darkness on the screen. The fine-tuning control should be adjusted for best display of colour because any improper adjustment of this control can result in complete loss of colour. The two colour controls, namely the colour control and the tint control, should be adjusted only after the monochrome controls have been adjusted for optimum results. The operation of the colour control will only change the saturation of the picture colours but will not affect the hue or tint of the picture. Any change in tint with the adjustment of the colour indicates a defective or misaligned chroma band-pass amplifier. The tint control should be adjusted to give correct flesh colours. An overall green, red or blue picture indicates improper setting of the tint control.

A colour picture tube is completely different from a monochrome picture tube. The colour picture tube has a number of components mounted on the neck of the picture tube. For proper *set up*, as it is called in colour TV, the adjustments for purity, convergence and gray scale should be correct. Appearance of a large coloured spot which does not change from picture to picture indicates poor purity or improper degaussing. Lack of convergence results in colour fringes around the objects. Improper gray scales is connected with the monochrome sections of the receiver.

If the trouble symptoms in a colour TV receiver still persist even after the operating controls and colour set up have been adjusted, the fault could be attributed to the colour circuits. These circuits include the chroma amplifier, the colour-burst amplifier, the colour killer, the sub-carrier oscillator, colour AFPC circuits and the colour-demodulator circuits. The troubles in the colour sections of the receiver can be located with the help of the troubleshooting chart given in Table 19.3.

19.23 TESTING OF COLOUR TV RECEIVERS

The procedure for the testing and maintenance of monochrome TV receivers already explained is equally applicable to colour TV receiver also. The test equipment required is practically the same for both systems. However, for the testing and adjustment of colour circuits, a special type of signal generator called the *colour bar generator* is required. The functions and use of the equipment are described below.

Colour Bar Generator

The simplest colour bar generator can produce three different types of patterns, viz. the dot-pattern, the cross-hatch patterns and the colour-bar pattern on the TV screen. The dot pattern and the cross-hatch patterns are used for the set up adjustments of the picture tube and the colour bar pattern is utilised for adjustment of colour circuits.

For static convergence, the generator is connected to the receiver to obtain the dot pattern on the TV screen. The permanent magnets on the convergence

Table 19.3 Troubleshooting Chart for Colour Faults

S. No. Fault Symptoms	*Possible Cause*	*Remedial Measures*
1. No colour	1. Chroma amplifier 2. Sub-carrier oscillator 3. Colour killer	Check chroma amplifier circuit and operation of colour killer and associated circuits.
2. Colour snow on monochrome picture	1. RF stages 2. Colour-killer stage	Check alignment of RF stage and also check colour killer and associated circuits.
3. Weak colour	Chroma bandpass amplifier	3. Adjust tuning of bandpass-amplifier transformer.
4. Drifting colours or colour bars	No colour sync	4. Adjust colour phase discriminator or burst amplifier.
5. One colour missing	1. Defective electron gun 2. Defective chroma demodulator	Test individual electron guns and also check the relevant chroma channel.
6. Incorrect relative hues	Phase error in sub-carrier oscillator	Adjust tint control and check sub-carrier oscillator for a leaky capacitor.
7. Abnormally intense colours	Automatic colour control (ACC) or defective colour control	Check ACC circuit and replace defective colour control.

assembly are adjusted till the red, blue and green dots converge into one white spot at the centre.

For dynamic convergence, a cross-hatch pattern is switched on. Convergence adjustments are made to obtain a pattern of white straight lines on the top, bottom, right and left of the screen without any fringes on the borders.

Colour Bar Pattern

The colour-bar pattern produced on the TV screen by a keyed chroma-bar generator is shown in Fig. 19.25.

This display of coloured bars on the screen of the colour picture tube is produced by a keyed chroma-bar generator. The use of this type of generator is quite common for the testing and servicing of colour TV receivers. The performance of the receiver is judged by the brightness and relative positions of the colour bars in the rainbow pattern. If the colours appear to be weak and dull even with the colour control fully advanced, the RF, IF and chroma amplifiers need alignment.

Standard test cards for colour television are transmitted by the TV broadcasting stations before regular transmissions. These test cards are designed to

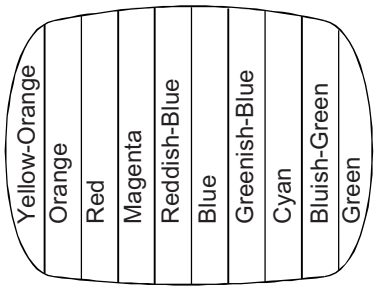


Fig. 19.25 A Keyed Colour-bar Pattern Displayed on the Screen of a Colour Picture Tube

provide visible information regarding the performance of the receiver. These cards enable an experienced technician to make necessary adjustments to the TV receiver.

19.24 COLOUR TV CIRCUITS

A colour TV circuit is more complicated than a monochrome TV circuit. This results from the fact that processing of colour signals requires the use of special components like colour picture tubes and technology used is also different from monochrome TV receiver.

Various types of colour TV receivers are available in the market but all these receivers are based on PAL system which is based on 625/50 CCIR (standards). Any other equipments like VCR used with these receivers must also use the VHS formats only. Receivers using all the three systems i.e. PAL, NTSC, SECAM are available but are very expensive and not for common use. A common feature of all colour receivers is that they employ a regulated power supply system (SMPS), electronic RF tuner and all solid state technology.

Various types of colour TV receivers used in the Indian market are ITT (German), Samsung (Korea), Philips, BPL, Videocon and Sony Trintron types. It is not possible to provide full circuit details for all these types but details of a typical circuit will be discussed here.

Block Diagrams

A block diagram indicates only the various stages in a TV receiver and the way these stages are interconnected. A block diagram does not show the details of components (both active and passive type) and their values etc. It only indicates the types and their circuit functions in the TV set. Figure 19.26 gives a block diagram of a Samsung (Korea) colour TV receiver commonly used in India. PIF signal from the RF tuner (not shown) is fed into the IC-TA 7607 AP which serves as the PIF amp. Det. keyed, AGC one of the outputs from this stage is fed to ICTA 7176 KAZ 10151 F which serves as the SIF amp and AF amplifier and discriminator. The final AF amplifier is a push-pull amplifier formed by 2SC 2073 transistors. The video amplifier consists of 5 video amplifier stages which processes the composite video signal. The output of the second video stage also gives video signals to colour processor and sync processor ICs besides feeding the third video stage.

IC-TA 7193/KA 2151 is the colour processor and produces the *R*, *G* and *B* signals by colour decoding. The colour signals are amplified by their respective video amplifier and feed the cathodes of the colour picture tube. The sync processor IC-TA 7690 processes the sync signals which controls the vertical and horizontal deflection stages in the picture tube. The EHT transformer produces 23 kV which after proper rectification feeds the EHT to the final anode of the picture tube. This also produces other two voltages required for other sections of the receiver.

Summary

The difference between a radio receiver and a TV receiver lies in the fact that a radio receiver is required to convert a modulated sound carrier into sound through the loudspeaker, whereas a TV receiver is required to convert not only a modulated sound carrier into sound but also a modulated picture carrier into a picture through a picture tube. A TV receiver is a combination of a FM sound receiver and a AM picture receiver. Of the three types of TV receivers commonly in use, viz all valve type, hybrid type and solid state type, only the solid state types using ICs and transistor, are in use these days A TV receiver consists of the following stages.

1. Tuner or RF section
2. Picture (video) section
3. Receiver sweep section
4. Sound section
5. Power supply section

Keyed AGC system is used in TV receiver and the sound section uses FM detection compared to AM detection used in video section.

A colour TV receiver is basically a monochrome TV receiver with a colour decoder for the colour signals. The two main features which distinguish a colour TV receiver from a monochrome TV are the colour picture tube and the chroma sector.

A regulated power supply (SMPS) electronic tuning and remote control operation are common features of modern colour TV receiver.

Review Questions

1. Explain the difference between a radio receiver and a TV receiver.
2. State the main categories of TV receivers and briefly mention the characteristic of each type.
3. Why is a solid state receiver more popular than any other type of receiver?
4. Give the block diagram of a solid state TV receiver and briefly describe the function of each stage.
5. What is electronic tuning? How does it differ from mechanical tuning?
6. Explain with a diagram how interfering signals are produced from adjacent channels in a TV receiver. What steps are taken to prevent this interference?

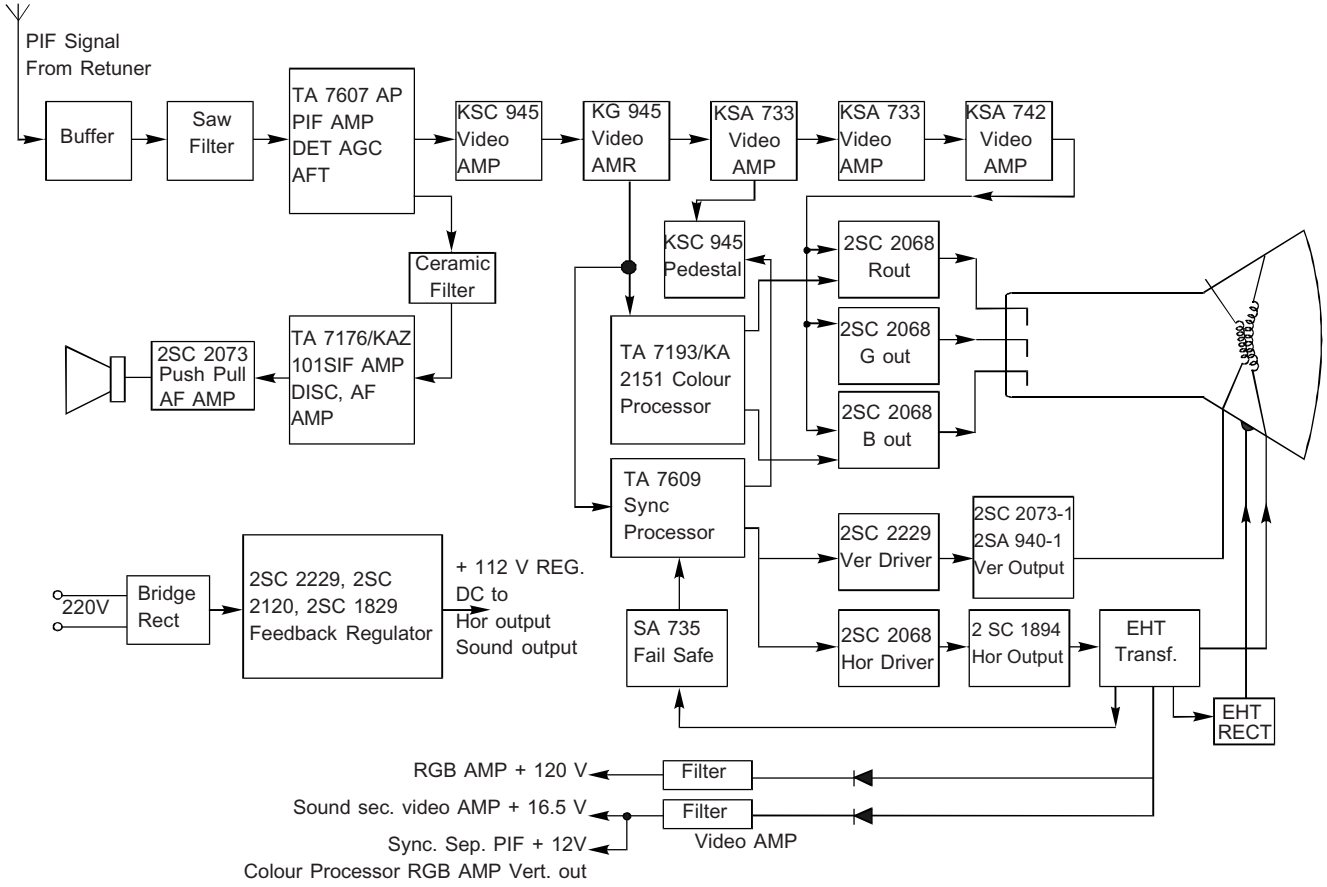


Fig. 19.26 Block diagram of Samsung Colour TV Receiver

7. What is keyed AGC? Explain with a circuit diagram the functioning of transistorised keyed AGC in a TV receiver.
8. Explain the use of ICs in a solid state TV receiver. Name some important ICs used in the (i) Sound section (ii) Video Section of a solid state TV receiver.
9. Give the construction of a delta-gun colour picture tube and explain its working. What are its shortcomings?
10. Define the terms (i) Purity, (ii) convergence, and (iii) degaussing as applied to a colour picture tube.
11. State whether true or false:
    (a) A primary colour cannot be produced by combining the other primary colours.
    (b) Cyan is complementary to green.
    (c) Colour-burst frequency for the NTSC system is 4.43 MHz.
    (d) Colour killer operates to shut down the chroma amplifier when colour burst is received.
    (e) White contains 59% red, 30% green and 11% blue.
    (f) Higher EHT is necessary for a colour TV than for a monochrome TV.
    (g) Degaussing demagnetises the steel fittings in a colour picture tube.
    (h) AFT maintains a constant colour level in the picture.
12. Give the block diagram of a Samsung colour TV receiver and indicate the important ICs used in this circuit.

chapter 20

Applications of Television

INTRODUCTION

Television has a direct and instant impact on the viewer. As such, it is considered to be one of the most powerful mass communication medium for propagating social objectives like family planning, social education, health care and improved agricultural techniques. Besides its entertainment value, the use of TV as a valuable teaching aid for schools and colleges cannot be underestimated.

A TV camera can be considered an extension of the human eye because of its ability to relay information instantaneously. The ability of a TV camera to reach and gather information from such hazardous locations as atomic energy plants, under water environments and outer space, has made it a very useful tool for research and development in various fields of education and research. Although the television technique was originally developed for commercial broadcasting, but the ability to reproduce picture electronically has proved extremely useful in various fields of education industry, business and visual communication in general. Some of the important applications of television will be briefly discussed in this chapter.

20.1 TELEVISION BROADCASTING

The use of electromagnetic waves for television broadcasting has already been discussed briefly in Chapter 12. Two carrier waves are separately modulated with video and audio signals and radiated by an omnidirectional transmitting antenna and received by an unidirectional antenna of the Yagi Uda type. Both the VHF and UHF channels are used for the TV transmission and the range of transmission is confined to line of sight distances and the useful service range is confined to 120 km for VHF stations and about 60 km for UHF stations. The use of satellite communication for TV broadcasting has made extended coverage of TV broadcasting possible. TV broadcasting which originally started as

a black and white transmission has now changed over to colour broadcasting which has a much wider appeal for its naturalness and aesthetic quality.

A wide variety of sophisticated equipment like the electronic news gathering (ENG) camera linked to the studios by means of microwaves are used to produce picture which not only produce a very popular source of further entertainment for millions of viewers but it provides a powerful means for influencing the social and political views of the public at large. Video tape recorders, telecine film cameras for display of movie films and slides are commonly used in TV studios to produce pictures of professional quality and standards.

The production of high quality TV broadcast pictures is limited due to restricted bandwidth allowed by the standards laid down by the International Radio Consulative Committee (CCIR) for TV pictures. Closed circuit TV (CCTV) and Cable TV (CATV) have no such restrictions and can produce high resolution pictures by making use of wider bandwidth.

20.2 CLOSED CIRCUIT TELEVISION (CCTV)

In closed circuit Television or CCTV, the camera signals are supplied to TV receivers or monitors either through a coaxial cable or through low power wireless or microwave links. Monitors are TV receivers without RF and IF stages which are located at suitable locations not very far from the camera set up, Fig. 20.1 shows one such cable link. More than one monitor can also be used through suitable distribution amplifiers and equalizers, power wireless transmitter or a microwave link when the distance separating the camera and the Television monitors (TR) are not very big.

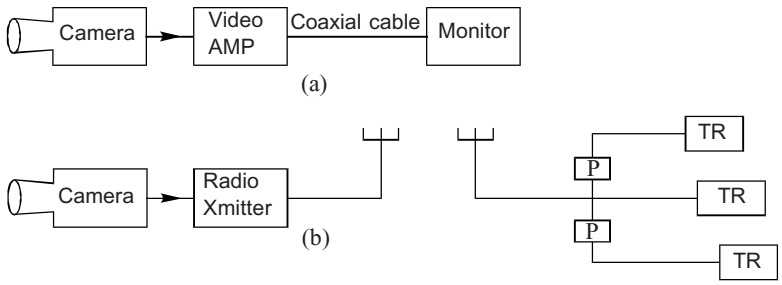


Fig. 20.1 (a) Direct Camera Link Through Coaxial Cable Used Through Suitable Distribution Amplifiers and Equilizers. (b) Shows a Wireless Link Using a Small Transmitter

In closed circuit TV, the signal is not normally radiated into space. These systems may employ more elaborate high resolution standards or use simpler industrial sync. standards. Only small RF powers are required and system costs are not very high. In view of its simplicity and lower cost the CCTV system finds wide ranging applications in education, medicine, industry, aerospace

surveillance, traffic control, security monitor, home security and data transmission. CCTV is particularly useful for remote monitoring and surveillance at hazardous surroundings where human monitoring may be quite risky and dangerous.

20.3 CABLE TELEVISION (CATV)

Cable Television or CATV is a special form of CCTV in which TV signal on standard channels are provided to viewers through coaxial cables on payment of fixed monthly charges as in the telephone system. When a common antenna system is used to deliver a strong signal through coaxial cable to every TV set connected to the system it is called Master Antenna TV (MATV). It is also called community Antenna TV (CATV) system when employed to supply TV signals to hotels, motels, schools, appartment buildings and community sets in small towns and localities.

The Cable Television System is becoming increasingly popular because it provides a strong signal which is free from noise and ghost interference due to multiple reflection from tall buildings in big cities. Line-of-sight propagation results from the use of VHF and UHF frequency and the signal becomes weak at remote or rural locations far from the transmitter. This results in severe interference and more problems and a low signal-to-noise ratio. In big cities, the strong signal in the variety of the transmitting antenna produce severe ghost images on the screen due to multiple path reflection from surrounding tall buildings. Cable TV seems to be the easiest solution to all these troubles. The signals on standard channels are distributed to subscribers through trunk amplifier and boosters, after reception from properly installed antenna systems or from programme studio directly through coaxial cables.

MATV system is also useful in hilly areas where the signal received is very weak in shadow zones. A common MA system installed on a suitable position on a hill can be used to distribute TV signal through coaxial cables to each house in the area providing the benefits of a Community Antenna Television (CATV) to the area concerned.

20.3.1 Ghost Cancellation in Television

Ghost images produced on the television screen due to multiple path reflection from the surrounding buildings near the receiving antenna cause a significant deterioration in the picture quality and is very annoying for the viewer. Till recently the only means available for reduction of ghost effects is to adjust the receiving antenna installation but this method is not very effective when multichannel signals are received from different transmitters on different frequency bands. However, methods have now been evolved by which the TV broadcasting can enhance the reception quality of TV pictures by incorporation of Ghost cancellation Reference (GCR) signals in the TV transmission.

GCR signals for different TV systems (PAL, NTSC, SECAM) have been recommended by the International Telecommunication Union (ITU). The method

involves the transmission of a GCR signal on a single line in the Vertical Blanking Interval (VBI). The ghost cancellation circuit on the receiver continually compares the received distorted GCR signal with a stored reference GCR signal and these attempts to apply an equal and opposite distortion so as to effectively remove or reduce the effect of ghosting on the received TV picture.

Experiments conducted by Doordarshan at DD-1 transmitter (Ch-4) at Delhi have yielded satisfactory results. It is expected that many countries including India will adopt a GCR system within the next few years.

20.4 EXTENDED COVERAGE OF TELEVISION— SATELLITE TELEVISION

Increasing the range of TV transmission either by increasing the height of the transmitting antenna or by increasing the power of the TV transmitter has its practical limitations. Even with the best of transmitting and receiving conditions the coverage of a TV transmitter seldom exceeds 120 km for VHF frequencies and 60 km for UHF frequencies. For extended coverage of TV, the use of microwave links and coaxial cables is often adopted but when the TV programmes are to be selected on National and International basis, the use of microwave links and coaxial cables becomes very expensive. The use of artificial satellites also known as communication satellites is becoming increasingly popular for extended coverage of TV transmission. These artificial satellites placed in equatorial orbits at 120° from each other can cover practically the whole populated land area of the world. How this is done in actual practice is further explained in this chapter.

20.5 WHAT IS A SATELLITE?

A satellite is a heavenly body resolving round a planet. Moon is a satellite of the Earth because it always revolves round the earth. Similarly, Earth is a satellite of the Sun.

Artifical Satellite is a box containing complicated equipment, which is continually flying round the earth. These satellites are used for relaying signals or information from one part of the would to other parts and hence called communication satellites. Communication satellites are of two types.

1. *Tracked Satellites* Eastern communication satellites used for telephone, TV and other data relaying, travelled high up around the earth and the earthly antenna had to follow them or "Track" them constantly. This was a difficult process but with the advance of technology problems have now been solved.

2. *Geostationary Satellites* Satellites for relaying TV programmes direct into peoples homes are called geostationary satellites. Geostationary is a Greek term meaning stationary with respect to earth (Geo-Earth). It can be shown by calculation that an artificial satellite placed at a height of 35803 km directly above equator and moving with a speed of about 11000 km/hour would remain stationary with respect to any point on the earth. At this high speed and high up in the heavens, the satellites goes round the earth once in 24 hours which is

also the time taken by earth to go round its axis once. Thus, there is no relative movement between the earth and the satellite and the latter is called geostationary satellite and no tracking is required for such a satellite. This orbit is also sometimes called the "clarke belt" after the famous scientist and writer Arthor C. Clark, who in 1945 pointed out such a position for an artificial satellite. Figure 20.2 illustrates the various points made above. Some of the technical terms used in connection with satellite communication are given below.

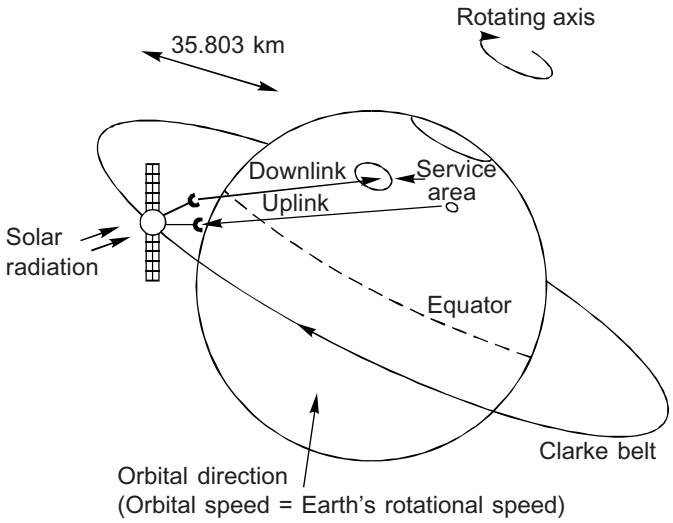


Fig. 20.2 The Clarke Belt, Uplink and Downlink

Up Link The uplink station is a ground station which not only sends signals up to the satellite but it sends signals at a differing frequency generally above Giga hertz band to avoid interference with down link signals. Another function of this uplink station is to control tightly certain other functions of the satellite itself such as its station keeping accuracy. Uplinks are controlled so that transmitted microwave power beam is extremely narrow in order not to interfere with adjacent satellites in the geo-arc. The powers involved are several hundred watts.

Downlink The satellite has a number of *Transponders* which are powered equipments used to re-broadcast the uplink signals down to earth-located receiver. The uplink normally sends signals at a higher Gigahertz range which are received and down converted into lower Gigahertz range. These signals are then boosted by high powered amplifiers for re-transmission to earth. Separate transponders are used for each channel and are powered by *Solar Panels* with back up batteries for eclipse protection. Depending on the power consumption of each transponder, the satellites are categorised as low power (20 W), medium power (45 W) and high power (100 W) satellites.

Medium used-Microwaves The medium used to transmit signals from satellite to earth is microwaves. These are electromagnetic waves which have a

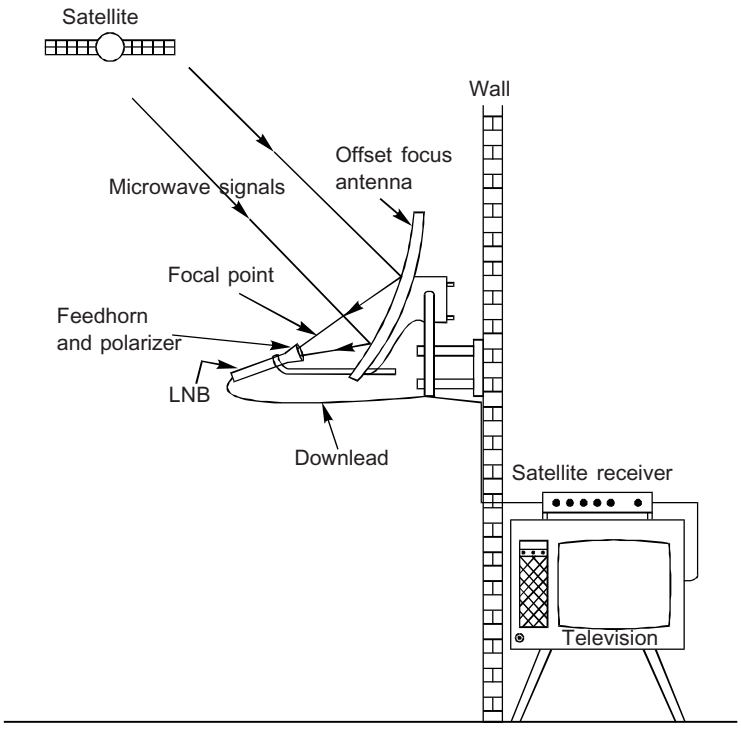


Fig. 20.3 Typical Antenna Mount

much higher frequency than the VHF/UHF bands used for normal broadcast of TV signals. Microwave signals from the satellite are very weak when they reach the earth and the attenuation by water vapour and any other obstruction in the line-of-sight of the antenna can make the S/N ratio very poor unless special precautions are taken by use of specially designed equipment at the receiving site to boost up the received signals. A Television Receive Only (TVRO) consists of an antenna designed to collect and concentrate the signal to its focus where a FEEDHORN is precisely located. This channels the microwave to an electronic equipment called a Low Noise Block (LNB). The LNB amplifies and downconverts the signal to a more manageable frequency for onward transmission through a coaxial cable to the receivers located inside the house etc.

Feedhorn The feedhorn, positioned at the focal point of an antenna, is a device which collects reflected signals from the antenna surface whilst rejecting any unwanted signals or noise coming from directions other than that parallel to the antenna axis. These are carefully designed and precision engineered to capture and guide the incoming microwave to a Polarizer located between the feedhorn. The LNB enables the number of channels to be doubled because two channels can have the same frequency provided these have different polarization (Horizontal or Vertical). This combination of feedhorn polarizer and LNB is termed a Head Unit.

Antenna The Antenna or 'Dish' collects the extremely weak microwave signals and brings them on to a focus. The antenna is in the shape of a *Paraboloid* which has the unique property of bringing all incident rays parallel to the axis to a focus as already explained in Chapters 13. Two types of antennas are commonly used. Prima Focus antenna, when the head unit is mounted in the control axis of the polarized paraboloid and Offset Focus when the head unit is mounted at the focus of a much bigger polaroid of which the observable dish is only a portion. An antenna mount is used to rigidly point the antenna at any chosen satellite. A Azismith/Elevation mount has only horizontal and vertical adjustments but a Polar mount enables the dish to be tracked across the entire visible geo-arc, stopping at any choosen satellite. Polar mounts are capable of receiving a fair number of satellites.

Satellite Receiver The purpose of a satellite receiver is the selection of a channel for listening, viewing or both and transform the signals into a form suitable for input to the domestic TV and stereo equipment. Down converted signals of about 1 GHz are fed by coaxial cable from the LNB to the input of the receiver. TV sets with built in satellite receivers designed to cover both the FSS (Fixed Satellite Service) and DBS (Direct Broadcast Service) will be a common feature of future satellite receiving setup. Figure 20.4 shows a typical satellite receiving setup.

20.6 WORLDWIDE TELEVISION THROUGH SATELLITE

In satellite Television, high power, higher directive transmitters from the land based uplink stations transmit wide band microwave signals to the geostationary satellites above the transmitter. The directive transmitting antenna on the satellite can direct the radiated beam on a narrow region on the earth called a "Foot Print" to provide satisfactory service in that region. High power satellites can provide strong field strengths in the illuminated foot print for national broadcasts. For international coverage more than one satellite may be required to be used which are all in the line-of-sight direction. FM, which has many advantages over AM, is used both for the uplink and downlink transmission.

For extended use of satellites for distribution of national programmes in a vast country like India, the satellites can be used in three different ways.

1. Programmes can be received by sensitive ground station as low power signals from satellites and then rebroadcast through conventional high power TV transmitters. The method is suitable for areas having large number of TV sets.
2. High power radiation from a satellite like NASA, ATS-6, can be received directly through low cost, medium size (3 metre) dish antennas on conventional television receivers fixed with a front end converter. This method was used in India in 1975-76 for Satellite Instructional Television Experiment (SITE) and is quite suitable for sparsely populated rural areas.

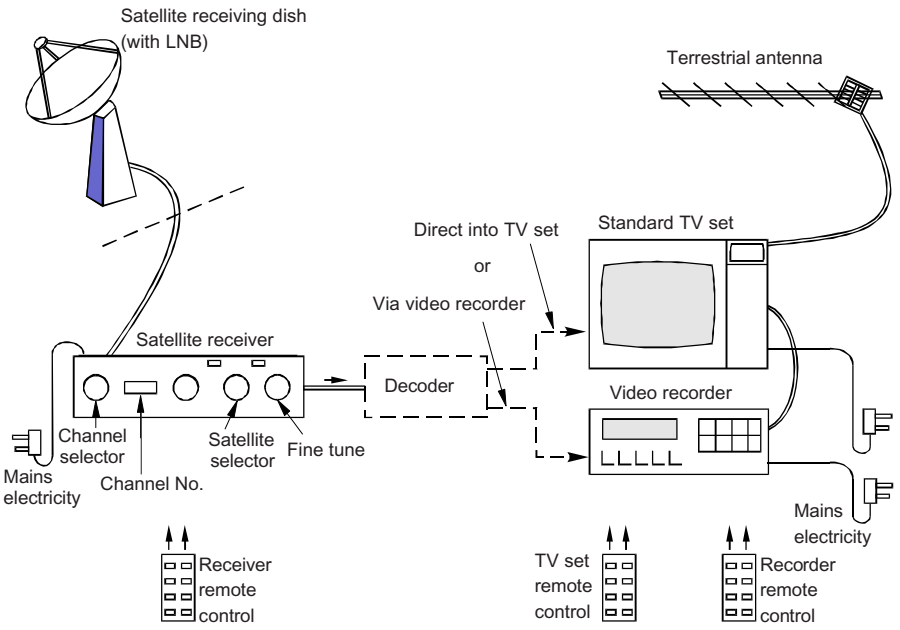


Fig. 20.4 A typical TVRO Layout

The third method is to employ some improved quality receiving terminal equipment and re-diffusion transmitter to broadcast the received signal on a standard channel for direct reception by conventional receivers. Japan was the first country to employ a medium scale satellite 'YURI' for experimental purposes towards direct reception. Advanced techniques towards reducing the cost of low noise front-end receivers may make direct individual reception possible in the near future.

Satellites for worldwide communication are now an established part of television and telephone services operated in the US by Communication Satellites Corporation (COMSAT) in cooperation with the 82-nation International Telecommunication satellite (INTELSAT) Consortium. There are satellites over the Atlantic, Pacific and Indian oceans operating as relay stations between 40 ground stations around the world. The first satellite used for television was TELESTAR in 1962.

As television standards differ from country to country (PAL-625/50 or NTSC-525/30), the transmitting station adopts the standards of the originating country and the ground station converts the received signal with the help of Digital Intercontinental Conversion Equipment (DICE) to other digital standards.

20.7 SATELLITES FOR DOORDARSHAN CHANNELS

Indian TV (Doordarshan) is making use of a number of satellites like INSAT-ID, INSAT-2B INSAT-2C, PAS-4 and INSAT-2D for the telecast of various DD channels.

DD-1 and DD-2 are telecast through transponders on INSAT-2C with footprints from South-EAST Asia to Middle Africa.

INSAT-2D was successfully launched by India Space Research Organisation (ISRO) in June, 1997. INSAT-2D carriers 23 transponders and with INSAT-2E to be launched next year, the satellite system will provide enough transponders for lease to private and government agencies in the country.

20.8 OTHER TELEVISION APPLICATIONS

20.8.1 Video Recording

Video recording is similar to audio recording. In video recording, the video signals are recorded for the reproduction of picture just as audio signals are recorded for the reproduction of sound. Like sound recording, video recording is done both on magnetic tapes and video discs. However, Magnetic Tape Recording is more popular than recording on discs. Magnetic tape recording is much simpler than disc recording and as such magnetic recording is used both for recording and playback but the recorded discs are used for playback only. In magnetic tape recording, the tape consists of a thin plastic coated with fine particles of magnetic oxide. This magnetic tape is pulled across the record/playback head by the transport mechanism of the tape recorder. The signal currents passing through the winding of the record/playback head produce magnetic flux proportional to the strength of the signal currents. The magnetic flux reacts with the magnetic tape to produce the same variations of magnetics on the tape as the variations in the input signal. On playback the moving magnetic tape induces the same signal currents in the record/playback head which are amplified to reproduce the signal.

Video recording has more problems than audio recording. One of the problems is the high frequency range of video signals (30 Hz to 5 MHz) compared to the lower frequency range (20 Hz to 20,000 Hz) of audio signals. This problem was solved by frequency modulating (FM) the video signals on a high frequency carrier before recording. The other problem was the recording of very high video frequencies requiring the use of very narrow head gaps and high tape speeds. This problem was solved by rotating the recording head at a high speed across the width of the tape which moves longitudinally at its normal speed. This increases the ‘‘writing” speed or the relative head-to-tape speed of the system.

Video discs are also used for video playback like the gramophone discs. Digitally recorded video discs called Compact Discs (CD) are also used for playback using large beams. Full details of video recording are available in Chapter 22.

20.8.2 Picture Phone

In picture phone system, the two persons can also see each other while talking on the telephone. The picture phone installation includes a small TV camera

and a miniature picture tube. A slow scanning rate is used for the picture and the highest video frequency involved is limited to 1 MHz. Picture phone service can also be used for showing still picture of drawings, photographs and equipment.

20.8.3 Facsimile

Facsimile is electronic transmission of visual information usually a still picture, over telephone lines. Since there is no motion in the picture the scanning for facsimile can be relatively slow. It is also called slow scan television. Facsimile (Fax) is employed for sending copies of documents over telephone lines.

20.8.4 TV Games

TV games are played on the TV screen by electronically generating video patterns to represent various aspects of the game like balls, bats, rackets and the court boundaries etc. To make the games look as natural as possible sound effects like shots, explosions etc. are introduced. Due to the introduction of sound and colour into these games, these have become a popular source of entertainment on the home TV, amusement parks and commercial arcadis.

TV games is a good example of application of digital electronics and IC technology to TV products. The earlier TV games used solid-state Transistor Transistor Logic (TTL) and provided logic for simple games like question answer games on the TV screen. However, with the introduction of microprocessors (μP) further sophistication has been possible. Although logic can be developed for any game to be played but most common games include tennis, football, squash and rifle shooting.

Figure 20.5 gives a block diagram of a functional TV games set up. The units required consists of (1) a players or users control unit (2) game cum control circuit logic unit and (3) RF oscillator and modulator. A normal TV receiver and a changeover switch completes the entire set up.

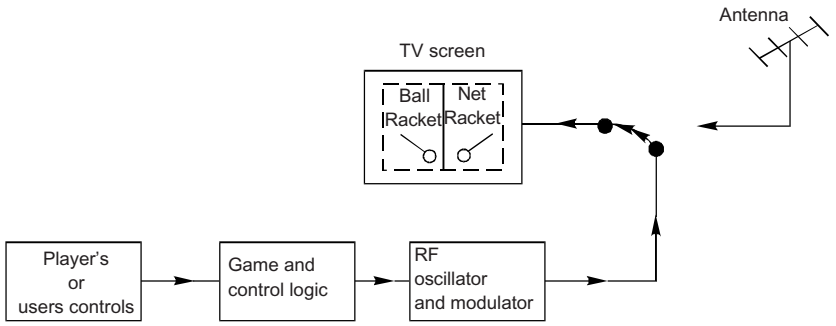


Fig. 20.5 Set-up for TV Games (Tennis)

The players control unit contains various control operation of the setup. The game and control produces necessary video signals for producing characters like rackets, balls and gun field etc on the TV. This section also produces logic

circuits for game-playing rules, scores. This section also produces horizontal and vertical sync pulses for tuning the composite video signal correctly.

The composite video signal containing full information feeds into the RF oscillator and modulator section. The oscillator can be set for any convenient channel (3 or 4) and the output of this is provided to the input of the TV receiver through the changeover switch 5 which can isolate the antenna when required.

As the number of games provided by the TTL circuits increased, the circuits became complicated. Accordingly manufacturers of TV games are now using special purpose or dedicated ICs and microprocessors for designing all types of complex games and their logic circuits. Dedicated n-channel MOS chips are now available both for the PAL (625/50) and NTSC (525/60) colour systems:

20.9 HD TV

HD TV or High Definition Television is a special type of television system which provides much bigger and better quality pictures than are provided by the conventional TV receivers based on PAL or NTSC systems. Merely increasing the size of the screen only increases the size of the picture but the picture quality suffers due to scanning lines becoming visible and the colour becoming blurred and fuzzy. A real HDTV will be required to include the following four important features:

1. Large Screen Size
2. Wider Aspect Ratio
3. Better Resolution.
4. Lack of Spurious Effects.

These features will now be considered one by one.

1. Larger Screen Size If the present cathode ray type picture tubes continue to be in use which is not certain, then HDTV tubes will have to be larger than the present day flat screen tubes of 68 cm size-construction of a one meter diagonal tube will be possible within the next few years and this size will be the most practical size for the domestic TV sets keeping in view the cost and the size of normal living rooms.

2. Aspect Ratio A one metre diagonal flat screen will be approximately, 86 cms wide and 52 cm high which gives an aspect ratio i.e of width to height as 5 : 3. The normal aspect ratio for a conventional TV is 4 : 3. However, keeping in view the wider aspect ratios used for cinema film displays, a slightly wider aspect ratio of 5.33 : 3 or 16 : 9 has been considered most suitable for HDTV.

3. Better Resolution Better resolution for finer details in the picture requires a larger no. of lines per frame. For worthwhile gain in picture definition, the no. of lines used should be more than 1000. Systems using between 1100 or 1500 lines have been suggested. Larger no. of lines scanned produce greater bandwidth for the video signal. For any increase in vertical resolution there has to be a corresponding increase in the horizontal resolution. Keeping in view the aspect ratio, the bandwidth required for HDTV will be about 30 MHz.

4. Lack of Spurious Effects The most noticeable effect of this type is cross colour in which the finely striped or checked spots cause spurious coloured patterns which move about with any picture movement. The luminance information breaks through into the chrominance channel, causing a fine crawling dot pattern known as cross-luminance. Any high quality television system should be free from these defects.

First HDTV system suggested was the Japanese 112/60 system. This system lacked compatibility as HDTV viewers will need to have two receivers in the same room unless the HDTV receiver also includes the circuitry of a conventional receiver in the same cabinet. This will make the set very expensive.

Along with the Japanese 1125/60 standard an European 1255/50 standard was also proposed with an American 1050/60 standard but no standard for universal adaption has so far been proposed.

Although the development of HDTV involves a large no. of technical problems but with the advance in TV technology a system compatible with the existing domestic receiver will soon be evolved.

20.10 REMOTE CONTROL

A TV or a VCR can be operated, with all its functions controlled from a distance by means of a hand-held unit called a remote controller or remote commander. A TV set can be put into any of its operational modes by a remote commander and the user need not even leave his seat. The TV and the VCR must have a separate remote controller and one cannot be substituted for the other.

Remote control can be effected in two ways. In corded remote control, a coaxial cable connects the remote control unit with the TV, but in cordless remote control, either an infra-red or an ultrasonic link is used between the hand-held remote commander and the TV set. Infra-red and ultrasonic waves travel in straight lines and for satisfactory remote control, the TV has to be in direct view of the remote commander. The corded remote control, though cumbersome, can be effective from any position in the room including round the corners. If remote control is used, both for the TV and the VCR, it is advisable to have one control corded, to avoid any chances of one remote controller accidentally upsetting the functions of the other, in case both are cordless controls.

A remote control panel duplicates all the functions of the control panel of the TV. It is a sophisticated piece of equipment, which makes use of digital encoder chips to convert the key-board signals into digital commands. These are communicated to the receiver unit in the infra-red channel and decoded, to produce pulses that are unique to the particular mode desired.

An infra-red remote commander has a transmitter (Fig. 20.6), which must be pointed towards the detector in the (TV) unit before any button is pressed on the key-board. The remote commander has a distance range of about 8 meters (25 ft) and an angular range of about 30° on either side of the line joining the transmitter and the detector. It is a battery operated unit which requires 2 dry

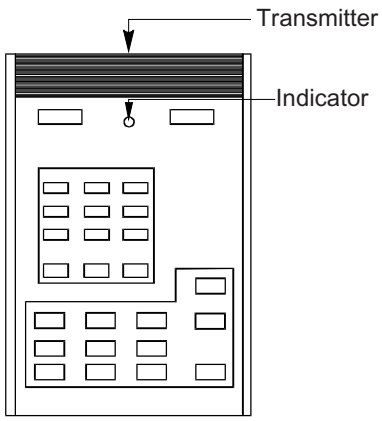


Fig. 20.6 Infrared Remote Commander

cells of 1.5 V which must be inserted as per the manufacturer's instructions. In normal use, the battery life is about 6 months. If the commander is to remain out of use for a long period, the battery must be removed to avoid any damage due to battery leakage. If the indicator light does not light up when a button is pressed, the batteries are discharged and must be changed immediately. Both cells must always be replaced by new ones. The commander must be kept away from hot and humid places.

20.11 TELETEXT

Teletext is a system of transmitting digitally coded alphanumeric (display letters or numbers) data in the field blanking interval (FBI) of a television signal without disturbing the normal vision or sound signals. The data signal is then decoded in the receiver and displayed as a page of information on the screen as an alternative to the video picture. Various methods for utilizing the spare time in the field blanking interval have been proposed as a means of transmitting additional information to the home but it was not till RAM (Random Access Memory) and ROM (Read only Memory) integrated circuits (ICs) became available that a practical TELEXT system became potentially viable.

One of the systems proposed by UK in 1974 had the following technical features.

1. Data pulses were transmitted on otherwise unused television line during the field blanking interval (FBI) using a bit rate of 6.9375 M bit/s (44x nominal line frequency)
2. Each television data line carried all the information for a complete 40-character display row.
3. Each page consisted of 24 rows of 40 characters, using both upper and lower case characters including a special top row called the "page header" carrying information for control and display purposes.
4. Using one data line per field, the system allowed two full pages per second to be transmitted.
5. All the data words were 8 bit in length and parity protection was used for the character data words.

6. Every page header carried check-line information for a line display and provision was also made for news flashes and sub-letters. A colour graphs facility was also provided. Control characters were used for colouring and flashing of selected words.

20.12 DIGITAL TV

In digital TV, both the sound and the picture signal have to be processed in accordance with digital technology. Whereas the use of digital technology of audio signals had long been established with the production of compact discs (CD), digital vision techniques have been difficult to establish because of the sheer volume of fast changing data involved necessiating the use of wide band transmission channel, large memory stores and very large scale integration ICs which have become available only recently. In digital TV the processes involved are the conversion of the analogue form into digital form, transmission and reception in digital form and the reconversion into analogue form to suite the requirements of the display tube. The following essential steps are involved in the process.

1. Sampling This is achieved by measuring the instantaneous amplitude of the waveform (sound or picture) at regular time intervals, short enough to ensure that the highest frequency required is adequately sampled. The sampling rate should be at least twice the highest frequency involved.

2. Quantizing The amplitude of the signal at each sampling spot is measured and assigned a number which represents the amplitude of the signal at that moment of time and is called the *quantizing level*. The number of quantizing levels depends on the nature of signal being sampled. In the case of TV signal the required levels for a good picture is 256 ($2^8 = 256$) but a noise free sound signal requires 1024 (2^{10}) quantizing levels.

3. Binary Conversion The quantizing levels are then converted into binary (I's, O's,) and are represented by series of pulses each of which has clearly defined time slot in serial transmission. These pulses can be distorted in transmission but they can still be identified as pulses and can trigger a generator to produce new pulses.

This process of converting analogue to digital form is called encoding and is done by an equipment called Analogue to Digital Convertor (A/D).

4. D/A Conversion-Decoding The digitally converted and processed signal is received at the TV receiver and reconverted into analogue form to suite the requirements of the display device, be it a picture tube or a solid state panel.

Advantages of a digital TV are usual advantages of a digital system.

SUMMARY

Television is considered to be one of the most powerful communication medium. Besides its entertainment value, it has a direct impact on the social, educa-

tion and healthcare habits of the masses. Some of the important applications of television are given in this chapter.

Television Broadcasting is similar to radio broadcasting but makes use of two modulated carrier in the VHF/UHF range, one for picture (video) and the other for sound (audio). It has a limited coverage but extended coverage is produced with the help of satellites.

Closed Circuit TV (CCTV): In this the camera signals are distributed through coaxial coils, to a number of monitor receivers placed at suitable locations. This is particularly suitable for surveillance, traffic control and remote monitoring particularly at hazardous places.

Cable Television (CATV) In this a common antenna or antenna system is used to supply strong signals to subscribers through coaxial cables. It is also called Master Antenna TV (MATV) or Community Antenna TV.

Satellite Television is one of the method for providing extended coverage of television by use of one or more geostationary satellites.

These satellites rebroadcast the interlinks TV signals to earth located receiver through down link station. The signals received from satellites are rebroadcast by ground stations in different ways to provide extended coverage for the programmes. Other application of television include picture phones, facsimile, TV games, etc. HDTV, Teletext and digital TV are some of the latest trends in television technology.

Review Questions

1. Describe a TV Broadcast system. What are its application?
2. Explain with a block diagram the working of a closed circuit Television. Give its application.
3. What is CATV? Give its merits and applications.
4. How are 'ghosts' formed on the TV screen? Explain the use of ghost cancellation reference signals (GCR) and how do these reduce the effect of ghosts on TV picture.
5. What is a geostationary satellite? Explain the terms 'Uplink', 'Downlink' and Transponder as used in satellite communication.
6. Explain the use of satellites for extended coverage of National programmes by Doordarshan.
7. Give the block diagram of a TV RO set up and explain its working.
8. What are the main problems of video recording and how are these solved?
9. Draw a functional block diagram of a TV games set up and explain its working.
10. Give briefly the features of the HDTV system. What are its advantages and limitations.
11. Write short notes on
    (a) Remote control (b) Teletext
    (c) Digital TV.

PART V

AUDIO AND VIDEO SYSTEMS

chapter 21

Acoustics, Microphones and Loudspeakers

21.1 NATURE OF SOUND

Sound is a sensation produced by nerve-pulses sent to the brain when the eardrum is subjected to variations of air pressure casued by some vibrating object. The human ear is an extremely sensitive pressure gauge which can respond to the minutest changes of air pressure, which constitute the sound waves. In fact, the word 'sound' is used for both the physical waves and the sensation produced by them. When produced by the regular motion of an object like a tuning fork, sound gives the sensation of a simple tone but the sensation becomes noise when produced by some irregular motion like the banging of a door or the bursting of a cracker.

How sound energy is transferred from one place to another, through a medium like air, can best be explained by the example of a tuning fork which is a device that produces a simple note or sound when its prongs are struck against some hard object. The prongs of the tuning fork start vibrating to and fro whereby the particles of air near the prongs are first pushed away and then sucked back. The particles of air also perform a to and fro shunting action like the prongs of the tuning fork and the air surrounding the tuning fork is divided into alternate regions of *compression* and *rarefaction* as shown in Fig. 21.1(a). The speed with which these regions of compression and rarefaction spread out in all directions is the speed of sound. This speed depends on the closeness of mass of the particles of the medium involved. For air at normal atmospheric pressure and temperature the speed is found to be approximately 1100 ft/s or roughly 750 miles/h. If the particles of the medium employed are heavier than those of air, sound will travel slower.

If the relationship between the variations of pressure above (compression) and below (rarefaction) the normal atmospheric pressure and the distance from the source of sound (tuning fork) is plotted as a graph it will have the familiar

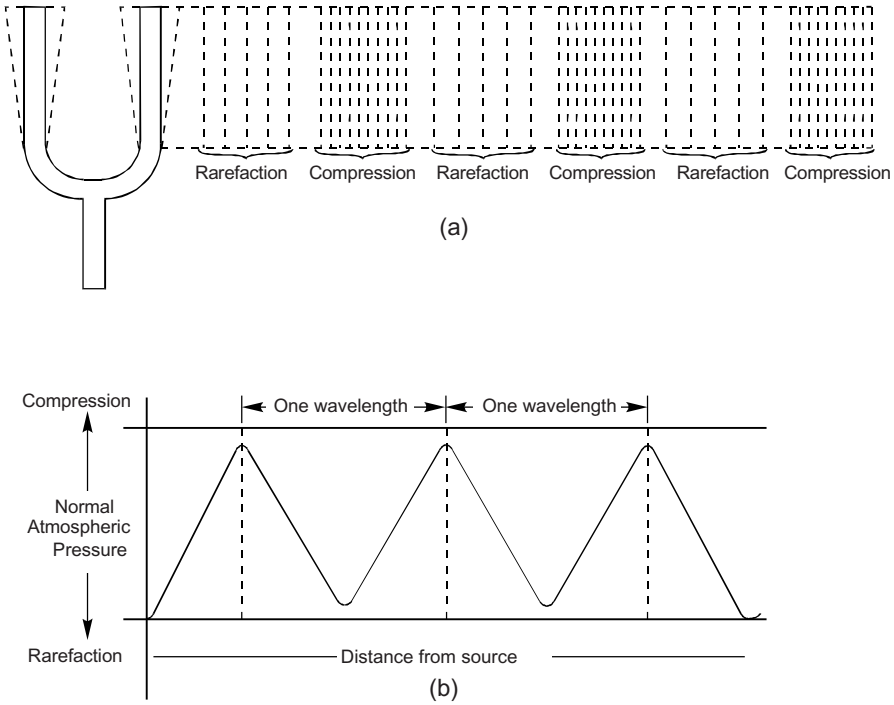


Fig. 21.1 Nature of Sound Waves (a) Regions of Rarefaction and Compression Produced by a Vibrating Tuning Fork (b) Graph Showing Variations of Pressure with Distance

form of a sinusoidal curve as shown in Fig. 21.1(b). The distance between two successive regions of compression (or rarefaction) constitutes a wavelength. The relation between the wavelength, frequency and speed is the same as in the case of radio waves and is reproduced below.

$$\text{Wavelength} \times \text{Frequency} = \text{Velocity of sound}$$

The main difference lies in the velocity of the waves which, in this case, is the velocity of sound waves which is much lower than the velocity of radio waves. Accordingly, for the same frequency, the sound waves are much shorter than the corresponding radio waves.

Example What is the wavelength of sound produced by a tuning fork vibrating at a frequency of 220 c/s at normal atmospheric pressure and temperature?

In this case, frequency = 220 c/s

and velocity of sound at NTP = 1100 ft/s

$$\therefore \quad \text{wavelength } (\lambda) = \frac{1100}{220} \text{ ft}$$

$$= 5 \text{ ft}$$

Music generally consists of high or low notes which are nothing but high or low frequencies. All musical instruments do not produce pure notes, as in the

case of a tuning fork. For example, if a note or sound produced by a violin is analysed, it is found to contain not only the fundamental but a number of other frequencies also which may be multiples of the fundamental frequency. Thus, the note corresponding to a frequency of 250 c/s will also contain frequencies of 500, 750, 1000, 1250 c/s. The multiple frequencies are konwn as *harmonics* or *overtones* and it is the proportion which each bears to the fundamental that decides the quality or 'timbre' of the instrument producing it.

21.2 LOUDNESS OF SOUND

Loudness of sound is the sensation or stimulation produced by the intensity of the sound falling on the eardrum. The intensity of sound is the total power contained in the sound waves hitting the eardrum. Whereas the intensity of sound is measurable by scientific instruments, the loudness depends on certain other factors like the sensitivity of the ear, the attentiveness, the mood and interest of the listener and, as such, is also psychological in nature besides being a physical sensation.

There is a certain degree of loudness above which the ear can hear a sound and below which it cannot. This degree of intensity of sound is called the *threshold of hearing*. Similarly, when the sound gets louder and louder a limit is reached beyond which a painful and uncomfortable feeling is produced in the ear. This limit is known as the *threshold of pain*. Since the ear is not equally sensitive to all frequencies both the limits of hearing mentioned above vary with the frequency. Futhermore, the range of human hearing extends from about 20 to 20,000 Hz. This range of frequency known as the audio frequency range varies with age, being higher in the case of children than other people who, in most cases, cannot respond to frequencies above 10,000 or 12,000 Hz.

Another important characteristic of human ear is that it is very slow to respond to increases of intensity of sound or loudness. Thus, a sound that increases in loudness or intensity by twice the original value is not heard twice as loud but hardly produces any sensation or change of loudness. In order that a sound is heard twice as loud as the original sound, its intensity must be increased ten times. In the same way to hear a sound three times and four times as loud as the original sound, the intensity of the sound must be increased a hundred times and a thousand times respectively over the original sound. Such a relationship between two quantities where one jumps up in steps of 1, 10, 100, 1000, etc. to make the dependent quantity change in steps of 1, 2, 3 and 4 is called *logarithmic*. The response of the ear to changes of intensity of sound is logarithmic. It is, therefore, natural to adopt a unit which is also logarithmic in nature for the measurement of sound intensity. Such a unit is called the *decibel*.

21.3 DECIBEL

The unit adopted for comparing or measuring sound levels is known as *bel*, named after Alexander Graham Bell, the inventor of the telephone. A more

commonly used logarithmic unit is the decibel which is one tenth (deci is one tenth) of a bel. Decibel is usually written as db. If the two power levels to be compared are P_1 and P_2, then the difference in the two levels expressed in db will be given by the mathematical expression

$$\text{db} = 10 \log \frac{P_2}{P_1}$$

P_2/P_1 is only a ratio and its logarithmic value can be determined from the log tables given in mathematical books. Thus, if $P_2/P_1 = 2$, then we know from log tables that log 2 = 0.301. So the above change of level from P_1 to P_2 when expressed in decibels will be $10 \times 0.301 = 3.01$ db or 3 db approximately.

It has been found that a change of level of 1 db is barely perceptible to the ear and an increase of 2 db is only slightly felt. An average person does not notice a change of sound level if it is less than 3 db.

If the change of power is a decrease in power, it is expressed by putting a minus sign before db. In a case where the ratio $P_2/P_1 = \frac{1}{2}$, the reduction in power is expressed as –3 db.

If the power level decreases from 4 W to 1 W, the decrease will be expressed as –6 db.

Decibels are also used to express the gain of amplifiers. If E_1 is the input voltage and R_1 the input resistance of the amplifier, then input power $W_1 = E_1^2/R_1$.

Similarly, the output power $W_2 = E_2^2/R_2$, where E_2, is the output voltage and R_2 the output resistance.

$$\frac{W_2}{W_1} = \frac{E_2^2}{R_2} / \frac{E_1^2}{R_1} = \frac{E_2^2}{E_1^2} \cdot \frac{R_1}{R_2}$$

If $R_1 = R_2$ (constant),

$$\text{the gain in db} = 10 \log \frac{W_2}{W_1}$$

$$10 \log \frac{E_2^2}{E_1^2} = 10 \log \left(\frac{E_2}{E_1} \right)^2$$

$$= 20 \log E_2/E_1$$

Example What is the gain of an amplifier in db when an input votage of 1 V produces an output voltage of 10 V, R being constant.

$$\text{Gain in db} = 20 \log \frac{10}{1}$$

$$= 20 \log 10 = 20 \times 1 = 20 \text{ db}$$

Decibels express only power ratios but to express the output of an amplifier in absolute units like watts, it becomes necessary to adopt a reference level

which will be called zero db. The reference level generally adopted is 0.006 watt or 6 milliwatt unless some other reference level is stated.

Example Express the power output in watts when an amplifier has a gain of 20 db, if the reference level = 0.006 W.

$$20 = 10 \log \frac{W_2}{W_1} = 10 \log \frac{W_2}{0.006\,W}$$

or $$\frac{20}{10} = \log \frac{W_2}{0.006}, \quad 2 = \log \frac{W_2}{0.006}$$

now, $\log 100 = 2$

$\therefore$ $$100 = \frac{W_2}{0.006}, \quad \text{or}$$

output $$W_2 = 100 \times 0.006 = 0.6 \text{ W}$$

When the reference level adopted is 1 mW in 600 Ω, the unit of power level is known as the volume unit or VU.

21.4 ACOUSTICS

All the sound produced by a source does not reach the eardrum directly, otherwise only one person will be able to hear the sound. A part of the sound that does not hit the ear directly strikes the walls, ceiling or floor of the room in which the source of sound lies. Like all other types of waves, sound waves also get either reflected or absorbed or a bit of each. The energy which is absorbed will not be heard any more but the reflected sound will again hit a second obstacle and suffer part reflection and part absorption and so on until after repeated reflections and absorptions it will reach the ear when it will have become very weak. There are two effects produced on the listener by these reflections. Firstly, the energy received by the ear will be both from the direct waves and the reflected waves and it will give the effect of greater "*fullness*". Seccondly, owing to the greater distance travelled by the reflected waves, the sound will be prolonged and this phenomenon is called *reverberation*. If the room is so constructed that there are many reflections, the time of reverberation may be quite appreciable, sometimes several seconds. A certain amount of reverberation is desirable in most cases. Materials which have different reflecting properties are used in the construction and treatment of the surfaces of buildings for obtaining different reverberation periods. The science which deals with the design and construction of rooms, studios and halls meant for particular purposes like speech, music, dance and cinema halls, is called 'acoustics'. A studio or a hall is said to be live or dead depending upon whether it has a high or low reverberation period. For good listening, the walls, ceiling, floor of the listening rooms, halls and auditoria have to be acoustically treated with suitable acoustic materials.

21.5 REVERBERATION TIME (RT)

The sound produced in any room (enclosure with walls, ceiling and floor) does not die immediately. This sound spreads out to various surfaces and is reflected speedily from one surface to another, suffering some loss of energy at each reflection. Eventually, after possibly a very large number of reflections, the sound dies down to inaudibility.

This process called *reverberation* is illustrated in Fig. 21.2.

The question that arises at this stage is as to how much time will the reverberation take to die down to inaudibility. The answer was provided by an American scientist W.C. Sabine. As a result of a large number of experiments conducted by Sabine, he concluded that the sound pressure dies down to inaudibility when the sound pressure has fallen to one-millionth (10^{-6}) of the original value. The time taken is called the *Reverberation Time (RT)* and is defined as the time taken for the sound in a room to decay through 60 db. Figure 21.3 shows this diagramatically.

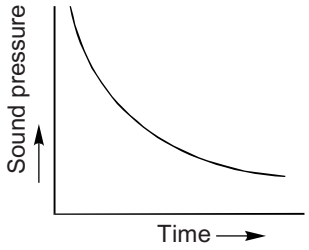


Fig. 21.2

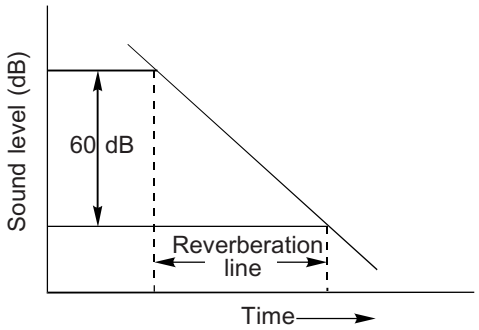


Fig. 21.3 Reverberation Time (RT)

The curve is a straight line instead of being exponential because vertical scale is in decibles instead of being in units of pressure.

Reverberation time is mainly affected by two things

1. The amount of sound absorbing material in the room
2. The size of the room. The bigger is the size of the room, the longer it takes for the sound waves to travel between reflections and more is the loss of energy due to reflection. In fact, RT is proportioned to the volume of a room.

21.6 SOUND ABSORPTION

All materials absorb sound to some extent. Hard inflexible substances with shining surfaces may absorb very little sound.

The term *Absorption Coefficient* is used to represent absorbent properties of materials. If the absorption coefficient of a material is represented by α, the

$$\text{Absorption Coefficient } \alpha = \frac{\text{Amount of sound energy absorbed}}{\text{Total incident sound energy}}$$

Thus, if a material absorbs only 10 units out of total of 100 units striking the surface, the $\alpha = \frac{10}{100} = 0.1$

Accordingly, a perfect reflecting material will have $\alpha = 0$ and a perfect absorber would have $\alpha = 1$.

It may, however, be pointed out that the absorption coefficient α generally varies with frequency. Most substances are better absorbers at high frequencies than at low frequencies. If the value of α at 500 Hz is taken as a representative value, then Table 21.1 gives the absorption coefficients of some typical materials used for the acoustic treatment of studios, halls and auditoriums, etc.

Table 21.1 Absorption coefficients of materials

Absorbent material	*Absorption coefficient (α)*
Brick wall	0.03
Heavy drape curtains	0.40
thick carpet	0.21
Wooden chair	0.17
Celotex	0.36
Acoustic tiles	0.55
Audience etc.	0.84
Wooden floor	0.09
Glass panes	0.25
Open window	1.00

21.7 CALCULATION OF REVERBERATION TIME (RT)—Sabine Formula

It was prof. W.C. Sabine who gave a formula for the relationship between Volume (V), absorption coefficient α and the reverberation time RT.
According to Sabine's formula.

$$RT \text{ (seconds)} = \frac{0.16\,V}{S_1\alpha_1 + S_2\,\alpha_2 + S_3\alpha_3 + \cdots}$$

$$= \frac{0.16\,V}{\Sigma S\alpha}$$

Where S_1 is the area whose absorption coefficient is α_1, S_2 the area whose absorption coefficient is α_2 and so on. For example, if we have a wall whose dimension are 5m × 8 m so that its area = 40 sq meters. If the average α for the wall is 0.4, then S_α for the wall would be 40 × 0.4 = 16 units. These units are called sabines after the pioneers of practical acoustics Prof Sabine.

Thus 1 sabine = 1 m^2 of perfect absorber with $\alpha = 1$

Sabines formula can now be written

$$RT = \frac{0.16\,V}{\text{Total no. of sabines}} \text{ seconds}$$

This formula is in MKS unit with V volume in cubic meters and S or the area in sq. meters.

Example An auditorium is 30 m long, 20 m wide and 5 m high. The ceiling is covered with celotex and the floor has a wall to wall thick carpet. There are four open windows of size 2m 1m and 5 drape curtains of size 2.5 m 1m covering the walls. If the audience consists of 520 persons, calculate the reverberation-time of the auditorium.

Total Volume of the auditorium is $30 \times 20 \times 5 = 3000$ cu.m.

Absorption by ceiling	$= 30 \times 20 \times 0.36 = 216$ sabines
Absorption by floor	$= 30 \times 20 \times 0.21 = 126$ sabines
Absorption by windows	$= 2 \times 1 \times 4 \times 1 = 8$ sabines
Absorption by curtains	$= 2.5 \times 1 \times 5 \times .4 = 5$ sabines
Absorption by audience	$= 520 \times .84 = 446$ sabines
Total absorption	$= 601$ sabine ≈ 600 sabines

$$RT = \frac{0.16 \times 3000}{600} = 0.16 \times 5 = 0.8 \text{ seconds}$$

Sabines formula is resonably accurate when the amount of absorption is small. If we assume that all surfaces in a room are covered with material with $\alpha = 1$, common sense says that the reverberation should be zero but the calculation does not work out to be that way. There is a modified form of sabines formula which is known as Eyring formula which gives better and more accurate results. According to Eyring formula

$$RT = \frac{0.16\,V}{-S \log e^{(1-e^{\alpha'})}}$$

Where S is the total area of all surfaces and α^1 is the average absorption coefficent.

21.8 OPTIMUM REVERBERATION TIMES

There is no simple law which allows us to calculate what the RT should be for any particular room, studio or concert hall. These are preferred figures but these have been obtained as a result custom, practice and history rather than any scientific study. Table 21.2 gives the average value of RT for some commonly used rooms, concert halls, auditoriums, etc.

21.9 ACOUSTICAL DESIGN OF STUDIOS AND AUDITORIUMS

A studio centre for broadcasting will have a number of studios meant for talks, drama, music, and news, etc.

Each of these studios will have a different acoustical design to produce the required quality of sound which is mainly governed by the reverberation time.

Table 21.2 *RT* Values

Name of the struture	*RT in second*
Household sitting room	0.5
Churches	12 seconds to 1 second at low frequencies
Concert halls	≈ 2 falling at high frequencies
Studios for orchestral Music	about 2 seconds
Studios for Chamber Music	1 second
Studios for pop music	0.5 second
Studios for speech, News and Talks etc.	0.4 second
Drama studios	0.1 to 0.2 second
Television studio	varies from 0.7 to 1.1 second
Theatres	1 second

Studios for talks, news and speech are generally treated to provide very low RT and are generally known as "dead". This is to provide clarity of speech. Drama studios have slightly high *RT* and the music studios have much higher *RT* to provide colour to the music. Temples, churches and other religious places have high *RT* suitable for devotional music.

Lecture halls, halls for factories and workshops should be less vibrant with *RT* varying from 0.5 seconds to 0.3 seconds.

After a suitable constructional design, the studios, theatres and auditoriums have to be acoustically treated with suitable acoustic materials which may also include sound absorbers.

21.9.1 Sound Absorbers

A sound absorber is a material which reduces air particle movement by friction. There is a range of commercially available sound absorbers—often in the form of some sort of tiles which can be fitted to the ceiling or walls. Some of the absorbers are porous absorbers or fibrous materials, wideband porous absorbers which were specially designed absorbers used for the BBC studios. Panel absorbers are made of plywood backed by an energy absorbing material. Helmholtz Resonators named after the German Scientist Hermann Helmholtz, are resonators which can either absorb or re-radiate sound depending on the friction in the neck. Other absorbent materials used are perforated/porous boards, felt, asbestos, cloth, maps and pictures on the walls. The floor should be covered with mattings, carpets etc. to provide adequate absorption to reduce the reverberation time to the required value.

21.9.2 Insulation

Insulation is the stoppage of unwanted sound, originating elsewhere, from reaching a studio or an auditorium. This is achieved by reducing the intensity of unwanted sound by use of absorbing materials. In fact, sound reduction of a structure is the combined effect of reflection from the incident side and absorption within the material of the structure.

Sound reduction property, or insulation property of a material is expressed in terms of its Sound Reduction Index which is defined as the ratio of the intensity of incident sound and the intensity of sound coming out of the material

$$\text{Sound Reduction Index } (SRI) = \frac{\text{Incident sound intensity}}{\text{Reflected sound intensity}}$$

SRI is generally expressed in decibels. The reduction of sound intensity or insulation is caused by the combination of its mass and stiffness and hence its resonant properties.

All good absorbers like thick curtains over openings, jute matting and carpets on the floor, padding for mounting machines etc prevent sound from leaking into other buildings. It may be said that all good absorbers are also good insulators. Celotex is a very good insulator of sound.

Doors of all Radio Broadcasting and TV studios are specially designed to insulate the studios from all unwanted sounds from outside.

Massive and rigid walls also act as excellent insulators for structure born noises.

21.10 MICROPHONES

The microphone is a transducer or a device which converts the variations of sound pressure in a sound wave into corresponding electrical variations in an electrical circuit. The electrical variations so produced are in the AF range and are further amplified by means of an AF amplifier to make them suitable for feeding loudspeakers, recording heads and for modulating carrier waves. In a good quality microphone, the electrical changes follow the sound pressure changes exactly and truly. Different types of microphones are available for proper reproduction of various types of sounds like music, talks, drama and sports commentaries; but a microphone must fulfill certain general requirements to be really useful for the purpose for which it has been designed. These general requirements are:

1. The response of the microphone should be independent of frequency, i.e. the output should not vary much with frequency.
2. The shape or body of the microphone should be such that the frequency response is reasonably independent of the angle of incidence of sound waves.
3. It should be free from harmonics.
4. The output of the microphone should be high compared to the self-generated and thermal noise.
5. It should remain unaffected by adjacent electric and magnetic fields.
6. The mechanical construction should be robust to withstand handling in service.

Besides these general requirements, the microphones are sometimes specially designed to be either omnidirectional or to discriminate between sounds coming from different directions. For broadcasts from industrial plants, sports

meetings, races and boxing matches, the microphones used are meant to keep off the unwanted strong voices from remote sources to avoid drowning of the voice of the commentator.

For the conversion of sound energy into electrical energy, the sound waves set up mechanical vibrations in some moving element to generate small AF voltages. The voltages generated by the vibrating element may be either proportional to the velocity or the amplitude of the moving element. Microphones are, accordingly, classified as *constant velocity* microphones or *constant amplitude* microphones. Of the five types of commonly used microphones, the carbon microphone, the crystal microphone and the capacitor (condenser) microphone are constant amplitude type whereas the moving coil and ribbon type of microphones are constant velocity microphones. Constructional details and characteristics of these microphones will now be described.

21.11 TYPES OF MICROPHONES

21.11.1 Carbon Microphone

This is the type of microphone commonly used as a transmitter in telephone equipment. Constructional details of a carbon microphone are given in Fig. 21.4.

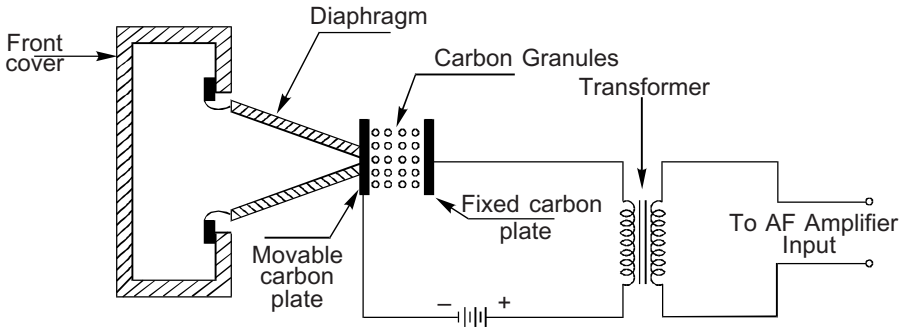


Fig. 21.4 Schematic of a Carbon Type Microphone

The operation of the microphone depends on the variations of sound pressure on the diaphragm, which in turn produces the contact pressure variations between the carbon granules which are enclosed between a movable carbon disc attached to the diaphragm and a fixed back-plate. The varying pressure on the carbon granules modulates a steady DC current that flows through the circuit due to a polarising potential of about 50 V applied from the telephone circuit. The fluctuating current varies at the AF rate and by means of a transformer these AF variations are applied to the input of an AF amplifier for further amplification, if necessary.

The carbon microphone has a limited use as a telephone transmitter because of a steady hiss generated by DC current. Its frequency response is also limited to about 3000 Hz and this makes it suitable for use in speech circuits only. It is

most suitable for telephone equipment because of its high output, simple design, low cost and durability.

21.11.2 Condenser Microphone

A condenser microphone or a capacitor microphone depends on its action on the variation of capacitance between two electrodes or conducting plates. If one of the plates is movable, sound waves striking this plate vary the capacitance between the plates and these variations are in step with the sound waves. Charging and discharging currents flow through the resistance *R* and these oscillatory currents produce oscillatory voltages which are applied to the input of the audio amplifier as shown in Fig. 21.5.

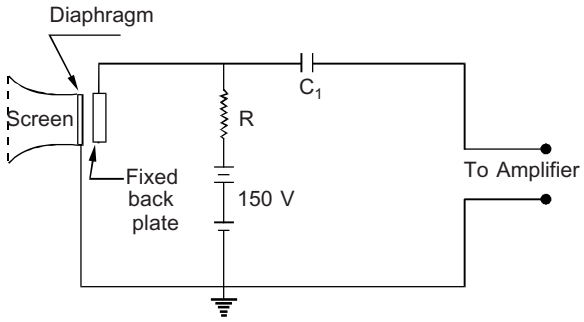


Fig. 21.5 Electrical Circuit of a Condenser Microphone

The diaphragm is usually of duralumin and may be as thin as 0.0025 cm. The separation between the diaphragm and the back plate is generally between 0.0025 and 0.005 cm. The capacitance between the plates is about 300 pF for a diaphragm of 38 cm diameter. A potential difference of about 150 V is applied between the plates from a battery through a resistance *R*.

A condenser microphone has a good frequency response from 30 Hz to about 9000 Hz. However, it has a high impedance and low output, making it necessary to use only short leads from the microphone to the amplifier. The amplifier is mounted on the microphone-housing itself, making the unit a bulky one. Lack of portability and ruggedness are the main disadvantages of a condenser microphone.

21.11.3 Crystal Microphone

A crystal microphone depends for its operation on the piezoelectric property possessed by certain crystals like Rochelle salt. When subjected to mechanical stresses like the variations of pressure in a sound wave, these crystals develop charges of opposite polarity upon their faces. The magnitude of the voltages developed depends on the displacement of the piezoelectric crystal caused by the sound waves. These voltage variations can be amplified by an AF amplifier. Besides Rochelle salt crystals which show the maximum piezoelectric

effect, certain ceramic elements like barium titanate also exhibit piezoelectric properties.

A conducting coating is applied to the two parallel surfaces (upper and lower) of the crystal, between which an emf can be produced by subjecting the crystal to the mechanical stresses. The application of an alternating force from the sound waves causes an alternating emf to be generated by the plate which is suitable as a microphone. Two such plates are cemented together to form what is called a *bimorph* element as shown in Fig. 21.6.

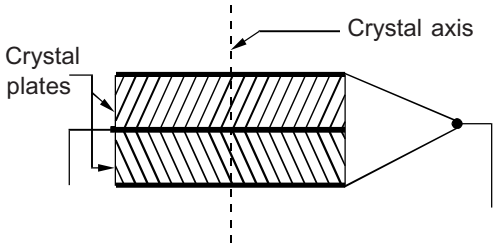


Fig. 21.6 Bimorph Element

One terminal of the element is connected to inner surfaces and the other to outer surfaces, so that the two plates are in parallel and the voltages generated are additive. The bimorph may be a bender type or a twister type which is generally used in diaphragm operated crystal microphones as shown in Fig. 21.7. The motion of the diaphragm caused by sound waves generates a voltage.

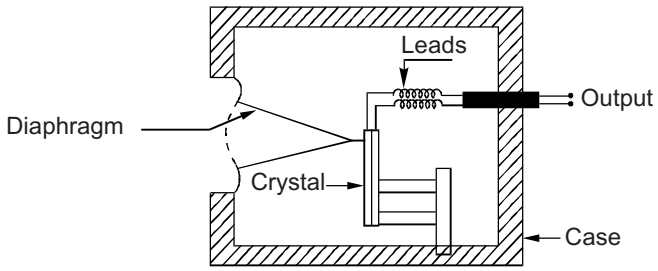


Fig. 21.7 A Crystal Microphone

The advantages of a crystal microphone are its light weight, ruggedness, easy maintenance and non-directional properties. It has a flat frequency response almost over the entire AF range extending up to 10,000 Hz.

21.11.4 Dynamic Microphone (Moving Coil)

A dynamic microphone, also called the moving coil microphone, operates on the principle of electromagnetic induction, i.e. the generation of an emf when a conductor moves in a magnetic field. In a dynamic microphone, the conductor is in the form of a coil placed between the poles of a strong permanent magnet. The coil is rigidly attached to a diaphragm made of duralumin metal and clamped firmly to the frame all round its outer edges. The coil itself is made of

a thin aluminium ribbon insulated from itself and the diaphragm by means of a suitable varnish. The construction of a dynamic microphone is shown in Fig. 21.8.

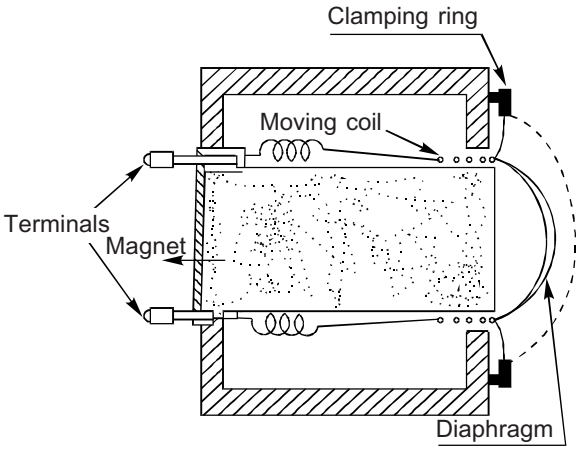


Fig. 21.8 Section of a Dynamic Microphone

When sound waves impinge on the diaphragm, it moves backward and forward moving the coil in a like manner. When the moving coil cuts the magnetic lines of force, a current is induced in the coil which, in frequency and amplitude, is a faithful replica of the diaphragm movement caused by the sound waves. The current thus produced is the output of the microphone. The output impedance of the microphone which is low (30 Ω) has to be matched to the impedance of the line feeding the amplifier (300 Ω).

A moving coil microphone has a uniform frequency response from 50 to 10,000 Hz. It does not suffer from any drawbacks like *hiss* or *pack* effect. It is robust and reliable and does not need any batteries for operation.

21.11.5 Velocity or Ribbon Microphone

The ribbon microphone operates more or less on the same principle as the moving coil microphone except that the moving coil and the diaphragm are replaced by a very light ribbon of corrugated aluminium foil. This aluminium ribbon is suspended loosely between the pole pieces of a powerful permanent magnet as shown in Fig. 21.9.

When sound waves impinge on the corrugated face of the ribbon, it vibrates back and forth and so cuts the magnetic lines of force. This causes alternating voltages to be set up across the ribbon conductor and these voltages constitute the microphone output. It is an extremely small voltage that appears across the ends of the ribbon. The problem of matching the small impedance (only 0.15 Ω) to the impedance of the microphone cable (300 Ω) is again solved by

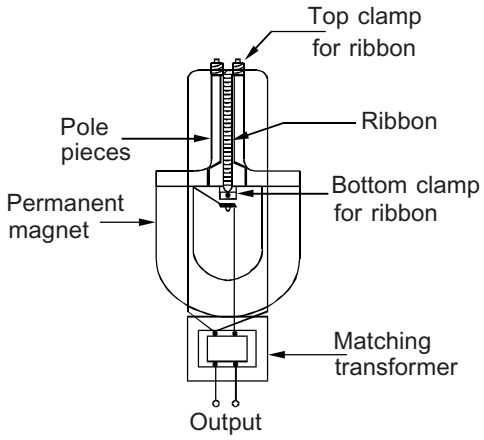


Fig. 21.9 Front View of a Ribbon Microphone

incorporating a matching transformer in the microphone itself. The reasons for using a corrugated ribbon are to increase the effective length of the ribbon for increasing impedance and to allow delicate tensioning by the spring action of the ribbon.

One important characteristic of the ribbon microphone is that it presents two faces—front and back—and is equally sensitive to sound on either face. On the other hand, the pole pieces themselves shield the ribbon from sound waves approaching from the sides. Such waves would also form equal and cancelling pressures on each face of the ribbon. The net result is that the microphone is virtually *dead* as far as sounds coming from these areas are concerned. The response of the microphone can either be represented graphically as in Fig. 21.10(a) or by a figure-of-eight polar diagram as in Fig. 21.10(b). This type of response is very useful in balancing orchestral musical programmes. The overall frequency response of the microphone is extremely good, being equally sensitive from 20 to 16,000 Hz. The microphone is always fitted with screens or covers which protect the ribbon from injury or dust.

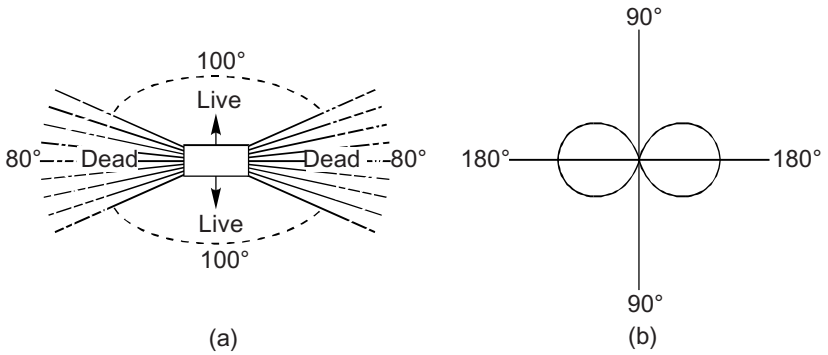


Fig. 21.10 Response of a Velocity Microphone (a) Graphical Representation of Response, (b) Polar Diagram as Figure of Eight

21.12 DIRECTIONAL PROPERTIES OF MICROPHONES

The directional properties of a microphone determine the way it responds to the sounds coming from different directions. These properties are determined by the *polar diagrams* which are graphical representations of the sensitivity of the microphone in different directions. The design of the microphone casing plays an important part in this case.

21.12.1 Omnidirectional Microphones

A simplified 'omni' microphone is shown in Fig. 21.11.

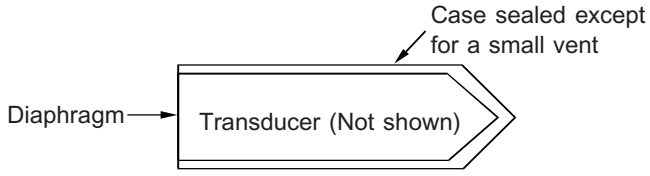


Fig. 21.11

The case is sealed except for a small vent to allow for temperature and atmospheric pressure changes. The transducer which is enclosed is not shown. The only way in which sound waves can get to the diaphragm is from the front. In other types of microphones the sound is allowed to reach both the front and the back of the diaphragm. If the wavelength of sound is greater than the microphone, we can assume that, because of diffraction (bending round the corners), sound waves will strike the diaphragm no matter from which direction they arrive. The microphone will then have an omnidirectional response which is a circle as shown in Fig. 21.12.

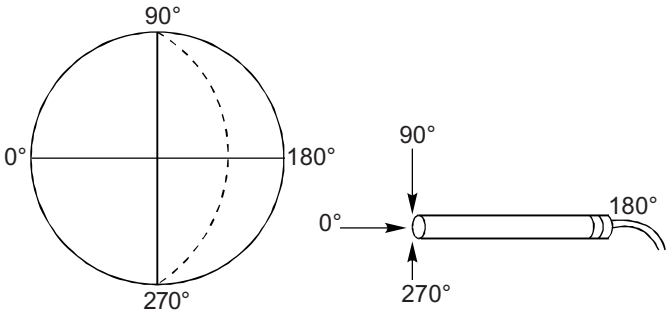


Fig. 21.12 Polar Diagram for an Omnidirectional Microphone

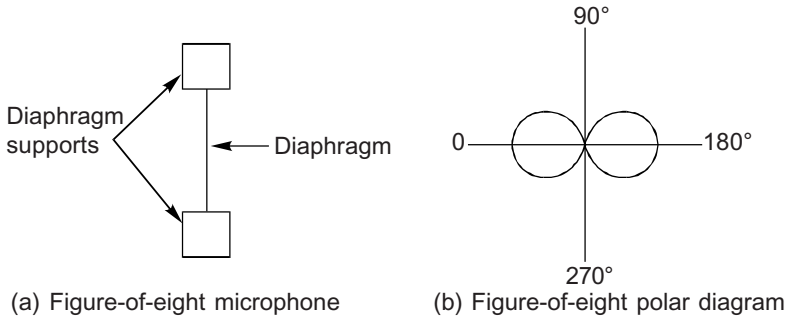
If the sound wavelength is smaller than the microphone diameter, then the diffraction process won't occur completely and waves from the back of the microphone will either not reach the diaphragm or, if they do, only partially. The microphone is not truly omnidirectional as shown by the dotted curve in Fig. 21.12.

Omnidirectional microphones have the following special features:

1. Most 'personal' micrphones those clipped to lapels, are omnis.
2. They are useful for hand holding for interviews where they do not have to be pointed accurately at the speaker.
3. Compared to other microphones, they tend to be less prone to 'rumble' and also to effect of noise. A wind shield is necessary when used outdoors.

21.12.2 Figure-of-Eight microphones

Figure 21.13(a) shows a microphone in which sound waves can reach both sides of the diaphragm.



(a) Figure-of-eight microphone (b) Figure-of-eight polar diagram

Fig. 21.13 Figure-of-eight Microphones

Sound waves arriving at the diaphragm from the left will have the same effect as those arriving from the right. Those which come from the top or bottom will pass the diaphragm with the same pressure on both sides of it, and there will be no resultant pressure on the diaphragm. It is obvious that the microphone will have a polar diagram which is (a) symmetrical front to rear (b) shows no output from sounds at right angles to front-rear axis. In fact this polar diagram is called figure of eight or bidirectional polar diagram as shown in Fig. 21.13(b).

A ribbon microphone described earlier has a figure of eight polar diagram. Such microphones are also called pressure gradient microphones.

Facts about pressure gradient microphones:

1. Ribbon microphones lend themselves to figure of eight operation.
2. Electrostatic microphones (condenser microphones) can be made figure of eight microphone (see variable polar diagram microphones).
3. Pressure-gradients microphones tend to be severely affected by vibrations, rumble, etc.
4. A figure-of-eight microphone can be very useful in discrimination against unwanted sounds.
5. The response of a figure of eight microphone can be expressed by the equation.

$$v = \cos \theta$$

where v is the distance from the origin at an angle θ.

21.12.3 Cardiod Microphones

A cardiod microphone is a combination of an omnidirectional microphone and a figure of eight microphone. The polar diagram shown in Fig. 21.14 is cardiod in shape which is a Greek word meaning 'heart' and hence the word 'cardiac'

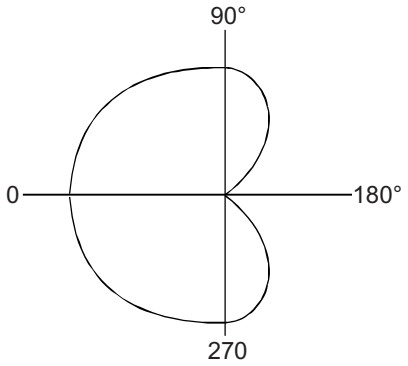


Fig. 21.14 Cardiod Polar diagram

or 'cardiagrams', etc. Mathematically, the expression for the cardiod polar diagram can be written as

$$v = 1 + \cos\ \theta \tag{21.1}$$

where v is the distance from the origin to the curve at an angle θ.

Equation 21.1 can be arrived at as follows

Equation for an omnidirectional microphone is $v = 1$ (21.2)

and for figure of eight microphone we have

$$v = \cos\ \theta \tag{21.3}$$

Adding (21.2) and (21.3) we have

$$2v = 1 + \cos\ \theta \tag{21.4}$$

which is the same as given in equation (21.1) (v or $2v$ does not alter the shape of the curve.

This only means that a cardiod microphone could be produced by combining an omni microphone and a figure-of-eight microphone as shown in Fig. 21.15.

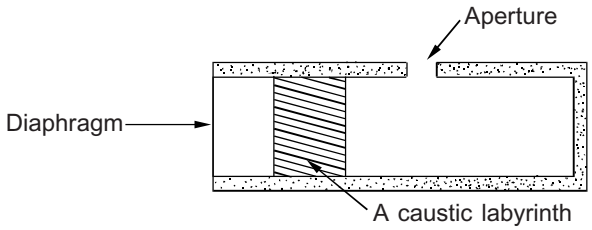


Fig. 21.15 Construction of a Cardiod Microphone

Also, because, a cardiod is part omni, part figure-of-eight it is obvious that this polar diagram can be obtained by allowing some sound to reach the rear of the microphone remembering the fact that in an omni no sound reaches the rear while in a figure of eight there is unrestricted access to the rear.

The labyrinth causes some delay in the sound reaching the rear of the diaphragm resulting in some phase-shift. The microphone is also an *acoustic phase-shifting network* and cardiod microphones using this principle are also sometimes called "phase shift cardiods". Facts about cardiod microphone:-

1. Because of their having partly pressure gradient operation, they show some bass top up.
2. They are prone to vibrations in sound.
3. Good wind shield is essential for outdoor operation and in some microphones this is built-in.
4. The 'dead' area at the back can be useful in discriminating against unwanted sounds and noises.

21.12.4 Hypercardiod Microphones

The polar diagram of a supercardiod or hypercardiod microphone is a mix between a cardiod and a figure-of-eight polar diagram which shows nulls at 45° off the rear axis. Many hypercardiods are constructed as a kind of cardiod but with greater sound wave access to the rear of the diaphragm.

Special features of hypercardiods:

1. It is sometimes useful to have two dead sides near the rear of the microphone.
2. The front lobe is slightly narrower than the lobe of a cardiod-this again is sometimes helpful.
3. Like other pressure gradient operated microphones hypercardiods tend to show bass tip off effect and are prone to pick up rumble and vibrations.

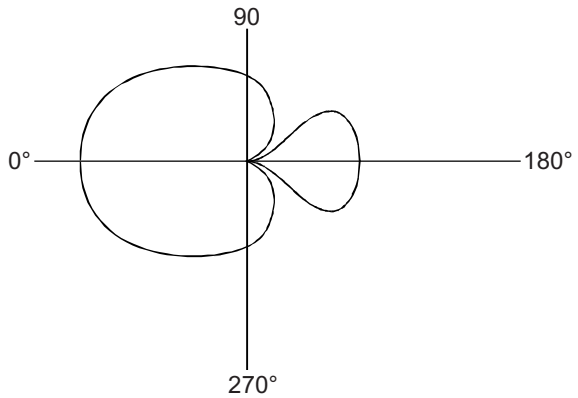


Fig. 21.16 Hypercardiod Polar Diagram

21.13 OTHER SPECIAL TYPES OF MICROPHONES

21.13.1 Gun Microphones

The main characteristics of these microphones are that they are very directional in their response. There are two types of gun microphones differing only in their lengths.

The true gun microphones are really metal tubes more than half a meter in length and with a diameter of about 2 cm. The tube is perforated or slotted along much of its length as shown in Fig. 21.17. The differences in path length travelled by waves entering different holes along the length result in cancellation effect.

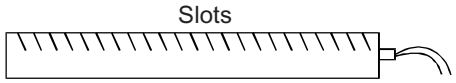


Fig. 21.17 Gun Microphone

Gun microphones of practical size are only very directional at frequencies above 3–4 kHz where the angle of pick up is about ± 15°.

The shorter gun microphones where the slotted part is some 20-25 cm long, have much wider angle of pick up. They behave as hypercardiods up to about 2 kHz.

These microphones are very extensively used for outside work such as news gathering, location drama and so on.

21.13.2 Electret Microphone

Electret microphone is a simpler version of the professional capacitor microphone.

A conventional electrostatic microphone needs a polarizing voltage of 50 V or more applied between the diaphragm and the back plate. This means that the power supply source needs to provide a low voltage supply for the amplifier and a relatively high voltage for the microphone capsule. However, certain materials exist which can be given a permanent electrostatic charge and they are called *electrets* (like magnets) so by using an electret material for either the diaphragm or the back plate, the need for a polarizing voltage is removed but there still remains the need for a low voltage supply for the amplifier. In all other respects an electret microphone behaves like a capacitor or condenser microphone described earlier.

21.13.3 Radio Microphone

A radio microphone is any high quality microphone used in combination with a battery operated transmitter which is either carried in the pocket or concealed somehow under the clothing. Usually the microphone is of personal type (say tie clip type) but given the right connections any kind of microphone can be used. The transmitter operates in the VHF range and is frequency modulated.

The receiver can be put in a convenient portion and its output may be fed into a studio microphone socket. The range of operation can be anything from a few metres at worst to perhaps 50 m at best. 200-400 m is achievable under suitable conditions in outdoor working. Reflections of radio signals in outdoor conditions can reduce the working range.

The receiver can be put in a convenient portion and its output fed into a studio microphone socket.

It is useful for in door and outdoor sports events.

21.13.4 Tie-Clip Microphone

These small microphones are fitted to the tie or lapel of the speaker by a clip. For small weight and size these are generally either capacitor or electret type units in which case they are provided by small button cells such as used in hearing aids and electronic watches. Most tie-clip microphones are omnidirectional and not suitable for PA systems.

21.13.5 Lavalier Microphone

This is a small moving coil type microphone which is hung round the neck of the speaker by means of a cord or a ribbon. It has its applications where mobility is important as in the case of lectures.

21.13.6 Noise Cancelling Microphones

These are used for announcements in areas of high ambient noise. They consist of two cardiods or hypercardiod microphones spaced a few inches apart and connected in anti-phase, speech is directed at one of the units but is also picked up by the other, hence producing some cancellation. As the first unit is nearer the announcer's lips than the second, there is difference in sound pressure level, and part cancellation occurs. Ambient sound coming from a distance arrives as a virtual plane-wave and affects both units equally cancellation is therefore total as shown in Fig. 21.18.

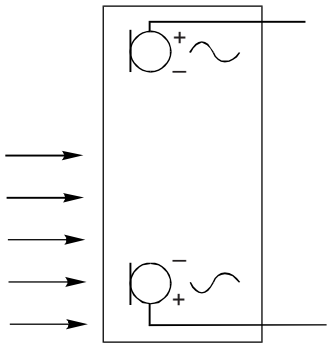


Fig. 21.18 Noise Cancelling Microphone

21.13.7 Boundry Microphones (Pressure Zone Microphones)

These are only small high quality units, often electrostatic capsules, mounted close to a surface. In some cases this surface is a small metal plate, 15-20 cms square. In others the microphone is fitted into a circular piece of wood. The latter arrangement is useful for discussions where satisfactory recordings can be made of the participants gathered around the microphone which can be on the floor. The boundary microphone offers many advantages over conventional instruments although it may not be an answer to all microphone problems but being of recent origin, there is much to learn about its most effective use.

21.14 PARAMETERS AND SPECIFICATIONS OF A MICROPHONE

Having considered the properties of various types of microphones, we shall now describe the main parameters and specifications used to describe a microphone. The requisites of a microphone are judged by its

1. sensitivity 2. impedance 3. frequency response 4. noise. 5. directivity.

These requisites will now be discussed in brief.

21.14.1 Sensitivity

There are several ways of expressing the sensitivity of a microphone. All relate to the electrical output to a given sound pressure. In other words it is the output that results from a given sound pressure level at the diaphragm. The unit is the mV/pascal or its equivalent the mV/(newton/m^2). In some cases the output in decibels below a reference of 1 volt is given.

Microphone sensitivities in some typical cases are given below

	Microphone type	*Sensitivity dB relative to 1V/(N/m^2)*
1.	Typical electrostatic microphones	– 20 to – 45
2.	Typical moving coil microphones	– 50 to – 55
3.	Some ribbon microphones	– 60 to – 65

To take impedance into account, an output power rating is often used instead of voltage. The usual reference standard is the milliwatt into 600 ohms. Thus, the sensitivity of a microphone can be expressed as – 60 dB where 0 dB = 1mW into 600 ohms.

21.14.2 Impedance

Microphone impedances fall within the values 30–50 Ω; 200 Ω; 600 Ω, 1000 Ω or 47 kΩ. The suitablity depends on the application. One of the determining factors is the microphone cable capacitance. Its reactance decreases with cable length and increasing frequency. So long runs of cable impose a low value shunt across the output of the microphone which decreases further as the frequency rises. So to maintain the maximum high frequency response with long cables low impedance is necessary and is commonly used in recording studio.

High impedance while offering a high signal voltage results in noticeable loss of treble (high notes) even over moderate cable lengths.

21.14.3 Frequency Response

Frequency response of a good microphone is the output of the microphone within ± 1 dB of the output at 1000 Hz. Although the audio frequency range extends from 20 Hz to 20 kHz, however a microphone which gives a flat response without ± 1 dB for frequency between 80 Hz to 8 kHz is quite acceptable for a normal programme but a better system is required for high fidelity work.

From design considerations, a microphone is considered to be a vibrating system consisting of electrical induction and capacitance corresponding to the mass and compliance of the vibrating system. Thus, the system has a natural resonant frequency at which the output of the microphone is suddenly boosted up thereby adversely affecting the frequency response of the microphone. The vibrating system is so designed that the natural resonant frequency falls outside the normal operating range of the microphone.

Frequency response of some of the commonly used microphones is given below.

Types of microphone	Frequency response within ± 3 dB
Moving coil	40 to 10,000 Hz
Ribbon type	16 to 15,000 Hz
Crystal	50 to 10,000 Hz
Condenser	20 to 20,000 Hz
Carbon	100 to 7000 Hz

21.14.4 Noise

Besides the external noise picked up by a microphone some noise is also generated inside the microphone due to resistance of the circuit components. This is quoted as signal to noise ratio in dB by the manufacturers along with other parameters of the microphone.

21.14.5 Directivity

Directivity of a microphone is determined by the way it responds to sound coming from different directions. Directivity has already been fully discussed in para 21.12 under the directional properties of a microphone.

21.15 MICROPHONE MIXERS

In broadcasting equipments or in PA systems it becomes necessary to mix the outputs of two or more microphones. The simplest method is to put the output wires in parallel and apply the combined output to the input of the amplifier. A changeover switch can be used for selecting or cutting out the output of any particular microphone. If, however, outputs are to be mixed in a gradual or controllable manner, use of more complicated mixing circuits is resorted to.

Such mixing circuits are called *faders* or *fade units* because the volume of sound can be faded in or faded out by means of these circuits.

The simplest fader would be a variable resistor introduced in series with one of the output wires as shown in Fig. 21.19.

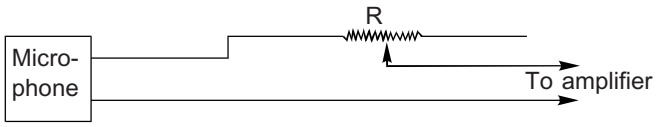


Fig. 21.19 Fader for a Single Microphone

This method could be extended to a mixing unit for more than one microphone as in Fig. 21.20.

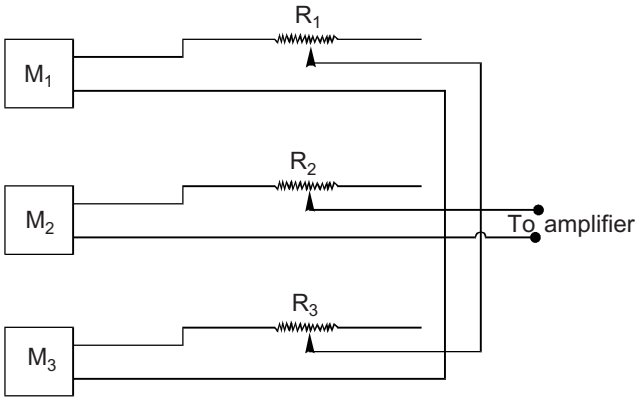


Fig. 21.20 3-channel Fader Unit

Two important requirements of a good mixing circuit are (i) the question of unbalance introduced by using a variable resistance in one leg, and (ii) the impedance matching. The problem of unbalance is solved by introducing a variable resistance in each leg of the microphone as in Fig. 21.21. When the knob of the fader unit is turned, both the sliding points move together. A high

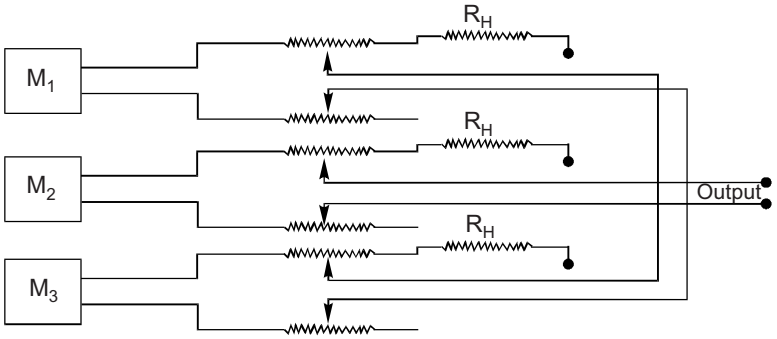


Fig. 21.21 3-channel Balanced Series Fader Unit

resistance R_H of about 20 MΩ is included between the last stud of the variable resistance and the off stud of one leg of each microphone. This resistance serves as a static leak and prevents a click being produced due to a sharp flow of current whenever the microphone is faded up.

The matching problem is solved by resistance networks called *attenuators* which provide constant input and output impedance at all positions of the mixer control. The input impedance is the impedance of the microphone and the output impedance of the attenuator is the input impedance of the amplifier. An attenuator of the T-type commonly used in mixer units is shown in Fig. 21.22.

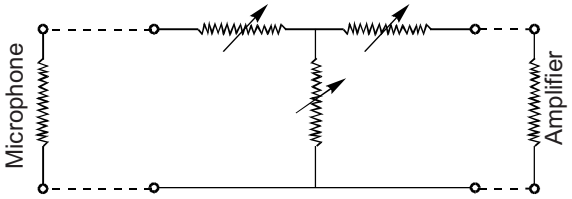


Fig. 21.22 T-type Attenuator

21.16 PICK-UP HEADS

A pick-up head or a phono pick-up is an electrical device for reproducing the sound recorded on a disc. A needle or a stylus attached to the pick-up follows the groove variations on the recorded disc and the mechanical variations of the stylus are converted into electrical variations which correspond in frequency and amplitude to the audio frequencies recorded on the disc. The AF voltages so generated by the mechanical movement of the stylus are amplified and reproduced by a loudspeaker as in the case of microphones already described.

Two types of pick-ups are commonly used in disc playback equipment. These are the magnetic type and the crystal or piezoelectric type.

21.16.1 Magnetic Pick-ups

This type of pick-up is based on the principle of electromagnetic induction. Change of magnetic flux produced by the movement of the pick-up needle as it follows the groove variations induces in a coil a voltage which corresponds in amplitude and frequency to the groove variations. Magnetic pick-ups are also of two types. In the *variable reluctance* type, the stylus is fixed on a magnetic armature which moves freely in the air-gap between the pole pieces of a strong permanent magnet as shown in Fig. 21.23.

As the magnetic armature is deflected from side-to-side, the magnetic flux through the coils wound on the pole pieces changes. When the magnetic flux increases on one coil it decreases on the other coil. Voltage proportional to the difference of the magnetic flux are induced by this push-pull action in the two coils which are connected in series. The emf generated is a few millivolts and the frequency response is uniform between 20 and 20,000 Hz.

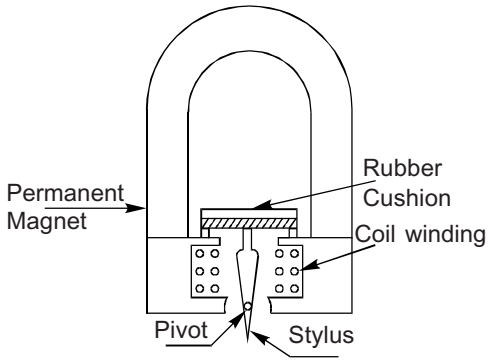


Fig. 21.23 Magnetic Pick-up

The other type of magnetic pick-up is the moving coil or the dynamic type which works on the same principle as the dynamic microphone. A light coil wound round the armature is caused to move from side to side as the stylus or the needle traces the wavy track on the disc. The number of magnetic lines of force threading the coil varies with the movement of the coil and small alternating currents are induced in the coil which may be amplified in the normal way and then fed to a loudspeaker. Since the moving system is very light, the frequency response is flat from 10 to 30,000 Hz. This is a high quality pick-up used in quality broadcasting and other hi-fi systems.

The magnetic pick-ups are fitted with sapphire or diamond stylus to increase the playing time and to reduce wear on the record groove which results in scratch and noise. The output of magnetic pick-ups being low, they require a preamplifier in reproducing an audio chain.

21.16.2 Crystal Pick-up

A crystal pick-up is based on the piezoelectric principle and makes use of Rochelle salt crystals for the piezoelectric bimorph elements to convert mechanical motion of the needle into electrical currents. When the needle follows the lateral motion of the groove modulations, the crystal faces get twisted and generate voltages proportional to the amount of deformation. The output of a crystal pick-up is high (1 to 2 V) and, therefore, no preamplifier is necessary in the playback chain. The impedance of the pick-up is high and it can be directly connected to the grid or base of the first stage in the audio amplifier. The response is not as good as that of a magentic pick-up (30 to 8,000 Hz) but because of the simplicity of construction and low cost, this is a popular type of pick-up used in normal playback systems. The crystal pick-ups are adversely affected by warm or damp climates and the crystal is damaged by high temperatures.

Modern pick-ups are available in sealed units called *cartridges.* A cartridge may have a single fixed stylus which can be used for the playback of only one type of disc which may be either 78 rmp or an LP record. Cartridges with two

stylus tips allow either of the stylus tips to be brought into position by rotating a lever arm fixed on the cartridge itself. The stylus with the thicker tip is for 78 rpm records and the stylus with lesser tip thickness is meant for the playback of 45 rpm and $33\frac{1}{3}$ rpm LP records.

21.17 TONE ARM

A tone arm not only holds the pick-up cartridge in position but it also guides the cartridge so as to keep it always tangential to the groove even with changing diameter as the arm moves across the disc. This is an extremely difficult requirement as the tone arm generally turns on a pivot and does not move radially in a straight line. Any departure from the tangential position of the groove at the point of contact is called *tracking error*. The tracking error must be reduced to a minimum by suitably designing the length, shape and angle of the tone arm with regard to its anchoring point on the turn table. This is necessary to reduce uneven pressure on the sides of the groove which geneally results in damage to the record.

Another important factor in the design of a tone arm is the tracking pressure exerted on the stylus by the combined weight of the tone arm and the pickup cartridge. The vertical weight on the record is reduced by counter balancing the tone-arm weight by fixed or sliding counterweights provided on the tone arm. Light tone arms prolong the life of a record and also give better reproduction quality. Long playing records require less stylus pressure than the 78 rpm records. Moreover, the design of a good tone arm makes it free from resonant mechanical vibrations so that it responds to only the groove modulations. Figure 21.24 shows the design and mounting of a modern tone arm.

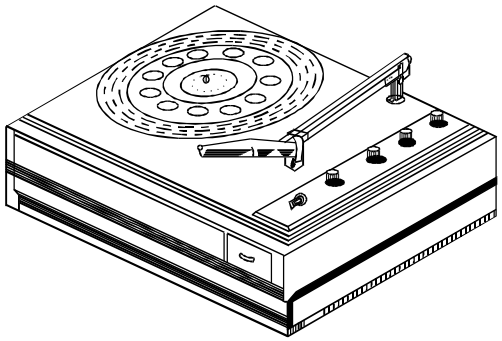


Fig. 21.24 Tone Arm Length Shape and Mounting to Reduce Tracking error

21.18 LOUDSPEAKERS

A loudspeaker is a device which converts electrical energy into sound energy. Audio frequency currents from an AF amplifier are converted into sound waves of corresponding frequency and amplitude by the loudspeaker. In principle, a loudspeaker is the converse of a microphone and the two instruments perform

complementary functions in a sound reproducing system. The constructional details in these two cases are so much similar that a loudspeaker can be regarded as a loudspeaking microphone.

A loudspeaker is the voice of any electronic entertainment equipment and, as such, it should be able to reproduce, as faithfully as possible, the original sound from the broadcasting studios. A good loudspeaker should be able to reproduce all sounds equally well irrespective of their amplitude frequency and waveform.

Sound waves are produced in air by a vibrating body. In the case of a loudspeaker, the vibrating body is a cone or a diaphragm which is attached to a driving unit which converts electrical currents into mechanical motion for the diaphragm to vibrate and produce sound waves containing the acoustical energy. The principle employed for converting electrical energy into mechanical motion is different in different types of loudspeakers. Based on the construction of the driving unit employed, the loudspeakers can be classified as magnetic, piezoelectric (crystal), electrostatic and dynamic loudspeakers. However, the type of loudspeaker most commonly used in radio, TV and other electronic equipment is the dynamic type and only this type of loudspeaker will be described in detail.

21.18.1 Dynamic (Moving Coil) Speakers

The principle of operation of a dynamic or moving coil loudspeaker is similar to that of an electric motor. The audio currents passing through the voice coil placed in a strong magnetic field produce a vibrating motion in the voice coil. This to-and-fro motion of the voice coil is transferred to the cone or diaphragm attached to the voice coil. The vibratory motion of the cone produces sound waves in the air which reaching the eardrum, produce the sensation of sound. Figure 21.25 shows the constructional details and main parts of a dynamic loudspeaker. It consists of the following parts.

Magnet A strong magnetic field is necessary for satisfactory operation of a dynamic loudspeaker. This magnetic field can be obtained by passing direct current through a field coil wound over soft iron pole pieces. Such a speaker using an electromanget for producing the magnetic field is known as an electrodynamic speaker. However very strong permanent magnets can now be made out of an alloy of aluminium, nickel and cobalt called *alnico* and the use of electromagnets with field coils has become unnecessary in the manufacture of loudspeakers. Speakers using permanent magnets are called PM speakers and this is the type used in almost all radio and TV sets. A very strong magnetic field exists in the annular air space between the outer pole piece formed by the yoke and the central pole piece. The voice coil is suspended in this stationary magnetic field.

Voice Coil This consists of a few turns of wire wound on a bakelite, mica or aluminium cylinder. It is suspended between the magnetic poles and held in position by means of a special device called a *spider.*

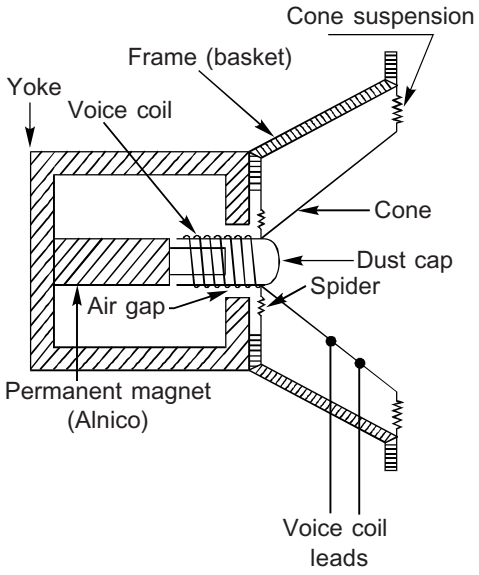


Fig. 21.25 A Typical Moving Coil (Dynamic) Loudspeaker

Spider This is a flexible membrane which is fixed on the coil form on one side and the speaker frame on the other side. The purpose of the spider is to keep the voice coil properly centered in the air gap betwen the pole pieces without allowing it to touch the pole pieces. The spring action of the corrugations on the spider brings the coil back to its normal position at the end of each back and forth motion.

Cone The cone or diaphragm of a loudspeaker is made of paper prepared by a special filtering process and moulded in the form of a cone of the required size. Corrugations provided near the outer rim permit the cone to move easily and over greater distances. The outer edge of the cone is fixed or cemented to the speaker frame and the inner edge is attached to the coil form. The cone is the vibrating element in a speaker which produces sound waves. The acoustical energy produced depends on the shape and area of the cone. Elliptical cones which give larger area and produce sharper directivity patterns for sound are sometimes used in preference to circular cones.

A dynamic speaker is the simplest of all types of loudspeakers. It can be made mechanically strong and in larger sizes to give more sound output at a lower cost than other types.

21.18.2 Loudspeaker Impedance

The voice coil of a dynamic loudspeaker has a certain amout of impedance depending on the size of the wire and the number of turns used for winding the voice coil. When wound with many turns of fine wire, the loudspeakers are said to possess high impedance and, if only a few turns of thicker wire are used for the moving coil, the loudspeaker coil is termed as a low impedance coil.

The actual impedance of a voice coil varies with frequency but for reference purposes the impedance mentioned is the effective impedance at 1000 Hz. In order to feed maximum undistorted power from the output stage of the audio amplifier into the speaker, the loudspeaker coil impedance must be perfectly matched to the output impedance of the output stage in the audio amplifier. The use of matching transformers for matching the output impedance of the amplifier to the loudspeaker impedance has already been explained in Ch. 4. Improper matching of the loudspeaker impedance not only results in loss of power but also affects the frequency response.

When a group of loudspeakers is to be fed from the same amplifier, the speakers are arranged in series-parallel arrangement as shown in Fig. 21.26 so that the effective impedance presented by the entire group is the same for which the matching transformer has been designed. For satisfactory results the impedance of all groups in parallel should be equal and the impedance of each speaker in any series leg should be equal. Moreover, for all loudspeakers to receive the same fraction of the total power, their effective impedance must be equal. If these conditions are not fulfilled, some speakers will have high pitched tones and other will have low pitched tones. If speakers with different impedances are used, each speaker will have its own matching transformer so that the total effective load presented to the load-end is the optimum load for the amplifier. Matching transformers providing tappings with different transformer ratios are available for two or more speakers of different impedances to be connected to the same transformer.

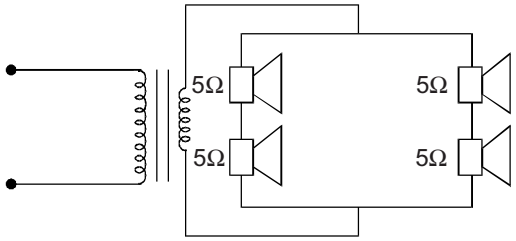


Fig. 21.26 Impedance Matching of Groups of Loudspeaker

21.18.3 Loudspeaker Baffles and Enclosures

Cone type loudspeakers are generally mounted on flat wooden board called *baffles* or enclosed in wooden cabinets of suitable dimensions called *enclosures*.

As will be clear from the technical reasoning given, the use of baffles and enclosures for loudspeakers is not merely to provide a mounting or decoration for the speaker but the use of a proper baffle increases the efficiency and improves the performance of the speaker, particularly at low frequencies.

In a loudspeaker, sound waves are produced as a result of the vibrations of the cone. When the cone moves forward, it compresses the air in front

producing a high pressure area (compression), and a low pressure area (rarefaction) is produced at the back. There is a natural tendency on the part of the compressed air in front to move around to the back position of the cone-equalise the pressure thereby preventing the formation of sound waves. This reduces the output of the speaker. One simple method of reducing the pressure neutralising effect is to increase the distance of air travel between the front and back of the cone by mounting it on a baffle as in Fig. 21.27(a). A relatively large baffle must be used to obtain adequate low frequency response. To obtain a good response down to 60 Hz requires a 137 × 137 cm (baffle) whereas a baffle 68.5 × 68.5 cm will provide a good response down to 120 Hz only.

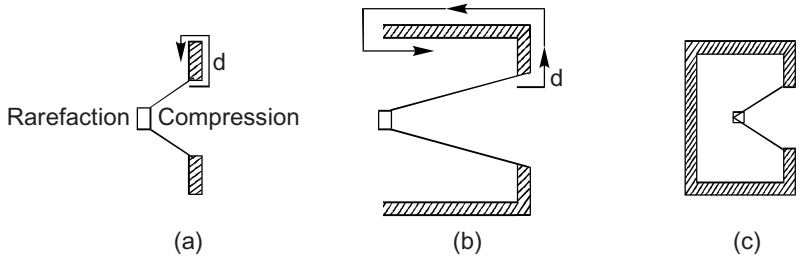


Fig. 21.27 Baffles for Loudspeakers (a) Straight Baffle, (b) Box Type Baffle, (c) Cabinet with Rear Closed

A more effective method of making a baffle without unduly increasing the size of the board is to fold the sides and make it in the form of a box with the back side open as in Fig. 21.27(b). The use of a cabinet with the rear closed, as in Fig. 21.27(c), provides almost an infinite front-to-rear path length for the sound waves. Excellent low frequency response may be obtained with a relatively small loudspeaker mounted in a small cabinet. However, the power handling capacity of such a loudspeaker is limited by the permissible excursion of the speaker cone.

21.18.4 Dual Loudspeaker Systems

A single cone type speaker is not able to provide uniform response and adequate output power over the entire AF range. A loudspeaker mechanism with a heavy and large diameter called *woofer* can reproduce low frequencies efficiently but fails to reproduce the higher frequencies properly. On the other hand a speaker with a light and small diameter cone and known as *tweeter* performs much better at the high frequency range of the audio frequencies. A combination of a woofer and a tweeter is, therefore, commonly used in radio, TV and other sound reproducing systesm to improve the quality of reproduction by covering as large a portion of the AF spectrum as possible by a dual system of loudspeakers.

For the proper functioning of a dual speaker system, it is necessary that the frequency range to be covered by the combination of speakers should be split into two ranges at a frequency called the *cross-over* frequency. A filter network

called a cross-over network is introduced in the output of the audio amplifier to divert the low frequencies to the woofer and the high frequencies to the tweeter. A dual loudspeaker system with a two way cross-over network consisting of inductances and capacitors is shown in Fig. 21.28. The values of the inductance L and the capacitance C can be calculated in terms of the loudspeaker impedance R and the cross-over frequency f by the following formulas:

$$L = \frac{R\sqrt{2}}{2\pi f}$$

$$C = \frac{\sqrt{2}}{4\pi f R}$$

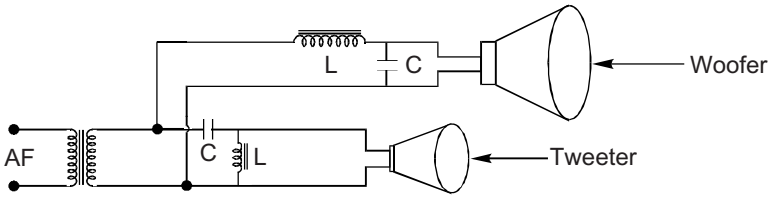


Fig. 21.28 Two-way Cross-over Network for Speakers

The large speaker for low frequencies and the small speaker for high frequencies are either mounted separately in the cabinet or these can be combined into a single coaxial unit in which the woofer and the tweeter are mounted coaxially in the same line. In certain cases, the same driver unit is made to operate two diaphragms of different sizes, one for high frequencies and the other for low frequencies. A multiple speaker system consisting of a woofer, a tweeter and a medium frequency range diaphragm with suitable cross-over network is sometimes employed for a uniform response over a much wider range of audio frequencies.

21.18.5 Horn Loudspeakers

Large-scale reproduction of sound involving several acoustical watts requires high power audio amplifiers which are very costly. The amplifier output can be reduced to a minimum by the use of horn loudspeakers which are high efficiency loudspeakers. The efficiency of a loudspeaker is the ratio of sound power output to the electrical power input. Horn loudspeakers are capable of providing an over-all efficiency of 50% compared to the cone type or direct radiator type loudspeakers in which case the efficiency seldom exceeds 5%.

A horn loudspeaker is shown in Fig. 21.29(a). It consists of an electrically driven diaphragm coupled to a horn. The principal virtue of a horn rests in its ability of presenting practically any value of acoustical resistance to the driving system thereby providing suitable matching conditions for the conversion of electrical energy into sound energy. A small diaphragm is able to produce a large volume of sound. For a good frequency response at low frequencies, the size of the horn becomes inconveniently large. But for its large size, a horn

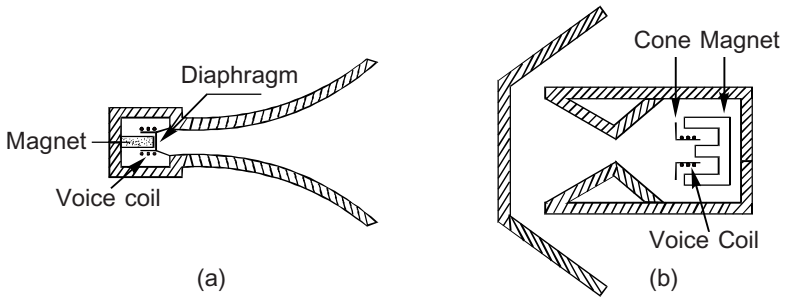


Fig 21.29 Horn Loudspeakers (a) Exponential Horn, (b) Folded Horn

loudspeaker gives a much better performance than a cone type loudspeaker. To economise in space, folded exponential horn type speakers shown in Fig. 21.29(b) are often used.

21.18.6 Column Loudspeakers

A column loudspeaker consists of a number of small direct radiator dynamic loudspeakers mounted in a cabinet as shown in Fig. 21.30.

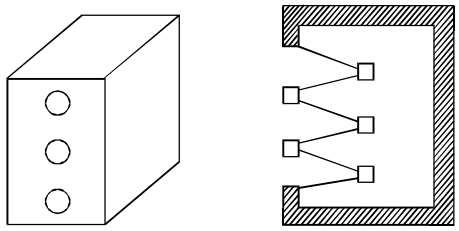


Fig. 21.30 Column Loudspeakers

These loudspeakers are capable of producing a narrow directivity pattern in the vertical plane and are useful in halls and auditoria where the purpose is to direct sound towards the audience rather than towards the walls and ceiling. In fact, column loudspeakers are useful in all cases of sound reinforcement where the sound must be confined to a narrow beam in the vertical plane in order to increase the ratio of direct to reflected sound. A uniform directivity pattern can be obtained in the vertical plane by employing a long column for the low frequency range, a medium length column for the mid-frequency range and a short column for the high frequency range with suitable electrical cross-over networks.

21.19 PUBLIC ADDRESS SYSTEM

21.19.1 What is a Public Address System?

A Public Address system or PA system for short, is a system for amplifying speech so that it can be comfortably heard by larger gatherings and at larger distances. If every one in the audience can hear the speech comfortably without

being aware of the PA system in use, the installation can be claimed to be quite successful.

21.19.2 Importance of a Public Address System

The importance of a good PA system can be judged from the following story.

The original form of a well known message to be relayed read "Send reinforcements, I am going to advance". Unfortunately when relayed through Military personnel in trenches during World War I, due the human fallibility throughout the relay stages, the resultant received message read "Send three and four pence, I am going to a dance".

Irrespective of the truth of the story, it does highlight some of the basic problems associated with any communication system particularly the one connected with human users.

21.19.3 Basic Requirements of A PA System

A good PA system should satisfy the following basic requirements:

1. *Comfortable level* The sound must be at a sufficiently high level or volume so that it can be heard comfortably at all parts of the hall or auditorium by every one with reasonable hearing power. However, volume should not be achieved at the expense of clarity.
2. *Intelligibility* The purpose of a PA system is to provide a good communication link between the speaker and his audience. Nothing is accomplished by a system that enables the audience to hear the sound without understanding what he/she is saying.
3. *Naturalness* The audience should just hear what appears to be the natural voice of the speaker without being aware of the equipment in use. However naturalness should not be attained at the cost of the other two requirements mentioned above.
4. *Reliability* The equipment should be thoroughly reliable particularly when large installations are involved. A breakdown during an important meeting or speech can be disastrous.

   Let us discuss these points in detail. Figure 21.31 gives a block diagram of a simple PA system.

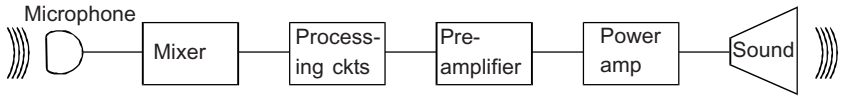


Fig 21.31 A Basic PA System

The microphone converts sound signals into equivalent audio signals which are based on a number of microphones or turntables/taperecorders to be applied to the inputs of the mixer. The mixer isolates different channels and mixes them before applying to the amplifier stages. The processing circuits contain master control, tone controls and graphic equalizers etc. The preamplifier is a

voltage amplifier which amplifies the weak input signal to a sufficiently high level to be able to drive the power amplifier. Power amplifier is the output amplifier to drive one or more loudspeakers which convert the electrical power back to sound waves of the required frequency and sound level.

21.20 HOWLROUND (FEEDBACK)

In an effort to increase the sound level we come across one of the worst problems in a PA system which is the problem of howl round or 'feedback'. This is the bugbear of all PA systems and results from repeated feedback between the loudspeaker and the microphone. Any random noise issuing from the loudspeaker is picked up by the microphone and the same is amplified and fed to the loudspeaker which reproduces it louder than before. Again it is picked up by the microphone, amplified to be reproduced louder still by the loudspeaker. This process continues repeatedly till the whole system bursts into oscillations and produces screeching sounds and whistles etc called howlround or feedback.

Figure 21.32 illustrates the process of howlround in a studio or an auditorium.

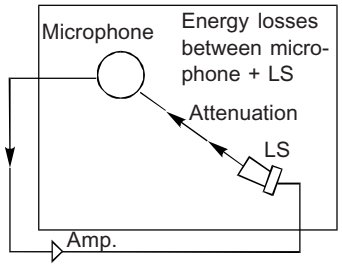


Fig 21.32 Howlround

The microphone, amplifier, room and the loudspeaker form a loop. The room can be thought of as an attenuator simply because of the reduction in sound wave intensity from loudspeaker to the microphone. If the overall gain in the loop is less than unity then the signal energy will die away. If, however, it is greater than unity then the oscillations are most likely to occur. It would appear to be a simple matter to make sure that the amplifer gain is set so that the loop gain never exceeds unity but the whole process is linked with the frequency responses of the various components of the loop viz. the room, the microphone and the loudspeaker.

Most modern microphones and loudspeakers have a fairly steady response but the studios, halls and auditoriums have an uneven response (± 10 dB or more) resulting in formation of standing waves. It would appear that at some frequency F in the combined response of room, loudspeaker and microphone results in several decibels of extra gain, say around 12-15 dB. This means that the loop gain is much higher than the mean gain and it is at this frequency that

howlround will occour first. Moving the microphone, which is the most portable part of the loop, may probably result in a change in the howl round frequency. Other factors for reducing the howl round are given below:

21.20.1 How to Reduce Howlround?

1. *Choice of Microphones and loudspeakers* Microphone Cardiods and hyper*c*ardiods directed towards the loudspeaker are the most obvious choice for the microphones. It is still possible that there may be problems in the case of outside broadcasts particularly when school halls are involved. That sound may be reflected on to the live side of the microphone. Small personal microphones, although usually omnidirectional may be a good solution of the problem because the body of the wearer has a screening effect.

   Loudspeakers The obvious choice for speakers is the line source speakers. The directional properties of these speakers allow the microphones to be placed in the relatively dead regions. The quality of these speakers may not be good enough for music but conventional good quality loudspeakers can also be used for PA; there is greater risk of howl sound, of course, but sometimes the risk is not as great as expected due, may be, to the flatter frequency response of the quality speakers.

In the matter of positioning the loudspeakers it is obvious that where possible line source speakers should, be mounted with the microphones on their dead axes. If these loudspeakers are some distance away from the microphone their angling is not so important, as the sound will get diffused due to reflections and reverberations.

2. ***Electircal Processing of the audio Signal***

   (i) *Frequency Shifters* These are units which can change the audio frequencies by a few hertz. The use of these units is based on the fact that the peaks in a room response are on the average, spaced by approximately (4/*RT*) Hz, *RT* being the reverberation time. If *RT* were, say, 2 seconds, then peaks would be spaced by about (4/2) Hz. If the frequency of the PA feed is shifted by half, that is by 1 Hz, then peaks will coincide with troughs and so on with the result that the room response will in effect get flattened. However, this method has practical difficulties and not very popular with singers.

   (ii) *Graphic Equalizers*

   The second method involves the use of *graphic equalizers*. A graphic equalizer is a device for eliminating peaks in the response of a room. It has a number of controls usually sliders, each controlling a narrow band of frequencies.

When advanced above the centre level postion, the control gives a boost to the particular band and when it is lowered, the response is reduced. It thus

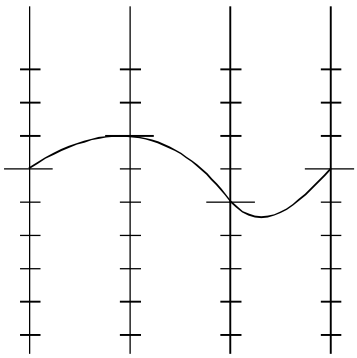


Fig 21.33 Graphic Equalizer

produces a peak or a notch in the overall frequency response. The positions of the potentiometer knobs show approximately the frequency response-graph which the system is providing and hence the name graphic equalizer.

Various types are termed according to the band of frequencies they control in the audio spectrum which is divided into ten 'octaves'. Each octave is twice the frequency of the one below it. The ten octaves are 16-31 Hz, 31-62 Hz, 62-125 Hz, 125-250 Hz, 250-500 Hz, 500-1000 Hz, 1kHz-2kHz and so on till 8-16 kHz.

The simplest equalizer has five bands i.e. there is only one band to two octaves. Such an equalizer is not suitable for a PA system and can be installed in a car.

A more common domestic type is the ten-band model. This provides one control for each octave. Again the resolution is insufficient. The chances that a feedback peak would cancel with one of the equalizer control frequencies is only twice as great as with the five band model.

The professional type is the third-octave type having thirty controls per channel where the chances of a band, coinciding with a peak are much better than the other types.

However a good PA system is often a matter of trial and error but an understanding of the principles can greatlly reduce the proportion of error.

21.21 OUTPUT POWER OF A PA AMPLIFIER

The output power of a PA amplifier should be enough to make listening comfortable at the farthest end of the hall or auditorium. The total output power of the amplifier required should be calculated on the basis that sound intensity equal to 80 dB over the threshold of hearing should be available to the audience at the farthest point. As the threshold of hearing is 10^{-12} W/m^2, the absolute intensity required for the audience is

$$10^{-12} \text{ W/M}_1 \times 10^8 = 10^{-4} \text{ W/m}^2$$

It we have an amplifier of Power P watt output, the power of the amplifier can be calculated from the formula

$$10^{-4} = \frac{P}{4\pi R^2} \times \frac{E}{100}$$

where E per cent is the efficiency of the loud speaker system and R is the distance in metres

$$P = \frac{4\pi R^2 \times 10^{-2}}{E} \text{ Watts.}$$

This power should be uniformly distributed over the loudspeaker system.

Amplifiers are classified mainly by their power output. These normally range from 20 to 250 W, but there are large ones also. For large installations, it is better to have several amplifiers of modest output power than a single high output one. This gives versatility levels that can be independently controlled. Different programmes can be relayed to different areas, and if there is break-down, only part of the system is affected.

Most public address systems except battery-operated ones range from 50 to 100 W, which is more than adequate for most systems other than large auditoriums and factories. The power rating of the amplifier should be adequate to avoid clipping and overloading. An underpowered amplifier will cause distortion and is also likely to overheat the output transistors resulting in thermal run always unless MOSFETs are used for the final stages.

21.22 PUBLIC ADDRESS AND SOUND REINFORCEMENT

The difference between Public Address and Sound inforcement lies in the fact that we ought to use the term 'Public Address' (PA) when a large number of people are being spoken to via loudspeakers, for example, in a commentary at a race meeting. Sound reinforcement on the other hand means that audience can hear something directly but loudspeakers are needed for complete audibility.

There is of course, a grey area where sound reinforcement merges into PA but that is of no significance here.

In terms of technology the best distinction to draw is based on whether or not the microphones can pick up the output of the loudspeakers. If they cannot it is sound reinforcement and if they can there is generally the major problem of hawl round. This is the familiar and embarrassings eqeal, whistle or shriek which is due to oscillations caused by a microphone picking up its own output from a loudspeaker. The effect is also called 'feedback'. This can also be sometimes beneficial and applied deliberately when a signal is returned to an earlier stage of the amplifier for stability and improving its performance.

21.23 INTERCOMMUNICATION SYSTEM

Wire telephony or the transmission of speech on lines is, perhaps, the most popular method of transmission of speech. Telephone systems can be broadly

classified into two main categories i.e. public Telephone System and Private telephone system.

(i) Public Telephone Systems an operated for commercial profit under some fixed schedules of tariff rates. Such system must serve all these who apply and are usually subject to some form of regulations as to rates and service by the public authorities.

(ii) Private Telephone Systems are those operated as auxiliaries to some other form of business or enterprise but not directly for commercial profit.

Private systems for general communication have numerous applications, they may be completely isolated i.e. not connected to other systems or they may be interconnected with public or other private systems. When interconnected with a public system they are called Private Branch Exchanges or PBX. A private branch exchange may be privately owned or leased from the operator of the public system. General communication systems operate on the same operating principles as the central office, i.e. they make use of operators or machine switches.

Intercommunication Systems

These are designed for isolated local service. This term applies to a particular method of switching where a conductor runs from each telephone instrument and the selection of the desired station is made by the operation of one of a set of keys, there being one key at each station for every other station.

Intercommunicating systems comprise entirely private telephone systems in most cases without any switch board or operating attendant for purposes of internal communication in factories, industrial plants, stores, offices, hospitals, residences and apartment buildings, etc.

Equipment Each station is equipped with a telephone set and a switching device mounted in a small box or a cabinet. For each station there is a telephone line or circuit extending to all other stations and so arranged that by depressing the appropriate switching key or button in the switching cabinet at any other station, connection can be established with this particular line, at the home station there is also a calling signal or bell (buzzer) and answering key arranged to connect home telephone set with the line.

Thus for a 10-line system, there will be 10 keys or buttons in each switching cabinet and 10 talking circuits in a small cable connected to all 10 cabinets. In addition to 10 line circuits there is another pair for supplying the talking current (common battery) to each station and still another pair for suppling ringing current to each station. Direct current is usually employed for ringing and dry batteries are used both for ringing and talking. Several keys in any switching cabinet are so arranged that the depression of one key automatically redcases or restores all others.

Operation In order to call station (extension) No. 7 from No. 3, the No. 7 button at extension No. 3 is depressed as far as possible and held there a

moment before releasing it, and meanwhile the bell connected to line No. 7 which is at station No. 7 responds. At station No. 7, the home button is depressed and the receiver taken from the hook, thus establishing communication.

Line Capacity For standard equipment varies with manufacturer. One manufacturer offers complete equipment for 6, 12, 22 and 31 station, another 6, 12, 16, 20 and 24 stations etc. These equipments are usually made in several styles for desk mounting, wall mounting or the flush wall style.

EPABX

This is a modern type of intercom system which is available in various capacities and provides different facilities. A block diagram of an EPABX is shown in Fig. 21.34.

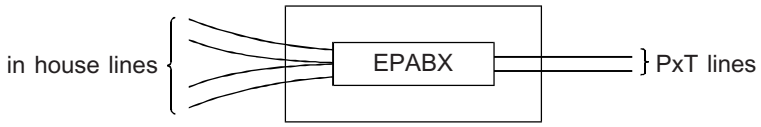


Fig. 21.34 EPABX

Facilities Besides connecting the telephone to the outside lines this system provides the following additional facilities.

1. **Extension to Extension call**
   This enables one extension user talk to another extension.
2. **Redial** Any extension user can repeatedly deal the last numbers without pressing all the numbers again.
3. **Automatic call back on busy extension.**
   If the called extension is found busy this feature automatically connects as soon as the called extension gets free.
4. **Automatic call batch on busy trunk line.**
   If all/any function line, are busy, this feature informs the user as soon as the junction trunk line is free.
5. **Call Transfer**
   Any internal or external call received at any extension can be transfered from that extension to any other extension.
6. **Call forwarding**
   This feature allows an extension user to receive the calls at any other extension.
7. **Call pick up**
   If another extension is ringing, this feature allows user to receive that call at his own extension without physically moving to that particular extension.
8. **Follow Me**
   Incoming call can be made to follow the extension user. In other words, the extension user can use any extension to receive in coming calls directed at his original extension.

9. **Call camp on.**
   This feature allows an extension to transfer call even to a busy extension.
10. **Call Parking**
    In case the extension desires to become free temporarily to attend to same important function, using this feature makes the extension free without loosing the call.
11. **Setting of alarm clock.**
    Each extension can be pre-set to ring at a pre-determined time.
12. **Conference**
    While talking to one party, it is possible to talk to a third party or a fourth party thus making a conference arrangement possible.

The method of operating the intercom equipment to make the above features possible is provided by the manufactures in a USER-MANUAL supplied with the equipment.

21.24 INTEGRATED SERVICES DIGITAL NETWORK (ISDN)

Recent technological developments in the telecommunication fluid have laid a firm foundation for the provision of new and modern services by means of Integrated Services Digital Network (ISDN). As a consequence, it is now possible to offer a range of powerful services that are of significance for both the business and residential subscribers.

The service will be offered initially in Delhi, Bombay, Calcutta, Madras, Bangalore and Ahmedabad followed by Jaipur, Ranchi and Hyderabad. ISDN will be offered through the new technology imported exchanges.

Is ISDN Different from the Existing Analog Phones

Integrated Services Digital Network (ISDN) has emerged as a powerful tool worldwide, for provisioning of different services- voice, data and image- by means of the existing telephone network. ISDN is being viewed as a logical extension of the digitalisation of the network and most developing countries are in different stages of implementing ISDN. In ISDN even subscriber voice is sent in the digital from and so the phone is called a digital phone.

An ISDN subscriber can establish at least two simultaneous independent calls on the existing pair of telephone line (Basic Rate ISDN), where as only one call is possible at present. The two simultaneous calls in ISDN can be of any type speech, data, image or video.

The call setup time for a call between two ISDN subscribers will be very short, of the order of 1 to 2 seconds.

ISDN will also support a whole new set of additional facilities, called supplementary services.

The ISDN subscriber will have full connectivity both nationally and internationally to other telephone subscribers.

What Equipment can be Connected to the ISDN Line?

In the ISDN, the telephone line is terminated on a common box, called the Network termination, provided at the subscriber premises. Beyond this box, on the internal wiring in the subscriber's premises, up to 8 ISDN terminals can be connected. These ISDN terminals can be of several types, for example, ISDN telephone, Personal Computer (PC), Video Phone, Video conferencing equipment, etc. In addition, existing terminals such as rotary and push button telephones, FAX machines and Modems can also be connected to the internal wiring with suitable adaptors Fig. 21.35.

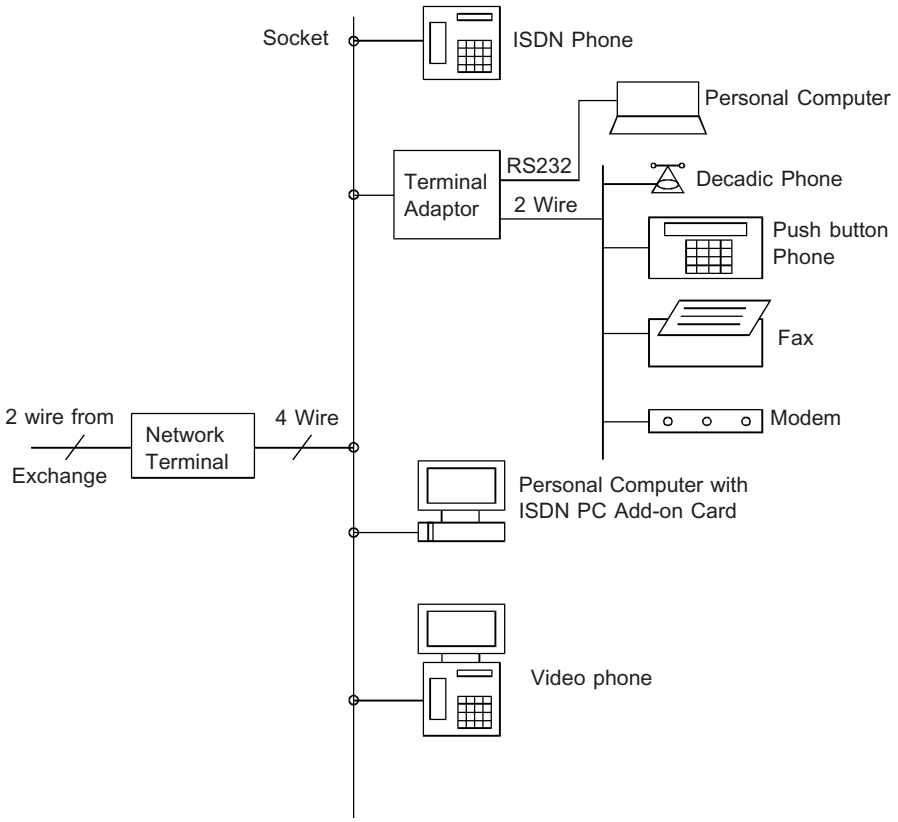


Fig. 21.35 Customer Premises Installation

Services Offered by ISDN A wide range of services catering to the needs of residential and business subscribers will be offered.

Data files between PCs will be transmitted at a high rate of 64,000 bits/s. This is more than 6 times the typical speeds possible at present.

An attractive service of ISDN being offered to subscribers is Video conferencing which has potential of curtailing travelling requirements of business executives. Video conferencing can be achieved between any two ISDN

customers on dial up basis on existing telephone lines. Two types of Video conferencing are being offered. For high quality video (348 kb/s), three ISDN lines will be required by the customer. In this case in addition to video image of the participants, still pictures of documents and drawings can also be transmitted. For ordinary video conferencing (128 kb/s), a single ISDN line will be sufficient.

'Phone plus' Facilities Offered by ISDN ISDN will also support a whole new set of additional facilities, called supplementary services for speech calls. The following services will be available for calls made between ISDN subscribers:

Calling Line Identification Presentation (CLIP)

When an ISDN subscriber receives a call, the calling subscriber number will be displayed on his ISDN telephone before the called subscriber answers the call. (The ISDN phone has a small LCD display resembling those available in calculators). Thus, the subscriber knows the telephone number of caller from the very beginning, even before answering the call. For example when the subscriber is already in conversation, he may choose to attend the second incoming call depending on the caller's number displayed. This service will be provided free of cost to all ISDN subscribers.

Calling line Identification Restriction (CLIR)

This service may be provided on subscription by one time payment. By means of this service, the calling subscriber will be able to prevent the presentation of his number to the called subscriber (Prevention of CLIP). However, this service will be overridden by certain agencies such as police and fire services, since they may need to know the identity of the caller in all cases.

Advice of Charge (AOC)

The amount charged for a call, in terms of call units, will be displayed on the calling subscriber's ISDN phone. In case of long distance calls, it is possible to see the count of metering pulses incremented for this call. This will be continuously updated as the call is in progress.

Multiple Subscriber Number (MSN)

As up to 8 terminals can be connected in parallel on the subscriber premises wiring, to call a specific terminal (PC to call a PC, and phone to call a phone), separate number can be allotted to each terminal. This will be particularly useful when the call is received from a normal (analog) subscriber. In case call is received from an ISDN subscribe, the terminal selection will be automatically made.

Call Forwarding Services (CF)

The call to a subscriber can be forwarded to another number under different criteria like, subscriber being busy or no answer. Calls can be forwarded unconditionally also.

Call Forwarding Busy (CFB)

If the called subscriber is busy, the incoming calls to his number can be diverted to another number specified by him.

Call Forwarding No Answer (CFNR)

If the called subscribers is not available (or does not answer the call), the incoming calls to his number can be diverted to another number specified by him, after a few rings.

Call Forwarding Unconditional (CFU)

All the incoming calls to a subscriber can be diverted to another number specified by him. The ring directly goes to the diverted number in this case.

Terminal Portability (TP)

In the subscriber premises up to 8 terminals can be connected to a single ISDN line. These terminals can be in different rooms and also can be on different floors. The internal wiring in the subscriber premises is terminated on sockets. During conversation it is possible to transfer the call from one terminal to another or even remove the terminal and connect it to another socket at a different location. This facility is available for calling as well as called subscriber.

Call Hold

During conversation, it is possible to hold atleast two more calls. The subscriber can switch between these calls.

Closed User Group (CUG)

Companies with offices in different cities can have their ISDN number in a closed user group. The subscribers can call each other using short numbers as if they are connected to a PABX. This group enjoys certain calling privileges like selective call barring and additional level of security.

ISDN Phone This terminal, in addition to having a handset and dialing key pad, also has an LCD display, additional key for storing frequently dialed numbers and other function keys.

(i) Display The ISDN phone has an LCD display resembling those available in calculators. The number dialed is displayed, so that the caller can leisurely enter the digits without mistake. This reduces wrong calling. In case of CLIP

service, the calling number is displayed. In case of AOC, the number of call units charged is displayed. This is also used for programming of MSN, CF, etc. In addition to providing tones on the status of the call (dial tone, busy tone, ringing tone, routing tone, etc.) the status is also displayed. This combined with speaker phone facility provides true hands free operation.

(ii) Logging: The logging facility provides for automatic storing of calling subscriber number, when the call could not be answered. The calling number can be recalled using the log.

Other facilities like redialing, memory dialing and speaker phone are also available. The phone is also called a digital phone, since signals are transmitted and received in digital form. So the phone provides clear and noisefree conversation. In ISDN, the line condition is always checked continuously, so that any fault in the line is immediately detected.

Terminal Adaptor

The existing terminals like rotary telephone, pushbutton, telephone, MODEMs, Personal computers and FAX machines can be connected. This is a quick solution to use many ISDN facilities with the existing terminals. Only a Terminal adaptor needs to be procured. The terminal adaptor provides connectivity to ISDN line on one side. On the other side a number of connectors are possible.

(i) Analog connector: Roto telephones, push button telephones (pulse type or tone type), MODEMs, FAX, answering machines, cordeless phones, etc can be connected.

(ii) Data port: Any PC with RS232C connection (serial port of the PC) can be connected to this port. Data transfer using standard software packages like Xtalk or Procomm is possible up to 9600 bps. MODEM is not required.

PC add-on ISDN Card

This card can be fitted in standard Personal Computers and can be used for data transfer at 64 kb/s. This card fits into vacant slot of any standard 386/486 PC. A software is also provided which will be installed in the PC. The connector from the PC is connected to the ISDN line. Using the software files from the PC can be transmitted or received at 64 kb/s. For e.g., the complete information on a 2 Mb floppy can be transmitted in 4 minutes. Using both the channels it is possible to send the data at 128 kb/s.

Video Conferencing Equipment This equipment consists of a computer, TV monitor, camera and other control units. This is a professional equipment owned by telephone Deptt. Users utilise this facility paying the usage charges.

This equipment will be connected to the network using 3 ISDN lines. The equipment establishes connection to a similar equipment on the other side, by dialing through the network. Moving video images of the conference participants can be sent as well as received along with their conversation. It is possible to send diagrams and photographs by a still picture camera. Data transfer can also be done simultaneously. The equipment works at 384 kb/s.

In addition video images can be sent or received on auxiliary equipment like VCR.

Using the control panel, the video camera can be moved or zoomed on the required participants of the conference. The transmitted picture can be viewed along with the received picture.

Desk Top Video Conferencing This is a compact version of the video conferencing equipment, usually PC based. The PC is upgraded by one or two add on cards. A camera is provided which can be appropriately placed. Only 128 kb/s of transmission capacity is required and therefore a single basic access ISDN line is sufficient. In many models, it is also possible to transfer files and jointly edit documents.

SUMMARY

Sound is a sensation produced in the brain by nerve-pulses sent by the ear drum when subjected to variations of air caused by a vibrating body. These variations of air pressure are sound waves which travel in all directions with the speed of sound. The relation between the wavelength, frequency and speed in the case of sound waves is the same as in the case of radio waves except for the speed of sound which is 1100 ft/s in air.

A part of the sound produced by a vibrating body reaches the ear after repeated reflections from the walls and ceilings and the sensation of sound gets prolonged. This phenomenon is called reverberation. A certain amount of reverberation is desirable and is produced by treating the walls and ceilings with materials having different sound reflecting properties. The science which deals with the design of buildings with regard to their sound reflecting and sound absorbing properties is called *acoustics.* Reverberation time (RT) is defined as the time taken for sound to decay through 60 dB. Various formulas are available for the calculation of RT—All materials absorb sound to some extent, the sound absorbing quality varying with the nature of the material. Materials with different absorption coefficients are used in the acoustical design of studios/ theaters, cinema halls and auditoria.

The response of the human ear to change of sound intensity or loudness is logarithmic and consequently the unit adopted for the measurement of sound intensity is also logarithmic in nature and is called the decibel.

The microphone is a device for conversion of sound pressure variations into corresponding electrical variations. For this, the sound waves set up mechanical vibrations in some moving element to generate small AF voltages which can be amplified. Microphones are classified as constant velocity or constant amplitude types depending upon whether the voltage generated is proportional to the velocity or amplitude of the moving elements. Of the five types of commonly used microphones the carbon microphone, the crystal microphone and the capacitor microphones are constant velocity microphones. A number of new types of microphones with directional properties are now available for

special purposes. A pickup is an electrical device for reproducing the sound recorded on a disc. The mechanical vibrations of the stylus attached to the pickup as it follows the groove undulations are converted into corresponding electrical variations. Magnetic pick-up and crystal pick-up are the two types commonly used in disc playback equipment. The principle of operation of each type resemble the operation of the corresponding type of microphone. A tone arm is used to hold the pick-up in position and to guide the pick-up cartridge properly through the grooves so as to keep it tangential to the groove at the point of contact.

A loudspeaker is a device which converts electrical energy into sound energy. Of the various types of loudspeakers, the most commonly used type is the dynamic or the moving coil type of loudspeaker. In this a voice coil placed in a strong magnetic field makes a diaphragm or cone vibrate when AF currents from an amplifier are passed through the voice coil. A matching transformer is used to match the output impedance of the amplifier to the impedance of the voice coil for maximum sound output. Baffles and enclosures are used to improve the performance of a dynamic loudspeaker. Woofers and tweeters are used with cross-over networks for getting a uniform frequency response over the entire frequency range.

Horn loudspeakers are used for large-scale reproduction of sound. These loudspeakers are high efficiency reproducers with good frequency response even at low frequencies. Column loudspeakers consisting of a number of dynamic speakers mounted on a cabinet are used where a narrow directivity pattern in the vertical plane is required.

A PA system is a system for amplifying speech so that it can be heard by larger gatherings and at longer distances.

Basic requirements of a PA system are comfortable level, intelligibility, naturalness and reliability.

Howlround is the bug bear of a PA system and can be prevented by proper choice of microphone, loudspeakers, and by electrical processing of the audio signal.

Intercom System is a private telephone exchange that provides communication between the inhouse users of a premises as well as between various extension and the P and T basis.

ISDN is a recent development that provides different services—voice, data and image by means of existing telephone network.

Review Questions

1. What is sound? How are sound waves produced in air?
2. What is reverberation? How can reverberation time of a room be adjusted by acoustic treatment of its walls?
3. What is a decibel? An amplifier is delivering 1 W output power to its load. Express the power level in decibels with reference to 6 mW as 0 db. (Ans. 22.2 db)

4. Describe with a sketch the principle of operation of a moving coil microphone.
5. What type of microphone is used in telephone equipment? Describe its working.
6. Explain the operating principle of a magnetic pick-up.
7. Describe the main parts of a dynamic loudspeaker.
8. Explain the use of a baffle in a loudspeaker.
9. What is a tweeter and a woofer? Explain their use in a high fidelity loudspeaker system.
10. What is a horn type loudspeaker? How is this type of speaker useful in large-scale reproduction of sound.
11. What is a PA system? Give its main requirements.
12. Write short notes on:
    (a) column loudspeakers,
    (b) impedance matching in loudspeakers,
    (c) tone arms,
    (d) microphone mixers.
    (e) Howlround in a PA system.
    (f) Inter com. system
    (g) ISDN
13. State whether true or false with brief reasoning:
    (i) A 1 kHz sound wave will have a higher wavelength than a 1 kHz radio wave.
    (ii) A music studio will have a higher RT than a Talks studio.
    (iii) A condenser microphone has a fig-of-eight polar diagram.
    (iv) A horn loudspeaker is more efficient than a radiator type of loudspeaker.
    (v) A woofer is a speaker that responds to higher frequencies.
    (vi) A PA system needs an omnidirectional microphone.
    (Ans: (i) False, (ii) True, (iii) False, (iv) True, (v) False, (vi) False.)

chapter 22

Sound and Video Recording Systems

22.1 INTRODUCTION

Sound and video recording systems play such an important part in all modern communication systems including radio, TV and other forms of audio-visual systems that it is necessary to understand the basic principles involved in both these recording systems. Sound recording, being the earliest and more common form of recording, is described first.

22.2 SOUND RECORDING

Sound recording is essentially a process of storage of sound. In sound recording, the acoustical signals are converted into some form of permanent or semi-permanent record from which these can be reconverted into sound. It was in 1877-78 that the earliest sound recording was made by Edison who found a means of converting sound pressure waves into mechanical motion to cut a varying groove on a soft wax cylinder. Reproduction was effected by allowing a needle to retrace the path along the groove and to make a diaphragm vibrate, thereby creating sound waves again. This was the beginning of gramaphone disc recording as we know it today. Various methods of sound recording have since been developed and it has reached a very high degree of perfection in the modern magnetic recording, popularly known as *tape recording*.

22.3 SYSTEMS OF SOUND RECORDING

Depending on the technique employed for the storage of sound, there are three established forms of sound recording in use at present. These are (i) disc recording, (ii) film recording, and (iii) magnetic recording.

Disc recording is a mechanical process in which storage of sound takes the form of a wavy track cut into a semi-rigid material, which the reproduction needle is forced to follow.

In film or optical recording, the record is produced in the form of a photographic image of variable area density. Compact Disc (CD) is the latest form of optical recording.

In magnetic or tape recording, the sound information is stored in a magnetic material in the form of elemental permanent magnets corresponding in length and strength to the signals to be recorded.

Of the three systems of sound recording mentioned above, tape recording is the most popular and universally employed system of sound recording. This system of sound recording will, therefore, be described first in detail. A brief description of the other systems of sound recording will follow at a later stage.

22.4 MAGNETIC (TAPE) RECORDING

The process of magnetic recording is based on the principle of electromagnetic induction. In magnetic recording, a magnetic material is magnetised to varying degrees of magnetisation along its length when moved in front of a coil carrying AF currents from an audio amplifier. The magnetic variations so recorded can be reconverted into electrical currents when the varying magnetic flux emitted by the magnetised material is linked with an electrical circuit. This completes the magnetic recording and playback process.

In the early days of magnetic recording, round steel wires and flat steel tapes were used but these were subsequently replaced by thin paper or plastic tapes coated with a finely powdered magnetic material of relatively high coercivity. These finely powdered magnetic particles of iron oxide behave like small magnets which have a random distribution when unmagnetised but are given definite orientation by the magnetising force of the audio currents of the recording head. These tapes are wound on reels or spools and are made to run at constant speed both during the recording and reproducing process. Very high quality tapes combined with rapid progress in electronic circuitry have made magnetic tape recording one of the most useful sound recording system of today. Tape recording finds wide applications not only in the field of entertainment and broadcasting but is extensively used in other fields like data storage, automatic control and computer applications.

One great advantage possessed by magnetic tape recording compared to other systems of recording is that the matter recorded on the tape can be completely *wiped off* by a process known as *erasing*. The same tape can be used over and over again for recording and playback of different programmes which is very economical. The magnetic system of recording lends itself to immediate playback, and simultaneous monitoring of the recorded programme is possible.

Editing is much simpler with tape recording than with disc recording. Stereo recording is easier to make with tape recording than disc recording. However, noise level is slightly higher with tape than with disc recording.

22.4.1 Tape Recorders

A tape recording machine is required to perform three important functions, viz. recording, playback and erasing. Figure 22.1 is a block diagram which indicates the basic functions performed by a tape recorder system.

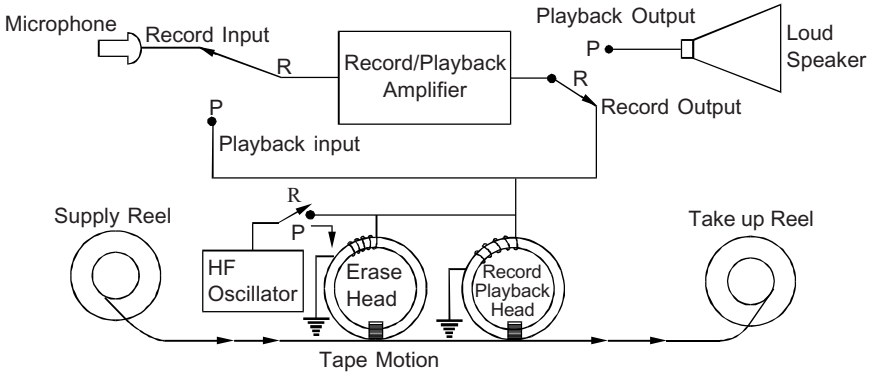


Fig. 22.1 Block Diagram of a Tape Recording System

In the block diagram, the same amplifier is used for recording and playback and is called the record/playback amplifier. During the recording process a ganged switch connects the input of the amplifier to the microphone and the output is connected to the record head so that the amplified audio signals are impressed on the moving tape. In the playback position, the magnetic signals picked up or detected from the tape by the same head are applied to the input of the amplifier for reproduction by the loudspeaker. One and the same head performing the combined functions of recording and playback is known as the *record/playback* head. The tape is wound on the supply reel and is moved at constant speed and under constant tension in front of the record/playback head during recording and playback operations. The tape is finally collected by the take-up reel.

An erase head located to the left of the record/playback head wipes off any information from the tape before the tape reaches the record head for a fresh recording. The erase head is energised by the HF oscillator which also applies an AC bias to the record head during the recording process for improving the quality of recording. The erase head is not energised during the playback process to prevent the recorded matter from being erased. A tape transport system consisting of a motor, drive assembly, belts and pulleys enables the tape to move at constant speed during record and playback processes and also to move faster during the fast-forward and rewind processes.

A tape recorder consists of two main sections, the electronic section and the tape transport section. The electronic section which performs the basic functions of recording, playback and erasing, is described first.

22.4.2 Recording

In the recording process, the audio signal from a microphone or some other source is amplified by the recording amplifier and fed into a ring shaped electromagnet called the recording head as in Fig. 22.2. The small air gap at the ends of the recording head emits a magnetic flux which impresses on the moving magnetic tape a magnetic pattern varying in magnitude and polarity in accordance with the audio signal.

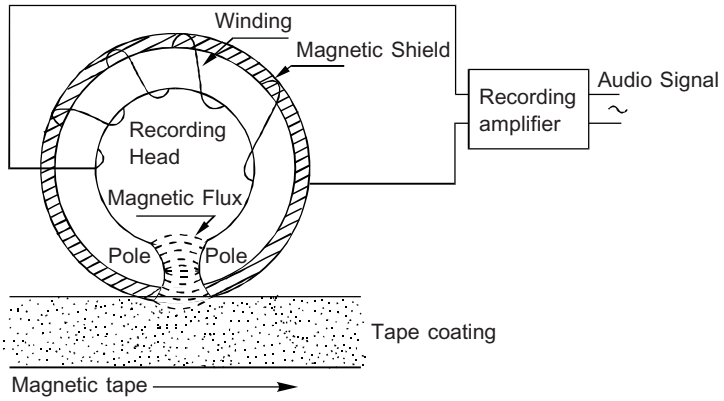


Fig. 22.2 Recording Head

The recording head consists of a coil wound round a ring shaped core material of high permeability (mu—metal). This material conducts the magnetic lines of force easily to the poles at the end, which develop opposite polarities when the current flows through the coil. The actual polarity of the poles depends on the direction of flow of the current through the coil. A magnetic shield surrounding the core and the coil protects the recording head from external magentic and electric influences.

The strength and polarity of the magnetism generated in the coil varies in accordance with the sinusoidal variations of current in the audio signal. The magnetism induced in the magnetic tape, which is moving at a constant speed in front of the pole pieces, also varies with the magnetic field. The amount by which the individual segments of the tape are magnetised depends on the strength of the magnetic field at the particular moment that the tape moves past the pole pieces.

A very fine slit or air gap is provided in the front portion of a recording head. The air gap enables the magnetic lines of force produced between the poles to pass through the magnetic tape which provides an easier path for the magnetic flux than the air gap. A magnetic pattern is recorded on the moving tape in the form of varying degrees of magnetisation corresponding to the audio signal. When the same tape is moved past the air gap of a similar playback head, the magnetised portions of the tape induce varying voltages in the coil of the playback head which correspond to the magnetically recorded

signal. The width of the air gap is very small and varies with the frequency response expected of the tape recorder. The width of the air gap is less for a playback head compared to a recording head. However, in most tape recorders operating on a slow speed (4.75 cm/s) a compromise gap width varying between 0.00025 cm and 0.00064 cm is used.

22.4.3 Biasing

Tape recording cannot be done by simply applying a varying magnetic field to the iron oxide coating on the tape. Due to a phenomenon called *hysterisis*, the tiny particles of the magnetic oxide coating on the tape resist any force tending to magnetise these elemental magnets. Once magnetised, these particles cannot be easily demagnetised by a demagnetising force. In other words, the relation between the magnetising force and the magnetism produced is not linear as shown in Fig. 22.3.

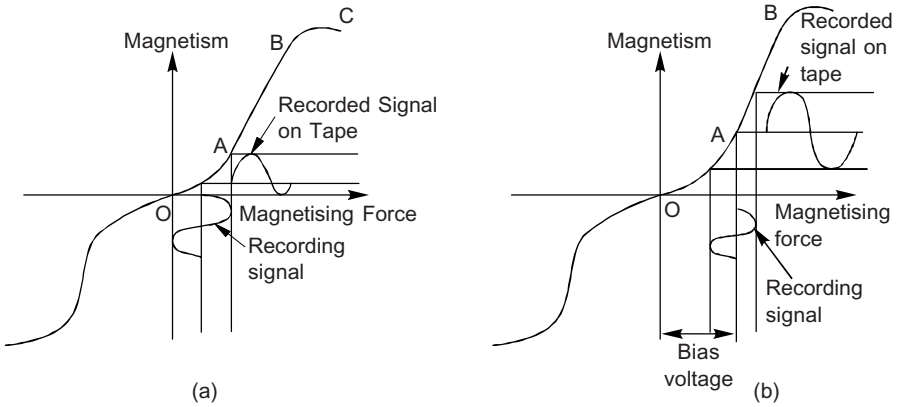


Fig. 22.3 Relation between Magnetising Force and Magnetism Produced (a) Non-linear Recording (b) Linear Recording

The curve rises slowly between *O* and *A* which is the non-linear portion of the curve. A large amount of magnetising force here results in only a small increase in magnetism. A sinusoidal recording signal does not produce a sinuosidal recorded signal on the tape and the output is distorted as in Fig. 22.3(a). The portion of the curve between *A* and *B* is straight, where the magnetism produced in the tape is in direct proportion to the applied magnetising force. This is the linear portion of the curve where a sinusoidal applied signal produces an output signal without distortion, as in Fig. 22.3(b). Therefore, to avoid distortion in the recorded signal a bias must be applied to the record head to raise its operating point to the linear portion of the curve. The bias will depend on the quality of magnetic coating on the tape. Biasing in tape recording is similar to the biasing of a vacuum tube to the linear portion of its transfer characteristics for distortionless output.

The bias applied to the recording head to avoid distortion can be either a DC bias or an AC bias.

22.4.4 DC Bias

In this method of applying bias, a DC voltage is fed to the recording head to raise the operating point to the linear portion of the transfer characteristics. This type of bias which was used in the early days of magnetic recording no doubt reduces distortion but the constant level of magnetisation produced on the tape results in a higher noise level. In modern tape recording, the method used for applying bias is the AC bias which produces much better results than the older DC bias method.

22.4.5 AC Bias

This method makes use of a HF AC bias whose frequency lies in the supersonic frequency range between 30 kHz to 150 kHz. Here the recording signals are superimposed on an alternating magnetising field whose amplitude and frequency are both considerably greater than those of the recording signals. This superimposition is easily effected by applying the HF bias current and the signal current either to separate coils or to one common coil on the recording head. Figure 22.4 shows how the HF bias signal raises the magnetising level of the audio signal into the linear portion of the magnetising curve. The input signal is in the form of a modulated wave formed by the HF bias and the audio signal. This waveform produces output signal on both sides of the magnetising curve. The modulation envelope remains undistorted by the non-linearity of the magnetising curve and whatever distortion occurs is only in the HF bias voltage as it crosses from positive to negative polarity. This is indicated by steps at E. However, the bias frequency being very high in the supernoise range will

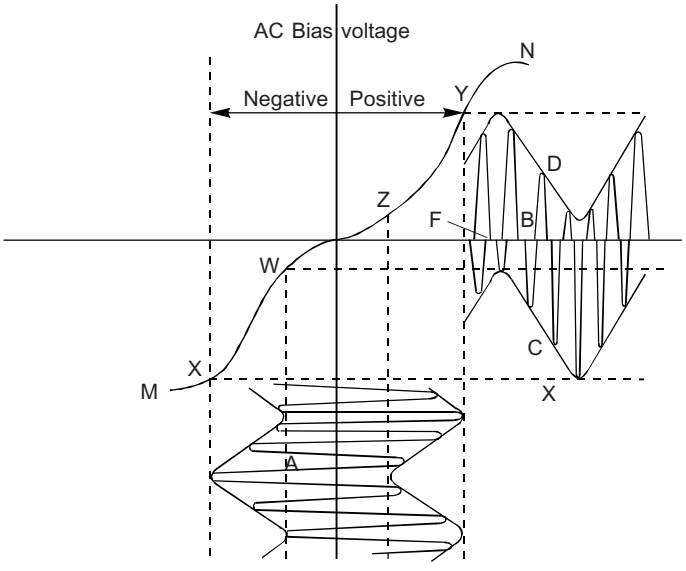


Fig. 22.4 Recording with AC Bias

not be heard in the playback. This can also be removed by filtering of the output signal.

In fact, there are now two undisturbed audio signals, one in each half of the curve which can be combined electronically to produce a powerful output signal. This not only results in a greatly improved signal-to-noise ratio but considerably reduces the natural tape "hiss".

The sensitivity as well as the maximum attainable undistorted output level are both increased while the absence of a DC component in the magnetising force ensures a low noise level.

It is important that the AC bias signal must be a pure sine wave. If the bias waveform is not balanced, the resultant DC component will cause higher noise level and distortion. The bias frequency chosen must be sufficiently high, preferably three to four times the highest signal frequency. This is necessary to avoid whistles during playback due to intermodulation products.

The amount of bias required depends on the characteristics of the tape and the recorder. The bias should neither be too low to remain in the non-linear portion nor too high to be in the saturation region. Ordinary inexpensive recorders like cassette recorders use a fixed bias level that is set for a particular recording head and the type of tape to be used with the recorder. However, the bias level in high quality tape recorders is adjustable to compensate for variations in characteristics. The bias is adjusted with a trimmer till the maximum undistorted output is obtained with a specific input signal to the recorder.

22.5 PLAYBACK

Playback process in tape recording is the reverse of the recording process. In playback, the magnetic field impressed on the tape during recording induces a signal voltage into the coil of the playback head. As the tape is pulled past the playback head, the varying magnetic field on the tape is converted into a signal voltage varying in accordance with the amplitude and frequency of the recorded signal. The induced signal is fed to the playback amplifier which amplifies it to the level required to operate a loudspeaker as in Fig. 22.5.

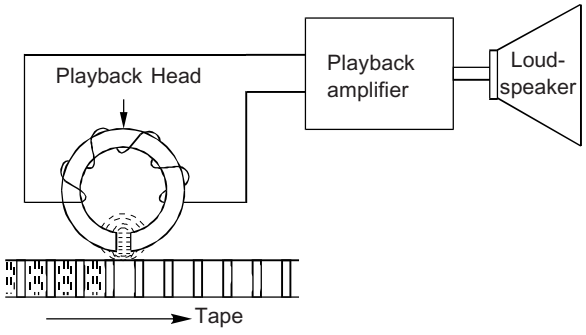


Fig. 22.5 Playback Process

The playback head is similar to the recording head except that the gap width is much smaller in the case of the playback head. However, in smaller tape recorders like the cassette tape recorder, the record head itself serves as the playback head also, and the combined head is known as the record/playback head. The gap width is a compromise between the gap width required for recording and playback heads when used separately. Where the same head is used both for recording and playback, the audio amplifier is also common for the recording and playback processes and is known as the record/playback amplifier.

The amplitude of the signal induced in the playback head depends on the frequency and amplitude of the recorded signal. The output from the playback head should be the same signal as was fed to the recording head but this is not generally the case. The playback signal is different from the recording signal because of a number of losses that occur during the recording and playback processes. Some of these losses are the magnetic losses due to hysteresis and the eddy current losses both of which increase with frequency. The self demagnetising effect of the recorded signal also increases as the wavelength of the recorded signal decreases and the neighbouring magnets become smaller. The AC bias also produces some demagnetising effect and causes reduction in the strength of the impressed magnets at higher frequencies.

As a result of all these losses the output during playback is considerably reduced at higher frequencies, particularly at the slow speed operation of the inexpensive type of tape recorders. In order to make the output quality as natural as possible, certain frequency compensation networks, to make up for the losses at high frequencies, are used during playback. With the equalisation circuits, a flat frequency response from 20 Hz to about 10 kHz is possible with ordinary cassette recorders but in high quality tape recorders a uniform frequency response from 20 Hz to 20 kHz is possible.

Tape speed is very important for determining the response at higher frequencies. The increased wavelength at higher tape speeds reduces magnetic cancellation effect for any particular frequency. High fidelity recordings like music recordings are always done at higher speeds compared to speech recordings. The gap width also affects the frequency response. The gap width must be a compromise for adequate response at high and low frequencies. Generally, the gap for the combination record/playback head is 0.0013 cm or less but for a separate record head with AC bias, the gap width used is of the order of 0.0025 cm.

22.6 ERASING

The process by which the magnetic information recorded on the tape can be removed or wiped off before recording new information is known as *erasing*. The erase operation is performed automatically during the recording process when the tape is moved in front of an erase head or an erase magnet before reaching the record head. This demagnetises the tape or renders it magnetically

neutral so that fresh information can be recorded on the tape. In fact, erasing is the biggest advantage possessed by tape recording so that the same tape can be used over and over again for a number of recordings.

Erasing is either done magnetically or by the use of an *erase head.*

In magentic erasing a permanent magnet or magnets come forward during recording and almost touch the tape which gets demagnetised before it reaches the recording head. During playback the permanent magnet recedes and remains away from the moving tape. The use of permanent magnets for erasing has, however, been superceded by the modern and more effective method of AC eraser.

In the AC erasing, an erase head located immediately before the record/playback head is used for erasing the tape. The erase head is fed with a HF signal from the same oscillator that is used for AC biasing except that the full output of the HF oscillator is applied to the erase head. Any particular area of the tape moving past the erase head will be subjected to a large number of cycles of the HF field before it leaves the erase head. Each small area on the tape is alternately magnetised, first in one direction and then in the other direction. The alternating field also decreases in intensity as the tape moves away from the erase head and finally the magnetism decreases to zero and the tape is completely demagnetised.

The erase head is similar to the record/playback head that it has a wider gap that ranges from 0.05 to 0.0076 cm. The wider gap in the erase head allows a particular area of the tape to remain under the influence of the demagnetising field for a much longer time and thus get completely erased. The erase head is energised only during recording by a suitable switching arrangement.

The tape can also be erased by feeding the erase head with DC. A strong DC field aligns the magnetic areas in one direction and so removes any audio information recorded on the tape. However, demagnetisation is not complete unless a demagnetising field of opposite polarity is also applied. DC erasing is not a very satisfactory method of erasing. Moreover, with a DC field the tape is not demagnetised but is uniformly magnetised in one direction at a particular level. As in the case of DC biasing, this increases noise and produces distortion. Except for some special applications, the DC erase method is not widely used. *Bulk erasers* use 50 c/s AC for erasing the entire tape which is placed on the bulk eraser. When power is applied, the strong magnetic field saturates the tape which is removed slowly and evenly to gradually reduce the intensity of the field.

22.7 THE MAGNETIC TAPE

The recording medium universally employed in tape recording is the magnetic tape. It consists of a thin plastic film or paper backing which is coated on one side with a dispersion or *paint* containing very finely powdered particles of a magnetic material of high coercivity. The thickness of the backing or base ranges from 0.0013 to 0.0018 cm whereas the coating is 0.00076 to 0.002 cm

thick. The standard width of a recording tape is 0.64 cm but the modern cassettes use tapes which are only 0.38 cm wide. The tapes are wound on reels which vary from 12.7 to 30.5 cm in diameter. The tape length employed may be anything from a few hundred to several thousand metres. A 18 cm diameter spool may contain about 360 m of the standard tape, 550 m of long play, 730 m of double play tape.

In both the recording and reproducing process, the tape is wound from a full to an empty spool at a constant speed. In a cassette, however, both the spools are enclosed into a plastic container which can be inserted into the cassette recorder.

The magnetic material invariably used for the tape is gamma ferric oxide (γFe_2O_3) acicular powder made into a dispersion of paint by mixing it with a binding agent. The binder which acts as a cement for holding the magnetic particles together generally consists of nitrocellulose but also contains certain other resins, solvents and lubricants to make the dispersion last longer and stick better to the base material. A coating of the disperion is applied uniformly to the base material by a special manufacturing process. Cellophone, kraft paper, cellulose acetate, polyvinyl chloride (PVC) and mylar (terelyne) have been employed as base materials for the magnetic tapes. Cellulose acetate (CA) is most commonly used as the base material but mylar is fast replacing cellulose acetate particularly in the case of extra long play tapes. Tapes coated with ferrichrome (cro_2. Fe_2O_3) and also those coated with a mixture of cobalt and ferrous oxide result in substantial improvement in S/N ratio and frequency response of the system.

22.8 ELECTRONIC CIRCUITS

Electronic circuits form a very important part of a tape recorder. These include amplifiers, oscillators, equalisers, recording level indicators, automatic level control and power supplies. Some of these circuits will be briefly described.

22.8.1 Amplifiers

Amplifiers are required both during recording and playback processes. While recording, the amplifier is used for amplifying the weak microphone signals whereas in the playback process the function of the amplifier is to amplify the signals from the playback head. The recording as well as playback amplifiers are generally good quality audio amplifiers whose characteristics can be modified by a switching arrangement to suit the recording and playback characteristics of the tape recorder.

Amplifiers in modern tape recorders mostly make use of solid state circuitry consisting of transistors, ICs and semiconductor diodes. Low noise silicon transistors, ICs are preferred to germanium transistors in the later models. In low cost tape recorders and particularly the cassette tape recorders using a common record/playback head, the same amplifier is used for recording and playback with a changeover switch for effecting necessary changes in the circuit arrangements.

The record/playback amplifier comprises two to three stages of amplification consisting of a preamplifier, a driver amplifier and an output stage. A record/playback amplifier used in a cassette tape recorder is shown in Fig. 22.6. The preamplifier consists of a BC149B transistor which has been specially chosen for the input stage where a low noise transistor is essential. Its collector is directly connected to the base of BC140C which forms the audio amplifier. A DC feedback path is provided from the emitter of BC140C to the base of BC149B through a 100 K resistor. The driver stage uses a high gain transistor type BC148B which drives a conventional push-pull output stage comprising a matched pair of AC128 transistors. Two transformers have been used for coupling and matching in the output stage thereby ensuring a higher gain from this stage. By a suitable switching arrangement, the output stage is made to drive a loudspeaker during playback and feed the record/playback head during recording.

The modern trend in tape recorder circuit design is to use a complementary pair (one PNP and other NPN) of transistors in the output stage with a suitable transistor to drive the complementary pair in correct phases for operation as a complementary symmetry pushpull amplifier. This circuit eliminates the two transformers and reduces the cost and weight of the equipment.

Equalisation has to be provided to compensate for the losses that occur during recording and playback processes. In cassette tape recorders operating at a fixed speed, negative feedback is used for providing the required boost through selected feedback components. During recording, the feedback reduces only distortion and noise and does not materially affect the frequency response. However, during playback the high frequencies as well as the low frequencies have to be boosted by reducing the negative feedback at these frequencies to ensure good frequency response over as wide a range of unsafe frequencies as possible.

In high quality tape recorders capable of operating at a number of speeds, different RC networks for equalisation are used at different speeds for satisfactory frequency response during playback. With proper equalisation, it is possible to get a uniform response from 50 Hz to about 8 kHz in the case of low speed cassette tape recorders whereas the tape recorders operating at higher speeds can provide good response up to about 20 kHz.

22.8.2 HF Oscillator

A built-in HF oscillator forms an important part of the electronic circuit of any modern tape recorder. This oscillator generates a supersonic frequency of about 50 kHz which is required for the biasing of the recording head during the recording process. The same HF oscillator also feeds the erasing head except in cases where DC or magnets are used for erasing. A single transistor is connected as a Hartley or Colpitts oscillator in most cases. In the circuit of Fig. 22.6 a transistor BC148B operates as an oscillator to supply HF bias to the reord/playback head through a capacitor and a variable resistor (100 K). DC has been used for erasing.

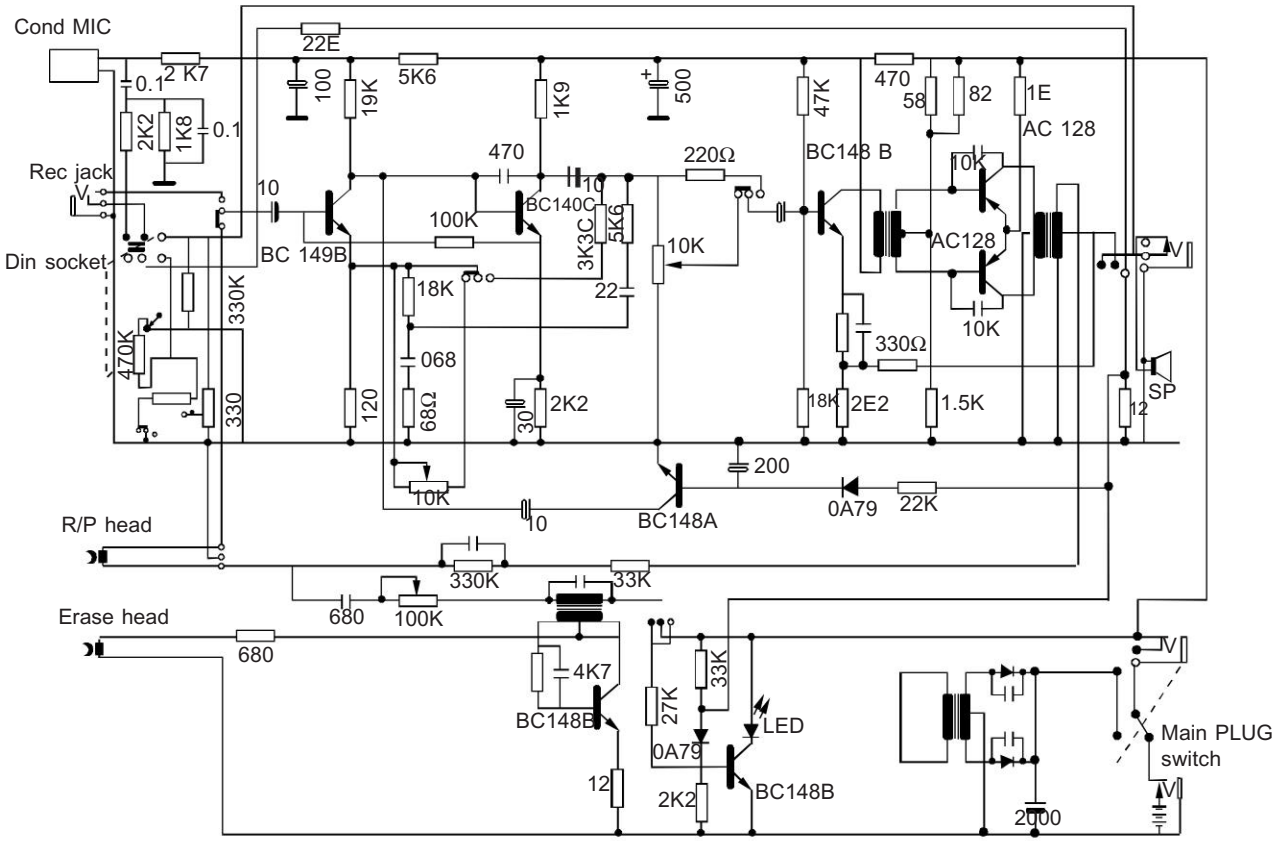


Fig. 22.6 Circuit Diagram of a Cassette Tape Recorder (*Courtesy:* Weston)

22.8.3 Other Auxiliary Circuits

Modern tape recorders include certain auxiliary circuits in addition to the amplifiers and oscillators. These auxiliary circuits include recording level indicators and automatic level control circuits.

The recording level indicator generally consists of a small meter to which a part of the output is applied after rectification by a diode. The meter indicates the correct recording level and can also be used to indicate the condition of the batteries. A light emitting diode (LED) is also used for indicating the recording level. The LED is connected in the collector lead of a transistor to the base of which the rectified signal voltage is applied (Fig. 22.6). The LED will glow when the collector current increases due to increase in the rectified signal voltage applied to the base of the transistor.

In automatic level control, the gain of the amplifier is reduced when the output level exceeds a certain limit. A transistor which is normally biased to cut off starts conducting when the output level rises above a certain limit and the bias applied to the control transistor increases. This produces a shunting effect thereby reducing the gain of the amplifier.

22.8.4 Power Supplies

Tape recorders employing solid state circuitry as in the case of cassette tape recorders, use either 6 or 9V batteries or they use battery eliminators for mains operation. The battery eliminators are either full wave rectifiers or bridge rectifiers with suitable filtering circuits. A mains plug switch (Fig. 22.6) enables change-over from battery operation to mains operation.

22.8.5 Tape Driving Mechanism

The tape has to be driven past the various heads at a constant speed both during recording and playback processes. For playback the recorded tape must be rewound on the supply reel. This is done at a much faster speed than used for recording or playback. For editing and quick location of any desired portions of recording, the tape can be moved forward quickly by an arrangement known as fast forward (FF). A tape driving mechanisms or a tape transport system is used for moving the tape in both directions at the required speeds. Small electric motors provide the driving power and a system of driving belts, idlers, capstans, wheels and pinch rollers make other desired operations possible.

Small portable tape recorders like the cassette tape recorder operate on a fixed low speed which is generally $1\frac{7}{8}$ inch per second (4.75 cm/s) but the quality of the recording is not very satisfactory. For high quality recording of music requiring better frequency response, much higher speeds of $7\frac{1}{2}$ inches per second (19.1 cm/s) and 15 inches per second (38.2 cm/s) are used. The use of more than one motor with suitable switching arrangements, makes the tape drive mechanism in these tape recorders a very complicated affair.

One of the problems that crops up in the design of a tape transport system is that the speed of the supply reel as well as the takeup reel varies with the diameter of the tape on the reel. The speed increases as the diameter of the tape wound on a particular reel increases. This not only makes it difficult to keep the speed of the tape past the head constant but also produces tape slack if the tape runs off the supply reel faster than it would be taken up by the takeup reel. Also the tape would break if it is pulled faster than it could be released by the supply reel. The problem is solved by an ingenious device called the *capstan*. It is a cylindrical piece located between the two reels and is driven either directly by the shaft of a motor or a gear pulley, idler wheel or a drive belt from the motor pulley. The capstan has a rubberised surface which will exert the necessary torque on the tape to pull it past the magnetic heads. A pressure wheel or a pad called the pinch roller keeps the tape pressed against the capstan and minimises the possibility of tape slippage or varying tape speeds. Besides moving the tape at a regular speed in the forward direction, the tape mechanism enables the tape to be wound quickly on either reel for fast forward (FF) and rewind operations. When the tape is rewound on the supply reel, the pinch roller is moved back away from the capstan. The pressure pads also move away from the tape to allow the tape to move freely and easily from one reel to another. Figure 22.7 shows the details of the tape driving mechanism in a tape recorder.

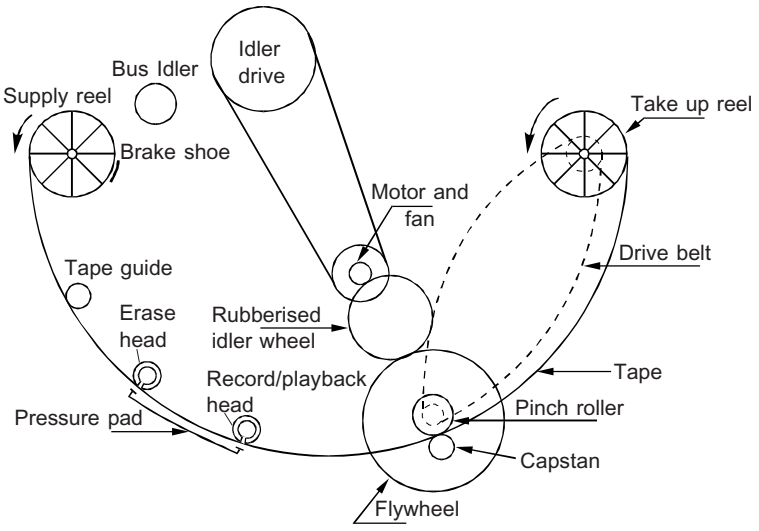


Fig. 22.7 Tape Transport System of a Tape Recorder

22.9 CASSETTE TAPE RECORDERS

A cassette tape recorder is a simple portable tape recorder in which a plastic container called *cassette* takes the place of the open type of reel or spool. The cassette contains two little spools or cores round which the tape winds and unwinds itself with the tape remaining inside the container all the time. When

the cassette is inserted into the tape recorder, that spindles on the tape transport system fit into the cores of the spools to provide free movement of the tape from one reel to another. Suitable openings provided in the cassette enable the tape to come in contact with the erase head, record/playback head and to engage with the capstan and the pinch roller etc. A single motor along with a driving belt and suitable idlers provide all the operational facilities required of a tape recorder. A built-in microphone is generally provided for recording but a socket is available for connecting an external microphone. Modern cassette tape recorders are suitable for operation from batteries or from supply mains.

A cassette tape recorder operates on the same principles of magnetic recording as any other reel-to-reel tape recorder. However, it has the advantage of compactness, simplicity and ease of operation. It is a completely portable machine and provides the simplest means of playback for prerecorded cassettes. The tape remains safe within the cassette itself without any risk of tape spilling or breakage of tape. The main drawback with a cassette tape recorder is that it is not quite suitable for the recording and reproduction of high quality music programmes. This is mainly due to the low speed of operation of a cassette tape recorder.

A cassette tape recorder operates at a fixed speed of $1\frac{7}{8}$ inches/s (4.75 cm/s) and this speed is more or less standard for all cassette tape recorders. The cassettes most commonly used are C-60 and C-90 which provide a playing time of 2×30 and 2×45 minutes at the above mentioned speed. C-30 and C-120 types of cassettes are also available. The tape used in cassettes has a width of only 0.15 inch (3.8 mm) compared to the 0.25 inch tape used in console and other portable tape recorders.

The cassette tape is divided into two tracks and only one track is used for recording and playback at a time. For recording or playback from the other track, the cassette has to be turned around and inserted again into the machine. The cassette is ejected by merely pressing a key provided for the purpose.

22.9.1 Switching

The switching arrangement in a tape recorder enables the machine to change from the recording mode to the playback mode and vice versa by effecting necessary changes in the electronic circuits. In a cassette tape recorder, piano type keys are used which, besides the changes or recording and playback modes, enable such other operations as rewind, fast forward, eject and stop. The circuit changes that take place with the operation of the record/playback switch are indicated in Fig. 22.6.

In the recording position of the switch, the microphone is connected to the input of the record/playback amplifier, the speaker is switched off and the output of the amplifier is connected to the record/playback head together with the bias voltage. The bias and the erase oscillator gets energised and the circuit for level indicator and the automatic level control are switched on together with equalising network for recording characteristics.

In the playback position the record/playback head gets connected to the input of the record/playback (RP) amplifier, the speaker to the output of the amplifier and the equalisers for the playback characteristics are brought into the circuit. The supply to the bias and erase oscillator and the circuits for automatic level indicator and the record level indicator are switched off. Figure 22.8 gives a general view of a cassette tape recorder with its switching keys and other controls.

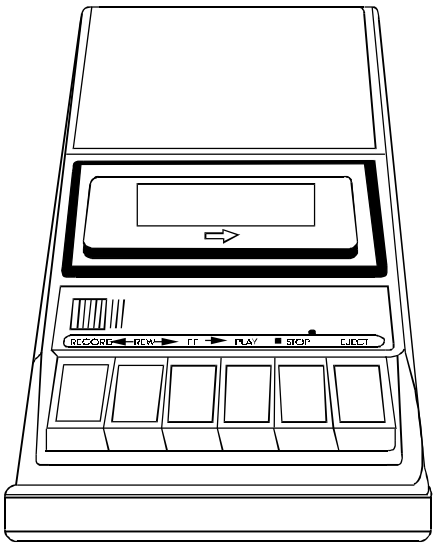


Fig. 22.8 A Cassette Tape Recorder Showing the Various Keys and Controls

22.10 TWO-IN-ONE

A two-in-one is a transistor radio-cum-cassette tape recorder. It is becoming very popular because it combines all the functions and facilities provided by a radio receiver and a tape recorder in a single compact portable unit. By a proper switching arrangement, it allows listening to a desired broadcast programme which can also be recorded on tape either with the listening going on or with the radio muted. Pre-recorded cassettes can be played back or recording from a microphone made as in a normal cassette tape recorder. A changeover switch allows the operation of the equipment as a transistor radio or a cassette tape recorder. Switches are also provided for various other functions. Push-button switches or piano key switches are mostly used for different functions and operations of the equipment.

The circuit arrangement in a two-in-one unit is such that the audio amplifier stage is common to the radio and the tape recorder. A changeover switch connects either the output of the detection stage of the radio or the output of the preamplifier from the tape section to the input of the audio amplifier. The radio or the tuner section normally consists of frequency changer, IF amplifier and the detector stage. The circuit of the tape section is identical to the circuit of a cassette tape recorder.

22.11 DISC RECORDING

The principle of disc recording is essentially the same as used for the recording of ordinary commercial gramophone records. The present day disc recording is only a refinement of the original method used by Edison who discovered a means of storing sound by converting sound pressure waves into mechanical motion and using this motion to cut a varying track in a wax cylinder. This was called the *hill* and *dale* method of sound recording. The next step was the changeover from the cylinder of the phonograph to the disc of the gramophone as we see it today. The use of modern electronic circuits has considerably increased the audio power required for the cutting tool but even with the electronic means of amplifying the energy, it remained the practice to record on wax because wax was still a very suitable material for making a *master* from which copies could be made. The method used for making the copies is known as *processing* and it consists of electroplating on the *master* wax a thin coating of copper. It is from this *master* that pressings are made and these are the commercial gramophone records sold in the market.

The method of processing takes a long time even for preparing a single copy of the record and, as such, not suited to the requirements of broadcasting where immediate playback after recording is required. The problem was to find a material for the disc which was soft enough to cut and also hard enough to withstand at least a few reproductions. The material found suitable for the purpose was a coating of a lacquer sprayed on a base of a stiff material such as aluminium, zinc or even glass. The base is merely for holding the lacquer and is not itself cut into. These discs were used for recording in broadcasting till the recording was completely replaced by magnetic tape recording which is superior to disc recording in every respect.

The essential parts of a disc recording system are:

(a) A motor to drive the turntable at a constant speed.
(b) A blank disc on which the record is cut.
(c) A cutter head (Fig. 22.9) with its chisel like sapphire cutter to convert the programme currents into mechanical vibrations.

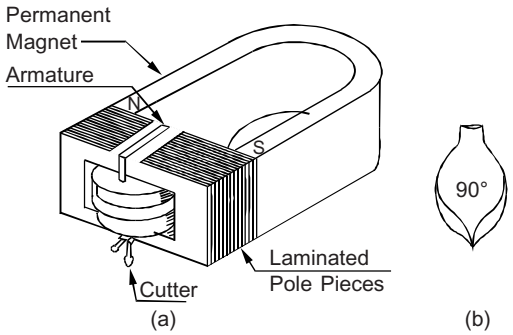


Fig. 22.9 Cutter Head for Disc Recording (a) View of a Cutter Head (b) Enlarged View of a Cutling Stylus

(d) A tracking mechanism to move the cutter head across the recorder and to produce the familiar spiral groove.

The cutter head is held in a frame so that the cutter rests on the disc with a weight of about $3\frac{1}{2}$ oz. (100 g) and this pressure (adjustable) determines the depth of the cut. A traversing arm with a lead screw (similar to that of a lathe) enables the cutter head to move from the outside edge to the centre of the disc. The circular groove is modulated laterally when the audio currents from the output of the recording amplifier are fed into the coil of the cutter head which has a moving iron balanced armature which holds the cutting stylus. The turntable rotates with a constant speed for the spiral groove to be cut into the disc. The standard speeds for rotation are 78 rpm or $33\frac{1}{3}$ rpm for long playback records.

The chief disadvantages of disc recording are the short time of playback and considerable amount of space required for storing. In order to find an answer to magnetic tape recording, which combines the advantages of long playing with ease of storing and also possesses the additional advantage that the same material can be used over and over again, certain disc recording companies have introduced long-playing records made of high grade plastic which will play up to 25 minutes. Another form of disc recording involves the use of Compact Discs (CD) which will be described later.

22.12 FILM RECORDING (OPTICAL RECORDING)

Film recording is a photographic method of sound recording mostly used in motion pictures or *talkies*. In this method, a beam of light is modulated in accordance with the variations of sound in the audio system. A photographic image of the sound modulation is formed on a sensitive photographic film which is made to move at constant speed in front of the light beam. Two methods of film recording are commonly used in motion pictures:

1. Variable Area Recording System (RCA) System

In this system, the area of the photographic film affected by the light beam varies in accordance with the sound modulation as in Fig. 22.10(a).

2. Variable Density Recording (Western Electric System)

In this type of recording the audio signals are passed through a variable area slit in such a way as to vary the intensity of light falling on the photographic film as in Fig. 22.10(b).

Both these methods require processing and are not suitable for immediate playback.

The sound track is printed on one side of the picture film.

For playback of the film recording, a concentrated beam of light is thrown on the sound track. The light emerging from the sound track is picked up by a photoelectric cell which converts the variations of light into audio currents.

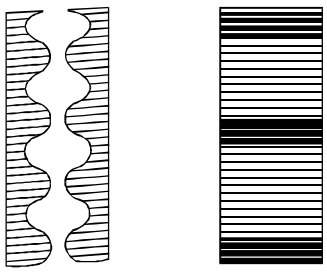


Fig. 22.10 Systems of Film Recording (a) Variable area Recording (b) Variable Density Recording

These weak audio currents are amplified by a suitable audio amplifier and reproduced by a loudspeaker.

22.13 HI-FI SYSTEMS

High fidelity (or hi-fi for short) is the name given to any audio system that can reproduce/record sound with a quality that resembles as closely as possible the original sound before recording or reproduction. For any hi-fi reproducing equipment, the performance of all the components in the reproducing chain should be of the highest acceptable standards. The essential components in the disc reproducing chain are the turntable, the pick up and the tone arm, the amplifier and the loudspeaker. A detailed description of most of these components has already been given but a brief account of their use and performance in an hi-fi system will be included here for the sake of continuity.

22.13.1 Turntable

A turntable for use in an hi-fi system usually consists of an accurately made heavy metallic disc which is rotated at a uniform and constant speed by a synchronous electric motor geared to the frequency of the AC mains. The turntable can be rotated at any fixed speed like 78 rpm, $33\frac{1}{3}$ rpm, 45 rpm or $16\frac{1}{3}$ rpm by means of different idler wheels that drive the turntable.

A good turntable should be free from *wow*, *flutter* and *rumble*. Wow is the variation in the pitch of the sound at a low frequency and flutter is the high frequency variation in sound. Rumble is the low pitched growling sound produced due to vibrations which get amplified along with the music. The hi-fi turntable must be a precision made turntable fitted with a steady and smooth running drive motor.

22.13.2 Pick-up and Tone Arm

The pick-up consisting of a cartridge fitted with a stylus enables the mechanical variations recorded on the groove, to be translated into corresponding audio frequencies. The pick-up is either a crystal pick up operating on the piezoelectric principle or a magnetic pick-up of the dynamic type which gives a very

high quality output. The pick-up is fitted at the end of a *tone arm* which guides the pick-up suitably across the record. A well balanced tone arm is necessary to avoid groove jumping, record wear and distortion.

Record changers are special type of turntables which, in addition to a pick-up and tone arm, are also fitted with a mechanism for playing a number of records automatically. In an automatic record changing mechanism, the record is selected in its turn, lowered on to the turntable and the pick-up placed in the first groove after the turntable has started rotating. The turntable automatically stops at the end of the record and the above procedure is repeated for the next record and the equipment is switched off at the end of the last record. The record changer mechanism is complicated and is often the cause of one trouble or the other. Record changers are not quite suitable for hi-fi reproduction as the quality of reproduction has to be sacrificed for the sake of automatic record changing facility. Moreover the use of long playback discs with a number of music items on each side has considerably reduced the importance of record changers.

22.13.3 Amplifier

The audio amplifier used in an hi-fi equipment usually consist of two stages—a preamplifier and a power amplifier. The preamplifier is a voltage amplifier which raises the voltage level of the input signal to a level suitable for the input of the power amplifier to develop sufficient output power to drive one or more loudspeakers. The preamplifier also contains the various controls such as the programme selection control, volume control and the tonal balance control for trebble and bass. The power amplifier should provide distortion-free output which is necessary for reproducing the entire dynamic range of a programme. The dynamic range is the difference between the column of the highest notes and the lowest whispers in a music programme. For comfortable listening in a normal living room, the output from the amplifier should be about 25 W per channel (there are two channels in a stereo system). The output required will increase with the size of the listening room. As the distortion of an amplifier increases with increase of output, the distortion figure should not exceed 1–2% at the maximum output level of any hi-fi amplifier. Another characteristic of the hi-fi amplifier is the frequency response which should be within ±1 db for the audio range extending from 20 Hz to 20 kHz. With the improved electronic circuitry and production of high quality components, it is not difficult to design and construct such high quality audio amplifiers.

22.13.4 Loudspeakers

The loudspeaker system is perhaps the weakest link in any hi-fi reproducing chain. Clarity of reproduction at the highest and lowest levels of output is the objective criterion for a good loudspeaker system. An hi-fi loudspeaker system will consists of a minimum of two loudspeakers—the tweeter and the woofer—with a suitable cross-over network to divert the high and low frequencies to the

tweeter and the woofer respectively. A third speaker called a 'SQUAWKER' is also sometime added to cover the mid frequency range. A suitably designed enclosure or baffle is necessary not only as a decorative piece of furniture but to improve the tonal quality of reproduction. Suitably designed and properly constructed high quality speakers are available to suit the requirements of hi-fi systems.

The listening room is as much a part of the hi-fi system as any of the components in the reproducing chain. The shape of the room, the acoustic treatment of the walls, the type of furniture in the room and the position of the speakers in the room play a very important part in deciding the quality of reproduction of the hi-fi equipment. The quality of even the best hi-fi equipment can be marred if the speaker system is not properly matched to the acoustic load provided by the listening room.

The source of sound for any hi-fi system could be an AM or FM radio (tuner), a tape deck or a transcription turntable as described above. All these units are sometimes contained in a single well-designed cabinet or *console*. This type of cabinet is also known as a radio phonograph or simply a radiogram. A properly designed console can give high quality sound reproduction but is quite expensive as the cost also includes the cost of the cabinet. More sophisticated consoles containing three units, radio-tape-TV are also available.

22.14 STEREO SOUND SYSTEM

A stereophonic (stereo for short) is a two-channel system of sound recording or reproduction which gives the feeling of depth and direction to the reproduced sound. Just as seeing with two eyes makes vision stereoscopic or three dimensional, hearing with two ears makes the sound three dimensional by producing the space effect. Close one eye and a stereoscopic picture becomes flat and two dimensional. Plug one ear and a stereophonic sound also loses its depth and direction and becomes a monophonic sound. Stereophonic sound system gives you the sense of participation in a music programme by enabling you to form an idea of the relative position of the musical instruments and the dimensions of the hall where the programme is being conducted. Compared to this, a single channel system of sound reproduction also known as the monoaural or monophonic system gives no idea of the point of origin of the sound.

Stereo effect is produced by the fact that listening is done by both the ears. The sound reaching the two ears from any point travels slightly different distances and is also of slightly different intensity. This difference of distance (or phase) and intensity of sound is computed by the brain to give the feeling of depth and direction which is the stereo effect. Simulating the stereo effect in actual practice will, therefore, require two separate recording or reproducing channels complete with two microphones, two amplifiers and two loudspeakers as shown in Fig. 22.11.

The sound picked up by the two microphones will be slightly different in phase and intensity and will be analysed by the brain to provide a three-dimensional picture of sound.

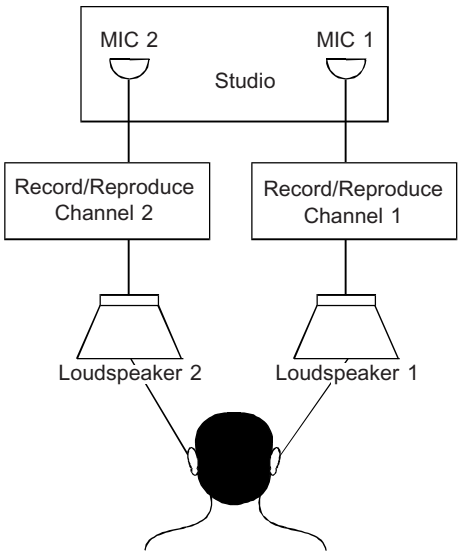


Fig. 22.11 Stereophonic Recording and Reproducing System

The fundamental difference between hi-fi and stereo must be understood at this stage. Whereas hi-fi pertains to the basic quality of reproduction of sound, stereo signifies a system of two channel reproduction to produce the three dimensional space effect in reproduction of sound. Stereo is not necessarily a substitute for hi-fi. A single channel monophonic system of sound reproduction does not become a stereo system by merely using two or more speakers in the reproducing chain. In the same way, a stereo system will not be an hi-fi system unless both the channels used for stereo reproduction are individually hi-fi channels. The conventional monophonic single channel reproduction can improve the tonal quality of reproduction to become hi-fi, but it cannot preserve the spatial character of sound which can be obtained only by the use of two channels to represent the two ears. Two-channel reproduction is, therefore, necessary for a stereophonic system. For stereo to be hi-if also, each channel must be a hi-fi channel in itself. Thus, hi-fi added to a stereo system makes it the most natural and satisfying system of sound reproduction.

22.15 PROCESSES INVOLVED IN STEREOPHONY

Several processes are involved in stereophony which is also sometimes called as natural hearing. The most important of these processes is the "time of arrival difference" at the ears as shown in Fig. 22.12.

Here sound arriving from A will enter listeners both the ears at the same instant. However, sounds arriving from B enter the right hand ear earlier than the left hand ear, thereby creating a "time of arrival difference." The brain can use this time difference to estimate the angle θ. This is due to the fact that from early childhood we learn to associate certain angles with particular time of

arrival thereby enabling us to locate the direction of the sound source. How small these time of arrival differences can be is shown by the following example.

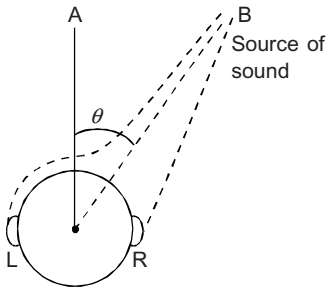


Fig. 22.12 Time of Arrival Differences

Imagine a sound coming from the extreme right. The path difference between the right and the left ear for an average human being is about 30 cm. Thus if the average sound velocity is taken about 300 meters/s, the time difference between the two ears will be approximately $\frac{30}{300 \times 100} = \frac{1}{1000}$ second or millisecond. It is also true that some people with very sensitive ear are able to detect a sound movement when the time of arrival difference is as low as 10 microseconds.

Other factors involved in the location of sound are

1. *Sound wave amplitude difference at the two ears*
   Due to diffraction (bending round the corners) effect, there may not be much difference of amplitude at the two ears at low frequencies but at frequencies above 700 Hz, there will be some amplitude difference and this can help the brain to asses the sound difference.
2. *Common sense.* This helps in knowing the direction of sound like the sound of birds coming through the window of a room.
3. *Visual clues.* If we hear music in a room we can see a loudspeaker and we assume that the sound is most likely coming from the loudspeaker. This can also be misleading.

Steady tones are difficult to locate and it is necessary that there should be some form of low frequency modulation to give the brain a chance to locate arrival times at the ear.

Factors like amplitude differences, visual clues and causes mentioned above are not very reliable and we are left with the time-of-arrival difference only. Therefore practical stereo system-must simulate time of arrival differences only.

22.16 PRODUCTION OF STEREO SIGNALS

Basically there are only two methods of producing stereo signals.

1. Using a pair of microphones.
2. Using a single microphone and electrically splitting its output into two normally unequal parts (PANPOT Method).

22.17 PAIR OF MICROPHONES

Figure 22.13 shows a pair of microphones with their diaphragms close together. Cardiod microphones have been shown in the figure but any type of microphones except omnidirectional microphone are suitable so long as their polar diagrams are as matched as possible. A mic is connected to channel A

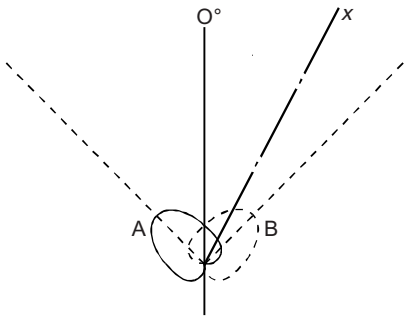


Fig. 22.13 A Coincident Pair of Cardiod Microphones

and B mic to channel B. The angle between the two microphones is shown as 90°. Of the sound approaching from the front direction O, the resultant signal in A and B will be the same and there will be no time of arrival difference at the listeners' ears. If sound arrives from X, there will be a larger amplified signal in B channel than in A and this will give the listener the impression that the sound image is somewhere to the right of the centre line. In fact such a pair of cardiod microphones provide a stereo image over an angle of 180° with a deminishing response up to about 270°. So a pair of closely spaced microphones meet the requirement of producing a stereo signal. What happens if the microphones are not closely spaced or they are not cardiod will be seen later.

22.18 ELECTRICAL METHOD

Figure 22.14 shows a single microphone whose output is fed onto a potential divider. This is a simplified diagram where amplifier and other items of equipment have been omitted. The potential divider is called a "PAN POT" or a panoramic potential divider. Clearly if the slider is in the centre of its travel, the mic output is split equally between A and B channels but it is nearer to one end than the other at Y say, the signal in B channel will be greater than in A channel and we have the required amplitude difference.

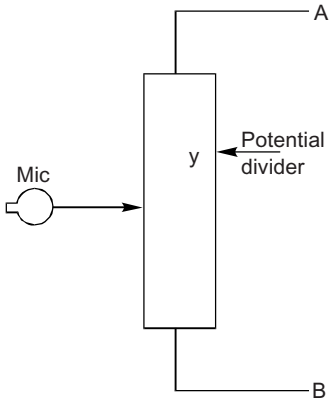


Fig. 22.14 Potential Divider Method

It is interesting to note that of the two systems discussed above there is no time difference involved and only the amplitude difference produces the required results. Both system have their practical applications.

22.19 PRACTICAL APPLICATIONS OF MICROPHONE PAIRS

There are two types of microphone pair arrangements—The coincident pair type and the spaced pair type.

(a) Coincident Pair Type A pair of closely spaced microphones is known as a coincident pair. It is the diaphragms that need to be closely spaced. Given the information of the angle between the microphone and their polar diagrams it is possible to predict reasonably well what the stereo sound image will be. The position of the stereo image will shift depending on the amplitude difference between the two channels.

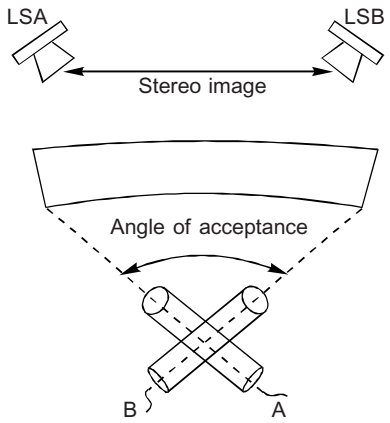


Fig. 22.15 Coincident Microphones

It can be shown by calculation and graphs that if one channel amplitude is greater that the other channel by 18 dBs or more the sound image is completely into that channel loudspeaker. Also for every 6 dB increase in the inter channel difference the image will be roughly one third away from the centre.

Given the polar diagram of the microphone pair, it is possible to predict accurately the resulting spread of stereo images. The angle of acceptance is the total sound source angle at the microphones to give a full loudspeaker to loudspeaker stereo. For example, the angle of acceptance for a figure of eight microphone pairs will be 90° at front with another 90° lobe at the rear and a pair of hypercardiod (cardiod with a lobe at the back) will have an acceptance angle of 130–140°.

(b) Spaced Pair of Microphones In this case there will be amplitude difference between A and B channels but there will also be time difference. If there is too much of time of arrival effect, the stereo image will be greatly exaggerated and the probable result is that the images will be only on the

extreme right or extreme left only with little or no central image. This gives what is known as the 'hole in the middle effect'. As microphones are brought closer together, the hole-in-the middle effect gets less obvious and may not be apparent when the separation is a meter or so. Many commercial recordings of stereophonic sound are made that way but at the cost of compatibility.

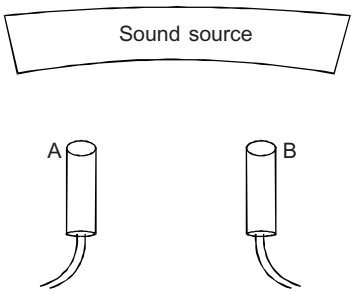


Fig. 22.16 Spaced Microphones

The hole-in-the-middle effect can also occur in domestic listening of stereophony if the two loudspeakers are widely spaced but this effect can be mitigated by using a third loudspeaker mounted centrally and fed with an attenuated addition of right and left signals.

To listen to stereophony properly, the two loudspeakers should be perfectly matched, equally balanced for volume and in phase. Ideally they should be at least 2 meters apart and equal distance from the listener, the two loudspeakers and the listener forming the three corners of an equilateral triangle.

22.20 COMPATIBILITY IN STEREO BROADCASTING

Compatibility in stereo broadcast system is as important as in the case of video or television broadcasts. Although stereo broadcasts are mainly directed towards listeners possessing stereo equipment, but it must be ensured that listeners with mono equipment should also be able to receive the stereo broadcast though without the stereo effect.

To see how the compatibility can be achieved we must first look at the variants of A and B signals. It is standard practice to refer to the sum of A + B as M and the difference A – B as S thus

$A + B = M$ and

$A - B = S$ or $B - A = S$ depending on which is greater A or B.

M stands for mono and S stands for stereo or directional component. A listener with mono equipment will receive only M without affecting the listening quality.

To see what M signal looks like involves the adding of the sensitivities of two microphones at different angles. Figure 22.17 shows two bidirectional microphones (Fig. of 8 microphone) arranged at ±45° to front axis. This is the basic BLUMLEIN pair named after A.D. BLUMLEIN whose work on two channel stereo is of historic importance.

Figure 22.17 shows the result of the addition of the sensitivities of the microphones.

It is a forward facing Figure-of-eight. Incidentally S signal is also a figure-of-eight but turned through 90°.

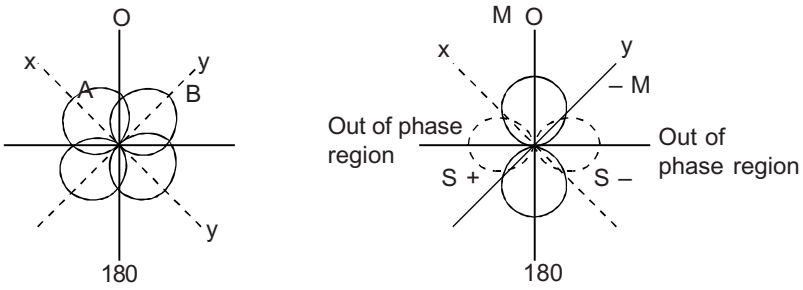


Fig. 22.17

We can now see how compatibility is achieved. Imagine there is spread of sound sources ranging from X to Y across the front of the microphone pair as shown in Fig. 22.17. The stereo image will spread from left to right loudspeaker with more or less equal loudness. The mono listener will receive only the M signal and this will be equivalent to a single figure of 8 microphone placed in the same position. Sound sources at X and Y will not be picked up as well as the central sounds. There is thus likely to be a change in the sound balance in going from stereo to mono. However, this is taken care of by the balancer on the control desk.

The problem is still there even when we use mono microphones with PAN POTS. Imagine a sound image has been panned-fully left (say). There is the A signal only. Now since.

$$A + B = M \text{ and } A - B = S$$

adding these gives $2A = M + S$ or $A = \frac{1}{2}(M + S)$. It is natural to omit 2 and simply write $A = M + S$. Microphones are available that produce M and S outputs directly.

22.21 M/S MICROPHONES

Microphones specifically designed for M/S working are usually end fire. They have the M-capsule in line with their principal axis and to produce M (Mono) signal, the S capsual for S (Stereo) complement is set at right angles across the width of the microphone as shown in Fig. 22.18.

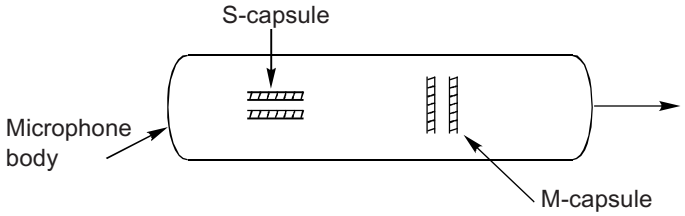


Fig. 22.18 M/S Microphone

There are some advantages of M/S pair. When the microphone is fixed on a boom in a TV, it is helpful for the operator to have one microphone pointing at the action. With a mono microphone with a boom and the M-microphone output is automatically the mono signal for the listeners not equipped with stereo.

22.22 TRANSMISSION OF STEREO (RADIO) SIGNALS

First A and B signals are converted into M and S signals as shown in Fig. 22.19.

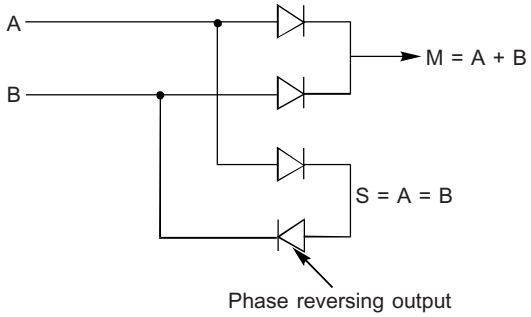


Fig. 22.19

The M-signal is transmitted normally so that any mono-receiver will detect it without much difficulty. The S-signal is employed to amplitude modulate a 38 kHz carrier. This results in sidebands which are 30 Hz, 15 kHz (audio range) above 38 kHz and the same band below 38 kHz. The S-signal now consists of a range of frequencies as given below:-

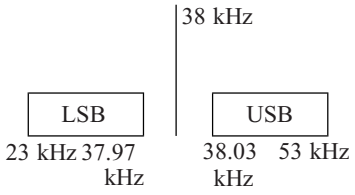


Fig. 22.20

The carrier is suppressed.

The pilot carrier of 38 kHz will be very close to the side bands 37.97 kHz and 38.03 kHz and it is difficult to design a filter for separating the carrier from side bands. The trick used is to halve the 38 kHz making it 19 kHz. This is reduced to a low level which can still be detected by the receiver and slot this pilot tone, as it is called, into the spectrum. The receiver then uses the 19 kHz pilot tone to synchronize an internal 38 kHz which produces what is in effect the missing Carrier.

Notice that 19 kHz fits neatly into a "green belt" between 15 kHz, the highest audio frequency and 23 kHz the highest side band. The 4-kHz separation is enough for the receiver to be able to filter out the suppressed carrier. The composite S-signal is then added to M-signal before being fed into the transmitter, the composite audio signal having the spectrum shown below.

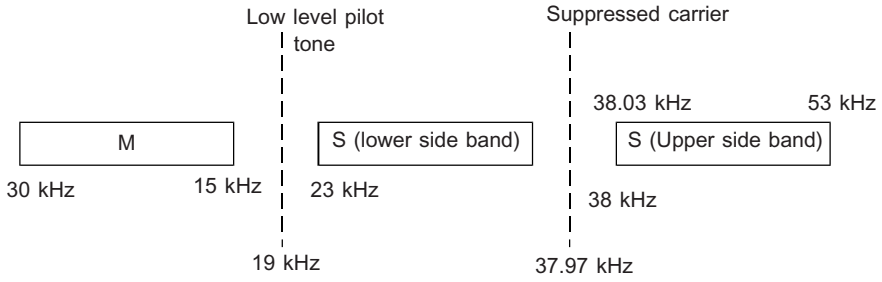


Fig 22.21 The Composite Audio Signal

22.23 STEREO RECORDING AND PLAYBACK

For stereophonic reproduction of sound, the recording of sound must also be done on a stereophonic recording system. The simplest method of stereo recording is the tape recording method where each channel can be recorded on a separate track of the tape used with a stereo tape recorder. Such a tape recorder has two separate recording channels complete with two microphones, two amplifiers and two record/playback heads. For stereo sound recording, the normal quarter inch width of the tape is divided into a number of separate tracks. Figure 22.22(a) shows a dual track tape in which the two tracks are separated by a narrow guard-band which is not coated with the magnetic material. For monoaural recording each track is recorded separately by a recording head with a narrow gap width. When one track is completely recorded, the spool or the cassette is turned around and reversed for recording on the second track. This only doubles the recording/playback time on the tape.

In stereo tape recording, the upper and lower tracks are further divided into two narrow tracks each with suitable guard bands as shown in Fig. 22.22(b). The left channel is recorded on the upper half-track and the right channel is recorded on the lower half-track with two separate recording heads. After the two upper tracks have been recorded on the stereophonic recording system, the tape is reversed and turned around and the other two tracks are recorded in the same way.

For the playback of stereophonic tapes, the playback channels must also be separate for each stereo track. Two playback heads or a dual track head picks up the sound from each track which is amplified by separate playback amplifiers and fed into separate loudspeakers. A stereophonic tape can also be played back on a monophonic tape recorder. In this case the playback head of the monophonic tape recorder scans both the stereo tracks as a single track and the

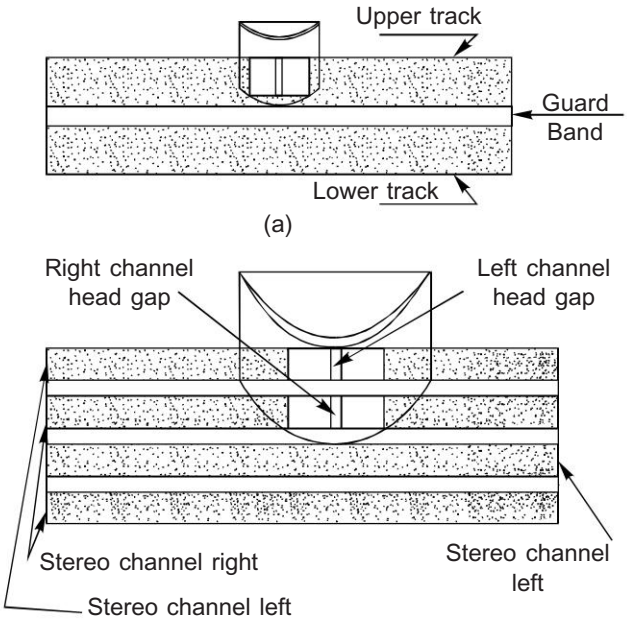


Fig. 22.22 Stereo Recording and Playback (a) Dual Track Tape (b) Tape for Stereo Recording

sound picked up by the mono playback head is reproduced by the single channel of the monophonic tape playback system. Thus, the four track system of stereo recording makes the stereo and mono playback systems compatible.

Tape recording is not the only method of stereophonic sound recording. In fact, stereophonic disc or phonograph records are now freely available. In stereophonic disc recording, the signals from each track of a stereophonic tape are recorded separately on the two walls of a single V-shaped groove on the disc as shown in Fig. 22.23. A master disc prepared by recording the left and right channels on the walls of the V-shaped groove is then used to prepare pressings or records commercially as in the case of monophonic disc recordings.

For the playback of the stereo disc recording, a special stereo cartridge or pick-up is required. This stereo-cartridge is so designed that it can indepen-

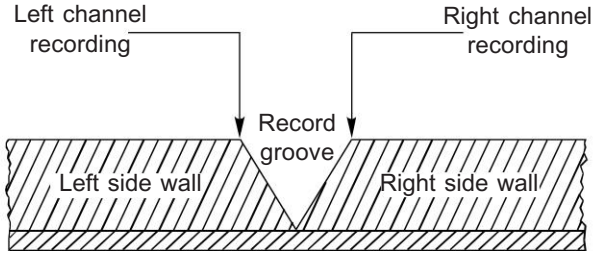


Fig. 22.23 Stereo Disc Recording

dently respond to both the lateral and vertical motions produced by the signals recorded on the side walls of a single groove on the stereo disc recording. The stereophonic cartridge acts like a pair of mini generators coupled to a single jewel stylus. The voltage produced by each generator corresponds to the resultant of the vertical and lateral vibrations of the stylus produced by the undulations on each side-wall of the stereo disc recording. The two voltages so produced are then separately amplified and reproduced by separate channels of the stereo playback system. The stereo cartridge when used with an ordinary mono disc will produce monophonic single channel sound like the mono cartridge.

22.24 DIGITAL SOUND RECORDING

An analog electrical signal is an exact replica of the original sound wave. The sound recording system that we have discussed earlier is analog recording system: The main problem with this system of recording is that any impairment or deterioration of the analogue signal caused by any reason whatever is virtually impossible to rectify. The main causes of impairment in analog recording are due to dust in the disc record grooves and bare patches in the magnetic oxide coating on the tape. Other causes of noise are the interference picked up from radio transmission and the multicopying system resorted to in tape recording system.

In digital recording the original analog signal is converted into pulses of voltage of standard height and duration. These pulses can get distorted or interfered with but as long as these can be identified as some sort of a pulse, these can be reconstructed and consequently converted back into a good analog signal.

22.24.1 Analog to Digital Coversion

The following steps are involved in the conversion of analog to digital signal.

1. Sampling The analog signal is split up into a large number of samples, the height or amplitude of which is measured at frequent intervals.

The sampling rate should be at least twice the highest audio frequency in the signal. The generally accepted rule for the sampling frequency *fs* is:

fs = 2.2 *fmax* where *fmax* is the highest frequency to be sampled. If the highest audio frequency to be supplied is 15 kHz, then the sampling frequency should be 2.2 × 15 kHz = 33 kHz.

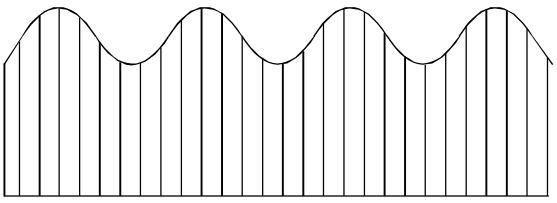


Fig. 22.24 Sampling

However, the universal standards for sampling are 44.1 kHz and 48 kHz, 44.1 being the standard for domestic recording and 48 kHz is the professional standard.

2. Quantizing The process of measuring each sample involves comparing it against a scale consisting of a number of discrete values called *Quantizing Levels*, each of which is assigned a reference number to represent its value. As the analog signal can have values which vary continuously over the whole range of measurements, the actual value and the quantized level will seldom coincide exactly.

The differences between the sample levels and the quantizing levels can be heard in the reproduced sound and is known as the Quantization Noise.

The more accurately the quantizing is done the better is the quality of reproduction. Good quality recording will need about 65,000 quantizing levels. For example 64 quantizing levels will give a S/N ratio of only 25 dB whereas 65536 quantizing level provide a S/N ratio of 85 dB.

3. Coding-Binary Code The magnitude of quantizing levels clearly shows that a digital tape or disc must store numbers up to 65,000 and that too at a high rate of 44,000 per second. This is a stupendous task but the problem is solved by the use of binary code—a code consisting of 1s and 0's.

After the quantizing process the resulting numbers are converted into binary, so that a number like 17,832 representing a sample of about a quarter of the height, when converted to binary becomes 100010110101000. Now, if each 1 represents a pulse of voltage, we then have a signal like the one shown in Fig. 22.25.

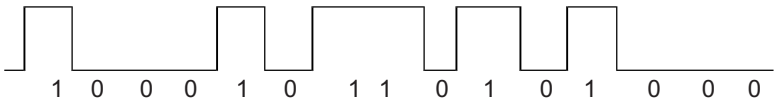


Fig. 22.25 A Binary Signal

In this type of signal individual pulses can be degraded but provided they are still just identifiable they can be reconstructed, if necessary. Each 1 or 0 in digital audio is called a bit which is short for *binary digit*. Figure 22.25 above indicates a 15 bit number.

Bit rate (bits per second) is calculated from the formula:-

Sampling frequency × number of bits/sample.

Thus a 10 bit system using 32 kHz sampling will have a bit rate of 10 × 3200 = 320 kb/second. Bit rate is of great importance in deciding the parameters of a digital system. The number of bits gives an indication of the quality of the system. CDs (Compact Discs) use 16 bits but telephone communication can be satisfactory with fewer than 10.

Error Detection Both the analog and digital systems are subject to interference during the process of recording. In the analog system there is no real effective method of detecting when error of any kind has crept in. The replay

equipment is not able to detect the cause of clicks heard in the replay although certain 'de-clicking' devices can reduce the effect to a small extent. It is somewhat different with digital audio.

Serious blemishes in digital audio reproduction could be caused by drop-out on magnetic recording, dirt on CDs, there can produce some starling effects but these can be identified by a process known as *Parity checking*. In Parity condition, the signal to be recorded or transmitted is subjected to a parity check. In this the number of 1's in the digital signal is counted and if necessary, made up to an even number by adding a parity bit. A 0 is added if there was already an even number of 1s and a 1 is added if there was an odd number of 1's. Thus every sample recorded or transmitted contains an even no of 1s.
Example Suppose the original sample is a 10-bit sample represented by 0011100101. The number of 1s is odd (5) and a parity bit (1) is to be added. The new sample becomes 00111001011. Similarly 1100110011 is a 10 bit sample with even no of 1s (6) and no parity bit but only a 0 is to be added. The new sample will appear as 11001100110. These might increase the bits/second but this is not significant.

22.24.2 Digital Recording and Playback

The main problem with digital recording is the enormous bit rate that is required to be recorded on tape compared to the analog system. Whereas the conventional tape recorder has to handle frequencies up to 20 kHz, in digital recording the maximum frequency will go up to 1.4 MHz. Thus the frequencies to be recorded are in the ratio of 1 to 70 and the tape speed in digital recording will have to be speeded up to 70 times which is not possible with existing spooling motors. There are two approaches to the problem and accordingly two types of digital recording machines are in common use.

1. R-DAT In this machine the problem of high tape speed is solved by using the relative speed between the tape and the head by moving the head rapidly rather than the tape moving rapidly. This method of obtaining high relative speed between the head and the tape is similar to the method used in video tape recording when the frequencies involved are as high as 5 MHz.

In such mechanisms the recording heads are mounted in a drum which rotates rapidly in a transverse manner across the tape which moves slowly. The axis of the drum is tilted slightly so that when the tape passes in front of it the recorded tracks are at a slant across the tape. The compact unit provides an encoder for analog to digital conversion (ADC) and also a decoder for digital-to-analogue conversion. R-H-DAT which is an abbreviation for Rotary-Head-Digital Audiotape uses a very small cassette, even smaller than an analog audio cassette. R-DAT is mainly meant for professional use only because editing with this machine is a complicated business.

2. Dash System A Dash (Digital Audio—Stationary Head) machine looks very much similar to an analogue one. The tape width and the tape speed are also similar to analog tape recorder. Here the tape moves relatively slow but it

is of a special type which fits smugly round the head. These several tracks carry different parts of the data and this is equivalent to an increase in the tape speed. These machines are very expenssive but this is compensated by the fact that editing with the machines is very simple. The tape can be easily cut and sliced as with an analog tape. A good example of digital recording and playback is a Compact Disc (CD)

22.25 COMPACT DISC (CD)

The first digital audio recording medium was a compact disc (CD) which was developed in early 80's. In a CD the digital recording is made by an optical process in which a *laser* beam is modulated with the digital audio to be recorded and focussed on to a rotating master plate. Laser beams are used for recording because these beams contain only one wavelength and can be accurately focussed compared to ordinary light which contains many wavelengths and cannot be easily focused with a lens. The master plate is coated with a suitable material which is etched by the intense laser beam to produce a pattern of dots corresponding to the digital signal. Accurate copies of the master are then made by a stamping process similar to the one used in making vinyl audio discs. In the replay process, a low power laser is focussed on to the track and the light reflected back is detected and converted into electrical signals. After suitable recording we obtain a very high quality audio signal. Figure 22.26 shows the optical process involved in the playback system.

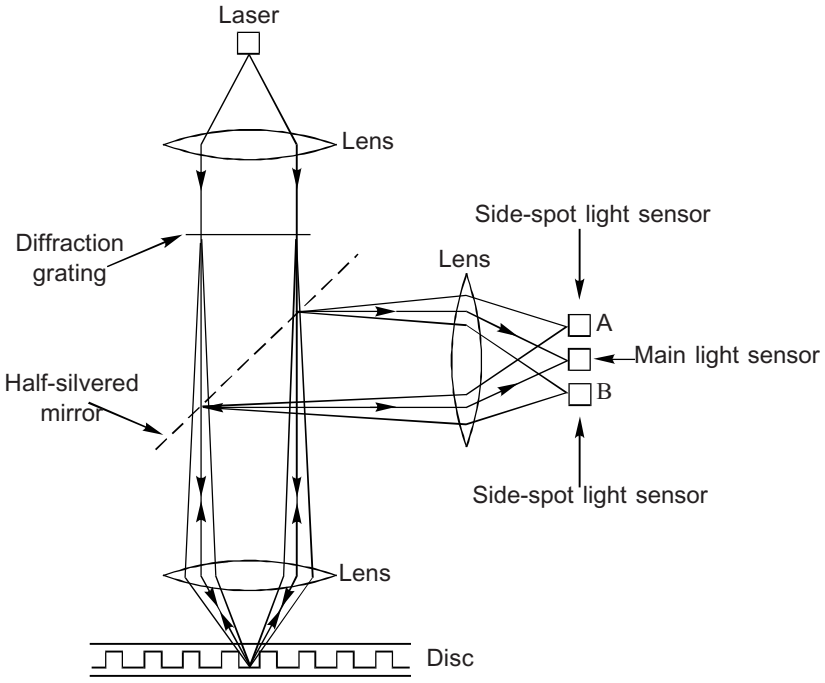


Fig. 22.26 Optical Process in a CD Player

CD Parameters CD format has many advantages like robustness, compact quick to use, offering random access, higher duration of recorded programme over the vinyl audio disc. A CD offers almost perfect reproduction of the recorded material with a frequency response within ± 0.5 dB from 20 Hz to 20 kHz and a dynamic range. Signal-to-noise ratio and separation between channels, all better than 90 dB. As it is a digital system controlled by a quartz crystal, wow and flutter are virtually non existent. Since there is no physical contact between the disc and the reproducing system, there is no deterioration of the quality due to wear. A CD is covered with a protective layer and therefore, the effect of dust and scratches is minimised. Further a powerful error correction system which can correct even large burst errors makes the effect of even severe disc damage insignificant in practice.

A standard Compact Disc is 12 cm in diameter and can provide a maximum playing time of 75 minutes (60 minutes normal) compared to playing time of only 30 minutes provided by a conventional LP (long play) disc.

CD has now become an international standard of audio quality. All India Radio has introduced CD players in a big way and at present most of AIR stations are using indigenously developed CD players.

22.26 VIDEO TAPE RECORDING SYSTEMS

1. Historical While audio tape recording was making rapid strides towards perfection, a serious thought was being given by the manufacturers to the possibility of recording TV video signals on tape. The main problem that confronted the designers of video tape recording machines was the relatively large bandwidth (4 to 5 MHz) of the TV signals to be recorded. Recording these high frequencies on tape necessitated the use of very high tape speeds and very narrow head gaps. Even with these limitations, video machines for the broadcast of TV signals were designed by BBC and Decca in Britain and by RCA and Ampex in the USA. In 1955, BBC used their VERA (Vision Electronic Recording Apparatus) for telecasting TV pictures. The tape speed was nearly 1000 cm/second and the head gap 20 microns. The machine designed by RCA used a tape speed of 600 cm/second and three separate tracks were used for simultaneous recording of R, G and B signals of the colour TV. All these machines using the longitudinal system of recording were cumbersome, difficult to operate and maintain besides being extremely wasteful of magnetic tape.

The limitations and problems of longitudinal video recording almost completely vanished with the revolutionary idea that the 'writing' speed of a tape recording system is actually determined by the *relative* head-to-tape speed and not merely by the speed of the tape past the stationary head. According to this idea, very high writing speed could be attained by rotating the head at a high speed across the width of the tape which moves at its normal longitudinal speed as shown in Fig. 22.27.

The rotating head could record a complete TV field on one pass and another field on the next pass and so on. The difficulty of wrapping the head completely

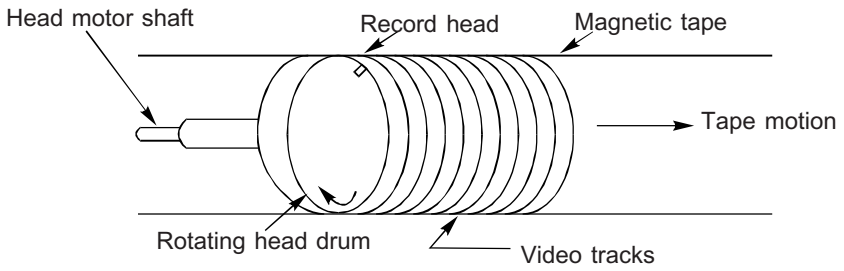


Fig. 22.27 Transverse Scanning Method with a Single Rotating Head

around the head drum was solved by the use of more than one head in a rotating drum assembly. In this multihead system, successive fields could be recorded by successive heads.

Based on the concept of a rotating head, Ampex in 1956, designed their, quadruplex transverse track recorder using four heads which spun across a 2-inch (5 cm) wide tape at a very high speed of 240/250 rps, while the tape itself moved longitudinally at 15 rps. The writing speed of the system was adequate for the monochrome TV. The quad system of video recording was immediately adopted by various TV organizations and with many refinements over the years, it still finds a place in many TV studios. Similar machines are also being produced by RCA and some other manufacturers.

Two problems that still perplexed the designers of good quality video recorders were:

1. The big difference between the audio range of frequencies(20 Hz to 20 kHz) and the video range of frequencies (25 Hz to 5 MHz) prevented the normal audio recording methods to be employed for video recording.

   The problem was solved by frequency modulating (FM) the video signal on a high frequency carrier before the signal is actually applied to the recording head. The frequency translation of the video frequencies up the frequency spectrum scale reduces the ratio of the highest to the lowest frequency without affecting the frequency spread. Full details of this method will be given later in the text.
2. The transverse tracks in the quad system were not long enough for the heads to record one field in one pass. This was achieved by what is known as the helical scan or the slant track system. The tracks make a slight angle with the vertical and are longer than the transverse tracks. This enables the head to record one complete field on each pass. This simplified both the switching arrangement and the complicated electronic circuitry. All VCRs now use the helical scan system.

Even with the refinements mentioned above, the video recording machines produced were of reel-to-reel type and not really suitable for domestic use. However, efforts were being directed in Japan to produce a simpler video cassette recorder similar to an audio cassette recorder. In 1970; Sony introduced their 3/4 inch. U-matic cassette system which has now reached a high

level of professional standard. Earlier, Philips had introduced the 1/2 inch. Cassette recorder based on the VCR system. This was soon replaced by their latest Philips/Grundig V2000 system using a two-sides cassette with 0.5 in tape—only 0.25 in of the tape is used on each pass, and the system provides a 2×4 hour playing capability. The VCC (Video Compact Cassette) system of Philips could not capture the consumer market mainly because it coincided with the colour TV boom.

Two significant developments that took place in Japan in 1975 were the introduction of the Betamax format by Sony and the VHS (Video Home System) format by JVC (Japanese Victor Company).

These cassette machines are based on the standards laid down by EIAJ (Electronic Industries Association of Japan) in 1968. VHS format soon became very popular in most countries of the world including India, with Betamax in the second place. Today the three most popular domestic systems in the field of VCR are: the VHS (JVC), the Betamax (Sony) and the V2000 (Philips). Details of these systems will be included in the following paragraphs.

22.27 VIDEO RECORDING

The main difference between audio and video frequencies is the dynamic range in the two cases. The dynamic range, which is generally measured in octaves, extends from the smallest signal that can be detected above noise to the largest signal that can be handled. The audio range of frequencies extending from 20 Hz to 20 kHz covers a span of about 10 octaves. (20 Hz to 20,480 Hz represents 10 octaves.). This dynamic range of 10 octaves is the maximum range that can be conveniently recorded directly on a modern tape without magnetic saturation of the tape.

The video frequencies cover a much wider range of frequencies than the audio range mentioned above. The video information extends from dc (0 Hz) to about 5 MHz and contains many more octaves than the audio range. For example, if the video range is taken to extend from 25 Hz to 5 MHz, this will represent an octave range of 18 octaves. It is not possible to record this high dynamic range directly on the magnetic tape. Moreover, the tremendous difference between the highest and the lowest frequencies in video recording makes it difficult to design a head gap that will reproduce equally efficiently, the high and the low frequencies in the entire video range. Some of the methods adopted to solve these difficulties are described below.

Reducing the Number of Octaves-Modulation A very ingeneous method is adopted to reduce the number of octaves in the video frequency range. The method employed is called modulation. In modulation, the signal is impressed upon a high frequency carrier, resulting in the production of new frequencies called sidebands. The sideband resulting from the sum of the carrier frequency and the signal frequency is called the Upper sideband and the sideband produced by the difference of the two frequencies is called the Lower sideband. By modulation, the modulating signal which normally lies on the lower end of

the frequency scale is translated into a signal occupying a higher range in the frequency spectrum. This considerably reduces the ratio between the highest and the lowest frequency in the modulated signal. As an illustration, if a carrier of 10 MHz frequency is modulated with a video signal occupying a bandwidth of 5 MHz, the two sidebands produced will be 10 + 5 = 15 MHz and 10 – 5 = 5 MHz as shown in Fig. 22.28. The ratio of the max. to the min. frequency is 15/5 = 3 and this represents less than 3 octaves. The frequency spread remains ±5 MHz above and below the carrier frequency. Recording less than three octaves is well within the capability of the modern tape recording system. Moreover, a head gap designed for 7.5 MHz can easily reproduce 10 and 5 MHz which are the limits of the frequency spread. This cosiderably reduces the need for equalization.

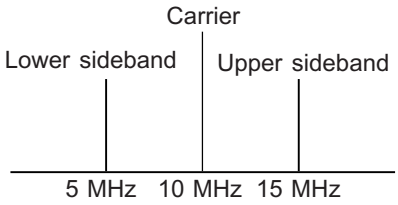


Fig. 22.28 Modulation Sidebands

Frequency Modulation (FM) Of the two methods of modulation available, the method employed in video recording is frequency modulation (FM) in preference to amplitude modulation (AM). In frequency modulation, the amplitude of the carrier remains constant but its frequency is varied in accordance with the intensity of the video signal and at a rate corresponding to the frequency of the video signal. FM is used because of its many advantages over AM, particularly in the realm of noise reduction. Also the amplitude fluctuations due to erratic contact between the tape and the head do not affect reproduction in FM and dropout problems are also minimised. These points will be discussed in detail at a later stage.

Head Gap Even after the problem of reducing the octave range of the signal is solved by FM, extremely small head gaps required for recording frequencies in the lower megahertz range present their own problems. One solution to this problem is to increase the tape speed and thereby make the wavelength of high frequencies longer, so that they fall within the reproducing capabilities of practical head gaps. However, increasing the tape speed means use of excessive tape length. With a head gap of 0.11 mil, the tape speed required will be 1000 inches/s for proper playback of a 5 MHz signal. With this tape speed, the tape length required will be 300,000 ft per hour. The size of the reels required for winding the tape will also be very big. All these factors will make the entire system cumbersome, expensive and unreliable. To solve these mechanical problems of longitudinal video recording, many other methods like splitting the video signal into a number of sections were tried. In fact,

the VERA machine developed by BBC and DECCA was successfully used in 1955 for the broadcast of TV signals.

The longitudinal recording had to face many problems, including the control of very high tape speeds. Besides being wasteful of tape length, the machines had a restricted bandwidth and would not allow long playbacks as required for feature films and sports events. The solution lay in increasing the writing speed by methods other than increasing the speed of the tape.

Writing Speed The writing speed of a recording system has been defined as the relative speed between the magnetic tape and the recording head. For recording very high frequencies and bigger bandwidths, the writing speed of the system must be very high. Increasing the writing speed by increasing the tape speed past the stationary head is fraught with many difficulties. One method of increasing the writing speed that completely revolutionised video recording was to rotate the recording head at a high speed across the tape in a transverse manner. Even with the tape moving longitudinal at its normal speed of 15 inches/s or 7.5 inches/s, a very high tape-to-head speed could be achieved. This method with certain modifications is the method that is now used in most video recording machines including video cassette recorders.

If a single rotating head is used, the tape has to be wrapped completely around the rotating head drum as in Fig. 22.29(a) so that the head is in contact with the tape almost for the entire revolution.

The speed of rotation is so adjusted, that the single head makes a complete revolution in 1/50 or 1/60 s, depending on the scanning system used. When so timed, the rotating head will record a complete TV field in each of there rotations. The loss of signal due to the gap in time when the head leaves the tape is so arranged that it occurs only during vertical blanking of the TV signal.

An improvement on the single-head scanning system is to use a two-head scanner in which the two video heads are mounted 180° apart. The tape has to be wrapped around the drum for a little over 180° as in Fig. 22.29(b). The speed of rotation is such that each head records a complete field during half the revolution when it is in contact with the tape. There is no time loss or gap in the recorded signal because each head is in contact with the tape for more than 180° and the heads are fed with the signal simultaneously. There is an overlap

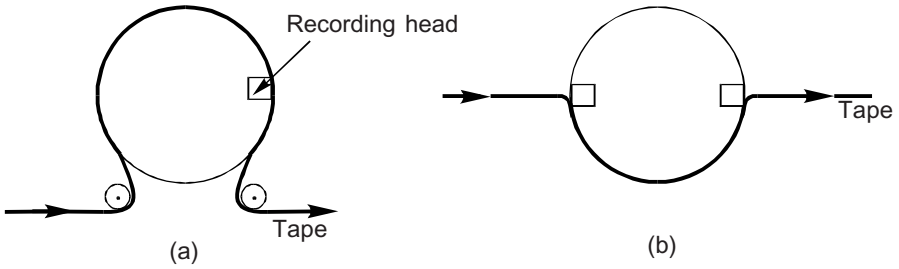


Fig. 22.29 Number of Heads and Wrapping Systems (a) Full Wrap, One head (b) Half Wrap, Two Heads

period during which the same information is recorded by both the heads. This overlap is taken care of during playback when a switching arrangement effects change over from one head to another.

A commercial video broadcasting machine using four heads mounted at right angles to each other on a drum rotating around a horizontal axis was designed by Ampex in 1956. The quadruplex or quad machine used a 2-inch wide tape moving longitudinal at 15 inches/s. The four recording heads spinning across the tape at a very high speed, produced a sufficiently high writing speed to record a full 5 MHz bandwidth required for colour TV. One hour of programme could be recorded easily on a tape length of 4800 ft. This machine was soon adopted by most TV broadcasting organizations, and with many modifications, is still the standard VTR machine used in many TV studios. The need to equalise and match four channels and the complicated switching circuits required for the four heads make it a system not suitable for domestic use.

Helical Scan Recording To make the VTR machines suitable for other than the professional broadcasting purposes, another system of scanning, called the helical scan system was introduced around 1960. In this system, the tape is wrapped around the head drum in the form of a helix. The heads are mounted on a drum which rotates round a vertical axis and the heads cross the tape at an angle from edge to edge as shown in Fig. 22.30.

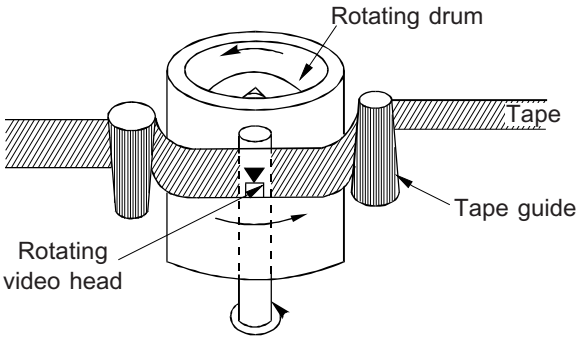


Fig. 22.30 Helical Scan Recording

Each rotating head records a slanting track on the longitudinal moving tape. The angle of the track depends on the diameter of the drum and the speed of the tape. The rounded tracks are longer and it is possible to adjust the speed of the rotating drum so that each head records a complete TV field on every pass of the head across the moving tape. With two rotating heads, each head will record its own field and the writing speed will increase. Since a complete field is recorded on each pass of the head, this simplifies the switching arrangement and other complications associated with a quad system already described.

A helical scan system is shown in Fig. 22.31. The audio and control tracks are recorded by stationary heads as in audio recording. Guard bands were initially used to separate one video track from another to avoid crosstalk.

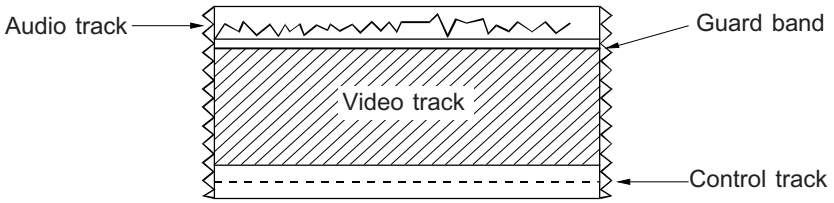


FIg. 22.31 A Helical Scan System

Guard bands were also used to separate audio and control tracks from video tracks. Methods used to get rid of guard bands to save tape space will be described at a later stage.

Video Cassette Recorders (VCRs) The helical scan system actually brought the video recording within the reach of the domestic user. The cumbersome reel-to-reel mechanism soon resulted in the introduction of video cassette recorders (VCRs) for domestic use. By placing the tape in cassette to protect it from damage by dirt and unskilful handling, the video cassette recorder became safe for domestic use. Further, by making the tape in the cassette capable of lacing and threading itself automatically, the VCR became a simple item of domestic use, like a camera or a washing machine.

22.28 VIDEO TAPE TRANSPORT SYSTEM

It has already been explained that audio tape recording and video tape recording are only two different forms of magnetic tape recording, the basic principles being the same. In both cases, the magnetic tape is driven past a series of heads at a constant speed, which is kept under control by the capstan and pinch-roller combination, helped by various other control devices. In the case of video tape recording, however, a high writing speed is required and this is achieved by rotating the video heads at a high speed across the horizontally moving tape. Very high writing speeds are necessary in the VTR because of the high frequencies and large bandwidths involved in the video range. For the recording/playback of frequencies in the megahertz range, the speed control must be of microsecond accuracy. The path followed by the tape in the record and playback modes must be exactly the same for satisfactory results. Moreover, the rotating video heads must maintain a correct spatial relationship with moving tape and a precise time-relationship with the video signals being recorded. This calls for a mechanical precision of the highest order and a very complex and sophisticated electronic circuitry. All machines of the same type must have the same mechanical precision for interchangeability of programmes.

For a clear understanding of the various mechanical processes and operations involved during recording and playback of the video signals, it will be interesting to follow the tape on various stages of its journey from the supply spool to the take up spool. For this purpose, a diagrammatic view of a VTR tape deck is shown in 22.32.

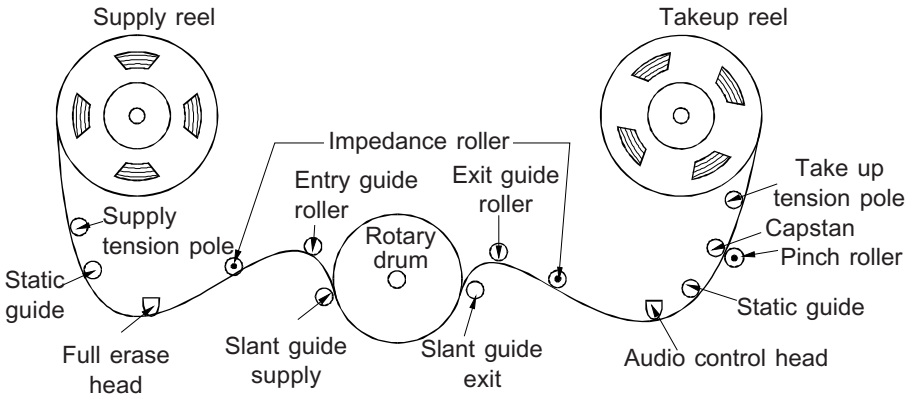


Fig. 22.32 Tape Transport System of a VTR

Most of the tape is wound on the supply reel and is pulled towards the take-up reel at a constant speed. In a VCR, both the supply reel and the take-up reel are enclosed in a plastic cassette.

Supply Tension Pole The tension of the tape is checked up and adjusted by the supply tension arm on emergence from the cassette. A braking system on the supply spool keeps the tape tension constant. This could either be a mechanical device or a servo control system. A similar tension pole adjusts the tension before the take up reel.

Static Guides These are similar to those available on an audio deck and help in the smooth movement of the tape.

Full Erase Head This erase head is meant to wipe the tape clear of all recorded tracks including the video, audio and control signals. Its head gap is large enough to span the entire or most of the width of the tape. The erasing is done by means of an RF current supplied to this head from a separate oscillator. In most machines, this oscillator also supplies the erase signal to audio and control heads. The full length erase head is energised only during the record mode.

Impedance Roller This is meant to smooth out any speed or angle fluctuations that might develop during the passage of the tape over the tension arm and the erase head, etc.

Tape Guides These are very accurately machined guide posts which enable the tape to take the correct slanting path around the head drum. These tape guides are located both, in front and immediately after the drum assembly. There are two types of tape guides—the slant guides and entry and exit guides as shown in Fig. 22.33. The slant guides are cone shaped poles which are slightly inclined to the vertical to create slightly more tension on the top than at the bottom. This is to enable the tape to sit firmly on the roller edge around the video head drum.

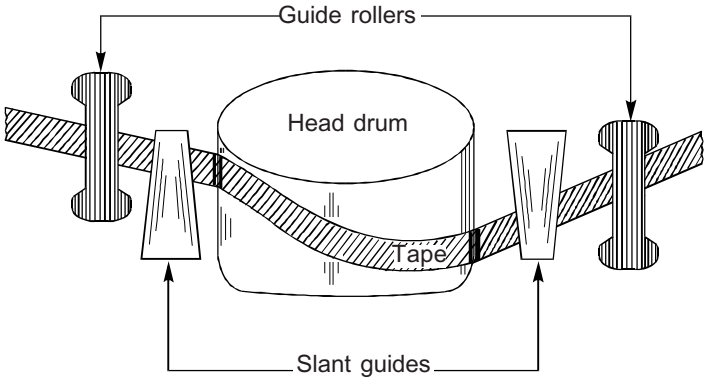


Fig. 22.33 Tape Guides

Entry and exit guide rollers are located immediately before and after the tapered guides. These are straight vertical poles with narrowed sections to accommodate the tape. These are set at an exact height to feed the tape to the slant poles and the drum at the correct height for scanning. The tape guides are very precisely machined components and must be identical in all machines of the same make. This is necessary for compatibility or interchangeability of tapes from one machine to the other of the same type. The tape guides are integral parts of a scanning drum assembly and are adjusted in the factory. These should never be disturbed.

Head-Drum Assembly The head-drum assembly, shown in Fig. 22.34, holds the rotating heads in a correct rotating plane. The drum assembly also supports and guides the magnetic tape around its circumference in the correct path. The two rotating heads are mounted on a rotating platform or disc which forms a part of the upper half of the drum assembly. The heads generally protrude a little beyond the surface of the rotating drum and so rotate in a diameter slightly greater than the drum diameter. The tip-protrusion must remain constant throughout the motion of the drum. The head-drum assembly calls for the highest degree of mechanical perfection which must be the same for all machines of the same type for compatibility. The head-drum assembly also supports various other components such as the tach coil or other sensing devices for the servo control of the head rotation.

The entire head-drum assembly can be removed or replaced. Changing the heads or tach sensors is the only repair work required for a head assembly besides cleaning the heads and tape guides as described later in this book. For a correct head wrap angle of the tape in helical scanning, the head drum itself is tilted at an angle of about 5° with the vertical.

Because of the large area of contact between the smooth surface of the tape and the drum, there is a tendency for the tape to stick to the drum. To avoid this phenomenon of "sticktion", air is blown through the head assembly from below to provide a form of lubrication between the tape and the drum.

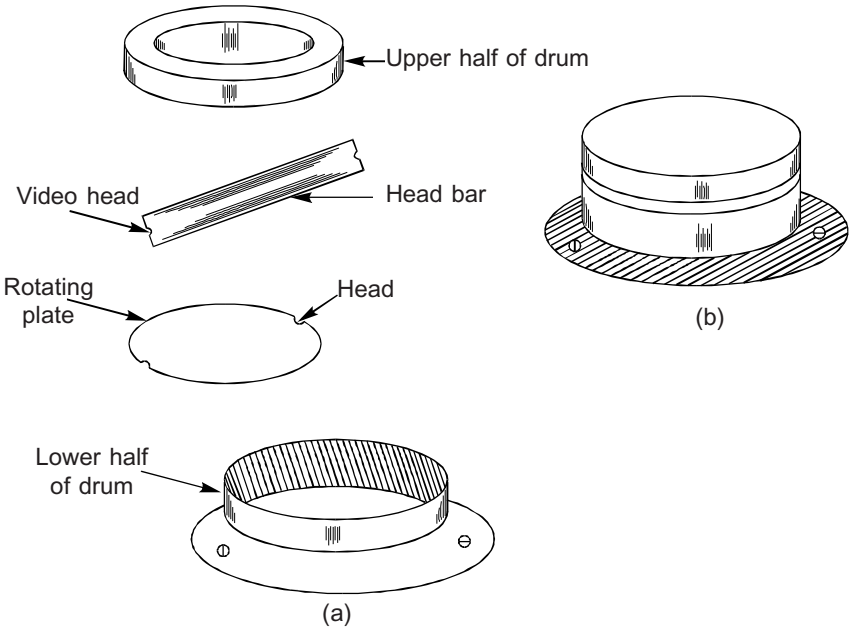


Fig. 22.34 Head Drum Assembly; (a) Exploded View (b) Assembled View

Video Heads The video heads are so mounted that they rotate in a horizontal plane and protrude a few mills from the surface of the head-drum. The head-drum is divided into two parts. The lower half is generally fixed to the chassis of the machine. The heads are mounted on a rotating horizontal plate (Betamax), attached to a central shaft that is driven by the head motor. The heads are mounted on a rotating drum in the case of VHS machines. The signal to and from the rotating video heads is carried either by slip rings and wire leaf springs or by a rotating transformer with stationary primary winding and the secondary winding rotating with the head-plate as shown in Fig. 22.35. The rotary transformers have the advantage of being free from mechanical wear or electrical noise.

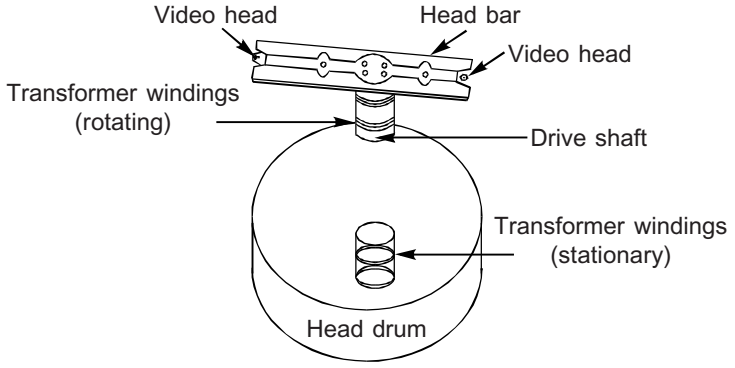


Fig. 22.35 Rotary Transformer

On leaving the head-drum the tape again passes round the exit, tapered guide to ensure that it is correctly seated round the roller edge. Passing round the exit guide roller, the tape is steadied by the impedance roller. Next comes the audio/control head which is a single block containing the audio and control heads. The gaps of these heads are so lined up that they scan the opposite edges of the tape. Another static guide keeps the tape properly aligned before it reaches the prime mover of the tape which is the capstan.

Capstan and Pinch Roller The capstan is a steel roller that actually drags the tape past the various heads and guides. The tape is held tightly against the capstan by a pinch roller or pressure roller actuated by a spring-loaded cam. The pinch roller is automatically disengaged during fast transport modes, stop and pauses. The capstan is generally belt-driven through a flywheel which is carefully balanced for smooth running of the capstan without introducing any flutter or wow. The pressure roller is made of synthetic rubber and has a mat surface to provide a good contact with the tape and the capstan. Leaving the capstan, the tape further enters the cassette again to be wound round the take-up spool which is gently driven by a slipping clutch or a direct driven motor.

For compatibility, all machines of the same type must conform to the same arrangement of components on the tape transport system. Even the slightest variation in the alignment of these components can lead to disastrous results.

Tape Threading In passing from the supply reel to the takeup reel, the tape has to pass through a number of tape guides, rollers and the capstan assembly besides being wrapped round the head-drum. The path followed by the tape is quite complicated and is different in different types of machines. The process by which the tape is made to go round the specified path is called tape threading. In reel-to-reel VTR machines, tape threading is done manually which not only requires expert handling but also involves the risk of damage to the tape. In the case of domestic VCRs, the entire tape threading is an automatic and fool proof process requiring no handling of the tape. As soon as the cassette containing the tape is loaded into the machine, a mechanism is set into operation by which a loop of tape is automatically pulled from the cassette and taken round the prescribed path only in a fraction of a second. Each system or format uses its own method of tape threading. This makes it necessary for tapes recorded on one machine to be played back on the same type of machine only. Various methods of tape threading have been adopted by different manufacturers but only two of these methods which are commonly used in domestic VCRs will be described.

VHS Method In the VHS method shown in Fig. 22.36, the tape is wound around the head drum in the shape of the letter M and the threading system is known as the M wrap. When the cassette is dropped into place by the loading process, double set of guide pins and guide rollers penetrate into the cassette shell and get behind the tape as shown in Fig. 22.36(a). As soon as the threading process is initiated by pressing the play button, the two moveable arms

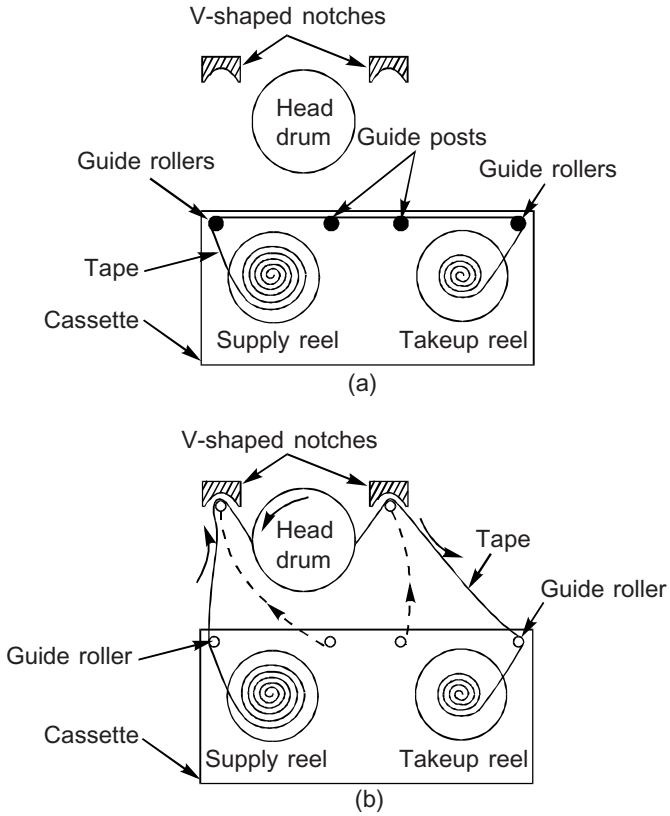


Fig. 22.36 VHS Tape Threading; (a) Before Start of Threading (b) After Completion of Threading

holding the guide pins swing into action and push the guide pins upwards. In this process, a loop of tape is pulled out from the cassette and wrapped around the head-drum. At the end of their travel, the guide posts get firmly seated in the V notches fixed on either side of the head-drum. The threading process is completed with the guide posts serving as guide rollers for the tape path as shown in Fig. 22.36(b).

Betamax System In the Betamax system of tape threading, a loop of tape is wrapped round the head-drum by means of guide posts fixed on a motor-driven threading ring. Figure 22.37(a) shows the position of the tape and the threading ring before the start of the threading mode. There are three guide posts beneath the cassette. Two of these posts are fixed to the threading ring and the third one is attached to the end of a jointed arm which is itself fixed on the threading ring. As soon as the cassette is dropped in position, a microswitch sets the threading operation into motion. The motor-driven threading rings starts moving anti-clockwise, carrying with it a loop of tape which gets wrapped around the head-drum. When the threading ring has completed nearly three-quarters of a turn, another microswitch acts and stops the ring in its final

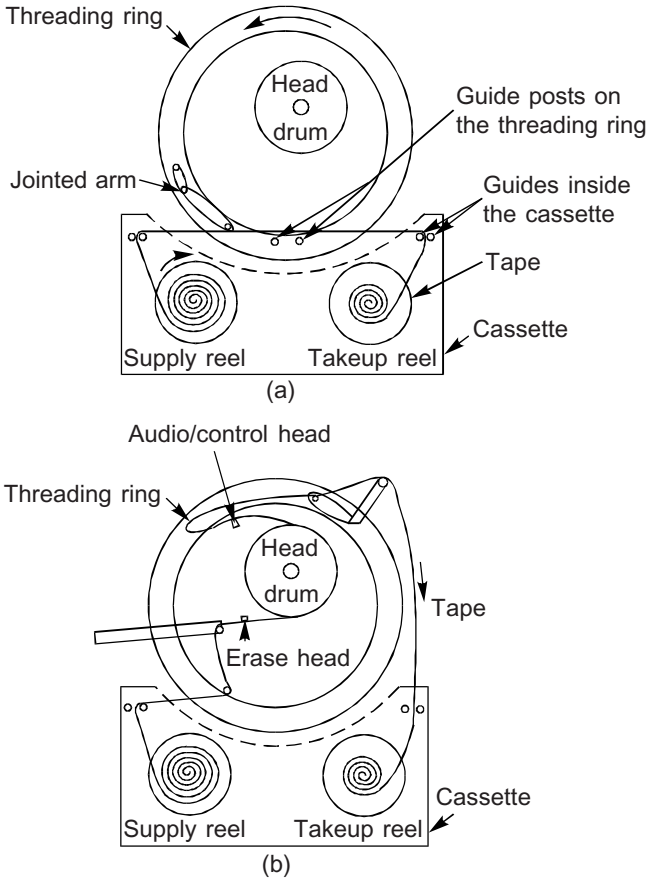


Fig. 22.37 Betamax Threading System: (a) Before Start of Threading Mode; (b) Final Position of Threading Ring

position, with the tape wrapped around the head-drum as shown in Fig. 22.37(b). The entire threading operation in Betamax takes less than 3 seconds. Once threaded, the tape remains in the threaded position for all the operating modes such as the play, fast forward and rewind etc. The threading process can be reversed and unthreading started by pressing the eject button.

Video Tracking In the early VTR machines using two head scanners, each scanner laid down a slanting track on the tape during half rotation of the drum.

This track contains one TV field complete with $312\frac{1}{2}$ horizontal lines and the sync pulses. The other scanning head B laid down a similar track adjacent to the track laid down by scanner A, leaving small empty space between the two adjacent tracks as shown in Fig. 22.38. The empty space between two tracks is called a guard band. The guard band is meant to avoid any crosstalk between adjacent tracks picked up during playback due to misalignment of the scanning heads. The width of the guard band depends on the linear speed of

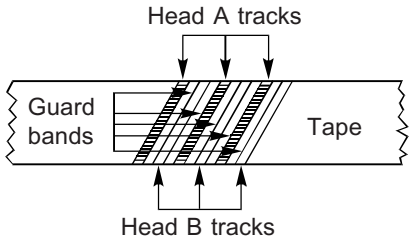


Fig. 22.38 Video Tracks with Guard Bands

the tape. Although preventing crosstalk, the guard bands considerably reduce the playing time due to low packing density of information. The playing time could, however, be increased by laying thinner tracks (less head gap) and bringing them closer by reducing tape speed. But the final solution of the problem was found in another technique called azimuth recording.

Azimuth Recording Technique Complete elimination of guard band without increasing the crosstalk between adjacent tracks is necessary for tape economy and increase of playback time. Elimination of guard bands can be achieved by slowing down the tape speed so that the adjacent video tracks but against each other or even overlap slightly as shown in Fig. 22.39(a). For eliminating the crosstalk, the azimuth angle (angle made by the head gap with the vertical) of each head gap is so adjusted that head A becomes quite insensitive to tracks laid down by head B and vice-versa.

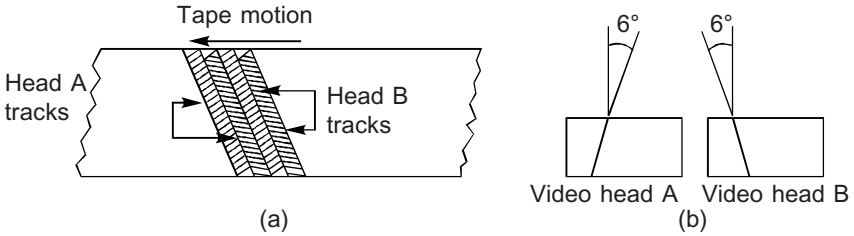


Fig. 22.39 (a) Azimuth Recording (b) Azimuth Angles for Heads A and B.

The phenomenon is similar to the one observed in audio cassette recorders when a particular cassette recorder plays back satisfactorily its own recordings but gives a low output, particularly at high frequencies, when playing recordings made on other cassette recorders. Adjustment of the azimuth angle of the audio head gap rectifies the defect.

In azimuth video recording, the head gaps of the two video heads are tilted on opposite sides of the vertical as shown in Fig. 22.39(b), to produce isolation between adjacent tracks. Although all cassette recorders use the azimuth system of recording, the angle of tilt for the head gaps is different in different systems or formats. In the VHS system, used in India; the video heads are tilted by 6° on opposite directions, thereby providing an azimuth error of 12° between the heads A and B. In Betamax system the tilt is 7° and Philip's in the

V2000 format uses an off-set of 15°. The playback time can be doubled or even trebled for the same size of the spool by the azimuth video technique.

Video head azimuth angles are normally fixed in the manufacturing process and are not adjustable.

Tape Formats A video format is a system of video recording which has characteristics different from other video recording systems. Tape width, tape speed, video head rotation, tape threading system, and placement of deck components is different in different systems or formats. The width and angle of video tracks, the width and number of audio and control tracks, the direction of head rotation and the tape movement together with the running time and picture quality are also clearly specified in a format. Besides the above specifications, certain special features like freeze-frame, picture search, frame-by-frame advance, Dolby noise reduction and remote control differ from system to system. Sometimes the same format a dopted by different countries uses different tape speeds to provide different playback times. Even with different methods adopted by different formats, the end results are more or less the same—the production of good quality pictures and sound. The multiplicity of formats in use today makes the inter-changeability of the cassette difficult because a programme recorded on one format cannot be played back on a machine having a different format. This incompatibility of formats makes the choice of a format difficult. Standard specifications for an $\frac{1}{2}$'' format have been laid down by the Electronic Industries Association of Japan (EIAJ) and till these standard EIAJ specifications are adopted, particularly in the case of domestic VCRs, the choice of a particular format or building a video library on any particular format seems to be risky for fear of that format getting out of date.

The three $\frac{1}{2}$'' formats that are most popular these days are the VHS, BETAMAX and the V2000 format.

VHS VHS (Video Home System) is the most popular system that is used in India and all over Europe. This format is also used in USA alongside the Betamax and was introduced in 1976 by JVC (Japanese Victor Company). It uses the helical scanning system on a tape width of 1.27 cm. ($\frac{1}{2}$''), with the two video heads rotating at 1500 rpm. Only one tape speed of 2.339 cm/s or 0.92 ips is used in India and Europe, providing a playback duration of 3 hours. In USA. Japan, however, three speeds—SP (Standard Play), LP (Long Play) and SLP (Super Long Play) are used giving 2, 4 and 6 hours of playback time respectively. One stationary audio/control head and a full track erase head are used. Other facilities provided are multi-day, multi-channel programmes, remote control facilities, freeze-frame, fast and slow motion with rapid search up to 10 times the normal speed. The additional facilities provided differ from model to model and the cost of equipment goes up as additional facilities are provided.

Betamax This particular format was introduced by Sony and is very popular in USA/Japan. The original Beta format could play for only one hour, but the

Beta-2 format offers up to $3\frac{1}{4}$ hours of play in Europe using a tape speed of 1.873 cm/s or 0.74 ips. In USA/Japan, the three speeds used are X_1, X_2 and X_3. X_1 is virtually extinct but X_2 and X_3 give 3 hours 20 minutes and 5 hours of play respectively. Like VHS, the tape width used is 1/2 inch and the linear tape speed of 1.873 cm/s or 0.74 ips is slower than the VHS tape speed, but due to the larger drum diameter, the actual writing speed is higher than the VHS writing speed. The tape path called B-load is much more complicated than the M-W rap used in VHS. Other facilities available are similar to those provided by VHS, but the audio quality needs special noise reduction system due to lower linear speed of the tape.

V2000 This is the latest and most sophisticated of the three popular VCR formats. This format has been evolved as a result of a joint venture of two leading European manufactures, Grundig and Philips. As a result of the use of new technical advances in the field of electronics, IC fabrication, tape performance and precision machining, this format is able to achieve a very high density packing of information. Consequently, a single 1/2inch tape in this format is able to provide 8 hours of recording and playback, compared to only 4 hours available from the VHS format. This is, however achieved by using only half (1/4 inch) of the tape on each pass and then flipping the cassette to use the other half of the tape. The video tracks are made much thinner and to ensure accurate scanning of these narrow tracks, the machines are equipped with the Dynamic Track Following (DTF) system. In the DTF system, use is made of piezoelectric crystal plates which automatically move the head in position for proper tracking in response to a sensing device installed in the heads themselves.

The tape path is more or less similar to the VHS threading system. The tape speed is 2.4 cm/s and with a 6.5 cm. head-drum, it provides a writing speed of 5.08 m/s which lies in between the writing speeds of the other two formats. This format is steadily picking up in Europe but has not yet been launched in the USA or India.

Many portable systems like the 8 mm Video, based on the popular domestic type formats are also gaining popularity.

22.29 CARE AND MAINTENANCE OF VCRs

Care and maintenance of VCR machines includes care in installation and operation of the machines, cleaning of video heads, tape decks and maintenance and care of the tape transport system. Occasional tests on the performance of the machines are necessary to keep the performance within prescribed limits. These tests are carried out by qualified and experienced technicians with the help of special test equipments. Cleaning of tape deck and video heads, maintenance and care of the tape transport system and other important items for the repair and maintainence of VCR machines will be described in Chapter 24. (For further details see VCR, Principles, Maintenance and Repair by the author).

SUMMARY

Sound recording is the process of storage of sound. In sound recording the acoustical signals are converted into some form of permanent or semi permanent record from which these can be reconverted into sound. Based on the technique employed, the three prevalent methods of sound recording are, disc recording, film recording and magnetic tape recording of these systems, tape recording is the most popular and universally employed method of sound recording.

In tape recording, a plastic tape coated with powdered magnetic material is magnetised to varying degrees of magnetisation along its length when moved in front of a recording head carrying AF currents from an audio amplifier, the magnetic variations recorded on the tape can be reconverted into electrical currents when the varying magnetic flux emitted by the magnetised tape is linked with an electrical circuit called the playback head. This completes the recording and playback process. One of the great advantages of tape recording is that the recorded matter can be wiped off by a process called erasing. A tape recorder thus performs the three main functions of recording, playback and erasing.

Disc recording is a process in which sound is recorded by modulating a groove cut into a soft lacquer surface formed on a hard aluminium or glass base. In film recording the sound is recorded photographically by modulating a light beam converging on a moving photographic film.

Hi-fi pertains to the basic quality of sound reproduction whereas stereo is a system of two channel sound recording or reproduction which lends depth and direction to reproduced sound. Stereophonic recording can be made on a stereophonic tape recorder using four track tapes. Stereophonic disc recordings are also available which can be played back only with a special stereophonic cartridge.

The main process involved in the production of stereophonic sound is the time of arrival difference of the sound at the two ears which is estimated by the brain to produce the stereophonic effect.

Basically there are only two methods for the production of stereo signals. One is the use of a pair of microphones and the other is the PANPOT method using only one microphone and electrically splitting its output into electrically unequal parts.

Compatibility in stereo broadcasting is achieved by using two bidirectional microphones arranged at 45° to the front axis to produce the M (Mono) and S (Stereo) signals. M/S microphones are also available to produce M and S signals. The M and S signals are combined to form a composite audio signal which modulates a 38 kHz carrier for the broadcast of stereo signals by radio.

Digital Sound Recording Digital Sound Recording system is superior to analog recording system in terms of dynamic range, fidelity robustness, freedom from degradation in recording and multicopying. The process involves the

conversion of analog to digital form by sampling quantizing and their converting into a binary code representing 1's and 0's. This is recorded on the tape in the form of pulses which if degraded can be reconstructed. Bit rate is the product of sampling frequency and the number of bits per second. Bit rate gives an indication of the quality of the system. Error checking and correction is possible with Parity checking method. Main problems with digital recording and playback are the enormous bit rate that is to be recorded on the tape. Specially designed machines are used for the recording and playback of the digital system. The important machines are R-DAT (Rotary Head-Digital Audio-Tape and Dash (Digital-Audio Stationery-Head).

Compact Discs (CDs) are the best example of digital recording. The recording as well as playback are done with a laser beam. A 12 cm CD can provide a max playback time of 75 minutes compared to only 30 minutes playback time provided by a conventional single disc.

Video Recording Systems The main problem in video tape recording is the large bandwidth (4 to 5 MHz) of the TV signals to be recorded. This requires the use of very high tape speeds and very narrow head gaps. The limitations of video recording completely vanished by increasing the writing speed of the tape recording system by rotating the recording head at a high speed across the width of the tape which moves with its normal slow longitudinal speed.

Two other problems of video recording were the high dynamic range (25 Hz to 5 MHz) compared to the dynamic range 120 Hz to 25 kHz of audio recording. This problem is solved by frequency modulating the video signal on a high frequency carrier before applying it to the recording head.

The use of two recording heads mounted at 180° to each other and the use of helical scanning enable one field to be recorded by one pass of each head.

Various improvements made in the video recording system resulted in the design of a VCR (Video Cassette Recorder) like the audio cassette recorder. Three popular formats of VCR are the Betamax system, the VHS and the V2000. Each format has its own characteristics and these are not interchargeable. The most popular format is VHS and this is being used in India.

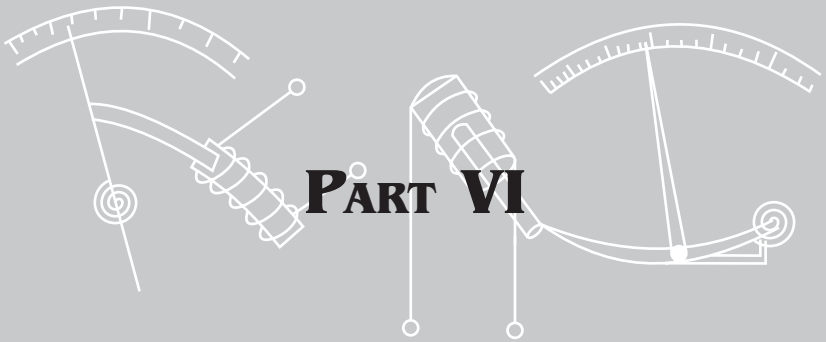
Care and maintenance of VCRs needs special test equipment and this will be described in Chapter 24.

Review Questions

1. What is sound recording? Name the three prevalent systems of sound recording.
2. What is tape recording and what are its advantages over other systems of sound recording?
3. Give the block diagram of a tape recording system and explain its working.
4. What is biasing in tape recording? Compare the two biasing methods normally employed in tape recording.

5. Define erasing and explain the different methods of erasing used in tape recorders.
6. Describe the tape transport system in a cassette tape recorder.
7. What is a two-in-one? Explain its working.
8. Explain the difference between hi-fi and stereo. Why is a stereo system not necessarily hi-fi?
9. Explain why tape recording is most suited for stereo recording and playback?
10. Describe the method of stereo disc recording? How does the stereo cartridge reproduce the stereo disc recordings.?
11. Explain the two methods used in the production of stereo signals. Give their relative advantages.
12. Discuss compatibility in stereophonic sound system. Describe how compatibility is achieved in the transmission of stereo broadcast signals.
13. (a) Give the advantages of digital sound recording over the analog sound recording systems.
    (b) Explain the important steps involved in the process of digital recording and playback.
14. What are Compact Discs (CDs)? How are CDs superior to normal audio discs.
15. What are main problems of video recording? How are these problems solved in actual practice?
16. (a) Give the block diagram of VCR tape deck and briefly explain the function of various components mounted on the tape deck.
    (b) Describe briefly the characteristics of the three important VCR formats. Which format is used in India?
17. State whether true or false with brief justification.
    (a) AC biasing is better than DC biasing.
    (b) Digital recording is better than analog recording because it needs less bandwidth.
    (c) Stereophonic sound is superior to hi-fi sound.
    (d) A compact disc is a good example of stereophonic recording.
    (e) Modulation in video recording reduces the number of ocatves.

    [Answer (a) True, (b) False, (c) False, (d) False, (e) True]



PART VI

ELECTRONIC MEASURING EQUIPMENTS, MAINTENANCE AND TROUBLESHOOTING

chapter 23

Electronic Test and Measuring Equipment

23.1 INTRODUCTION

Electronic test and measuring equipments are essential for the repair and maintenance of any modern entertainment equipment, Radio, TV and VCR. These test equipments include simple meters for the measurement of current, voltage and resistance together with more sophisticated equipments like oscilloscopes, RF signal generator and pattern generators required for conducting general performance tests on the electronic equipment under test. The working principles and the methods for the use of these measuring equipments will now be described.

23.2 ELECTRICAL MEASURING INSTRUMENTS (SIMPLE METERS)

The most important quantities that one comes across in the study of electricity and electronics are current, voltage, resistance and power. For successful operation, installation and repair of electrical and electronic equipment like radio and TV receivers, it becomes necessary to measure one or more of these electrical quantities. The instrument used for the measurement of these quantities is called a *meter*. Since all the four quantities are interrelated by Ohm's law, the meter used for the measurement of one quantity, generally current, can be suitably calibrated to read other quantities as well. Meters used for the measurement of current, voltage, resistance and power are known as ammeters, voltmeters, ohmmeters and wattmeters respectively. Basically all these meters are current meters with suitable modifications to measure the other three quantities. The correct use of these meters requires a clear understanding of the working of these meters. The principles underlying the working and use of the common types of meters are described in this chapter.

23.3 CURRENT METER

Any of the effects produced by current can be utilised for the measurement of current. Two of the effects that are utilised for the measurement of current are the magnetic effect and the heating effect. The strength of the magnetic field and the heat produced by the flow of current are both proportional to the amount of current flow and so can be used to measure current. Each of these effects will give rise to a different type of current meter. However, the most commonly used type of meter is the one based on the magnetic effects of the current and this will be described first.

23.3.1 Moving Coil Meter or Galvanometer

The principle of a Galvanometer is the same as that of a DC motor. A coil carrying current develops a rotary motion when placed between the poles of a permanent magnet as shown in Fig. 23.1. The amount of motion and its direction depend on the strength of current and its direction in the coil. The direction of motion can be reversed by reversing the direction of current. If the coil is suspended by a spring which opposes the motion of the coil, it will come to rest in a position of equilibrium. The strength of the current flowing through the coil can be read on a calibrated scale with the help of a pointer that moves with the coil. This is the basic principle of a moving coil current meter also known as a *galvanometer* and named after the Italian scientist Galvani. A simple device consisting of a moving coil and a stationary permanent magnet was originally invented by a Frenchman, Arsene d' Arsonval, and is called d' ARSONVAL meter movement. Figure 23.2 indicates the details of a simple moving coil or d' Arsonval meter movement. The coil is wound on a soft iron ball called the armature. The coil and the armature are delicately pivoted on a jewelled bearing. A small spiral spring (not shown in the figure) opposes the movement of the coil. A light aluminium pointer attached to the coil moves on a calibrated scale. Wires are brought out from the two ends of the coil on the binding posts and the entire meter movement is enclosed in a glass-faced case which prevents entry of dust and damage to the instrument.

Moving coil meters are used only for DC measurement. This type of meter will not read AC currents.

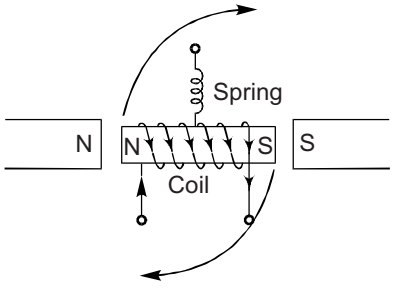


Fig. 23.1 Principle of a Galvanometer

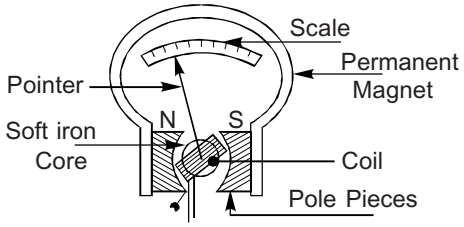


Fig. 23.2 Moving Coil Meter

23.3.2 Moving Iron Type Meter

A moving iron type of current meter is based on the principle of repulsion between similar magnetic poles. If two soft iron bars *AB* and *CD* are placed inside a current carrying coil wound on a non-magnetic former, the two soft iron bars will get magnetised in the same direction with the two ends facing each other having the same polarity as shown in Fig. 23.3(a). The two similar poles will repel each other. If the direction of the current through the coil is reversed as in Fig. 23.3(b), the polarities of the two iron bars will be reversed but similar poles will again be facing each other and these will again repel. If one of the two soft iron pieces is fixed and the other is free to move, this motion can be used to measure current by attaching to the moveable iron bar a pointer that will move over a calibrated scale.

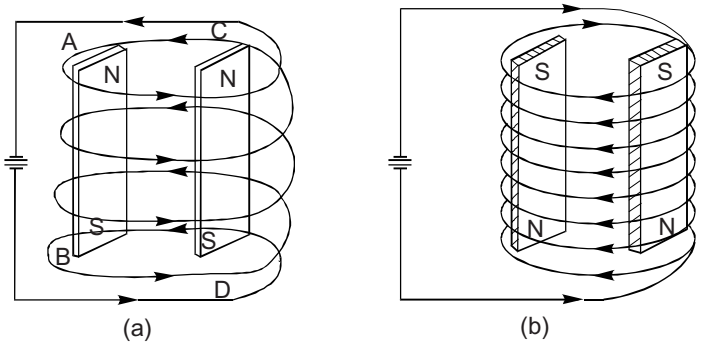


Fig. 23.3 Principle of a Moving Iron Meter

In the actual construction of the moving iron meter, the soft iron bars take the form of two rectangular soft iron pieces called *vanes*. One vane is fixed and the other one is free to rotate round a pivoted edge. The movement of the rotating vane is controlled by springs attached to the axle. A pointer attached to the rotating vane indicates the value of the current on a graduated scale as shown in Fig. 23.4.

In the arrangement shown in Fig. 23.4, the two vanes placed inside the coil point along the radius of the coil and this is called the radial vane movement.

In the concentric vane type of movement (not shown) the two vanes are semi-circular in shape and form segments of two circles concentric with the coil. The principle of operation in both the types is the same and both are known as repulsion type of meter movements.

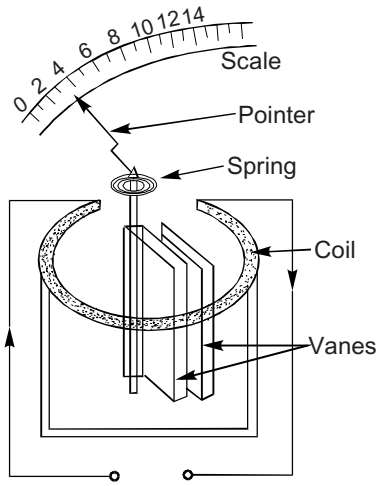


Fig. 23.4 Radial Vane Meter Movement

A third type of moving iron meter movement, known as the Plunger type, is based on the principle of magnetic attraction between opposite poles. In this type a soft iron plunger is pulled into a fixed coil when current flows through the coil and the plunger gets magnetised due to the magnetic field in the coil produced by the current. The distance that the plunger moves into the coil depends on the amount of current flowing through the coil. The plunger which is pivoted has a pointer attached to it and the movement of the pointer on a calibrated scale indicates the amount of current. Figure 23.5 shows the details of this type of meter movement.

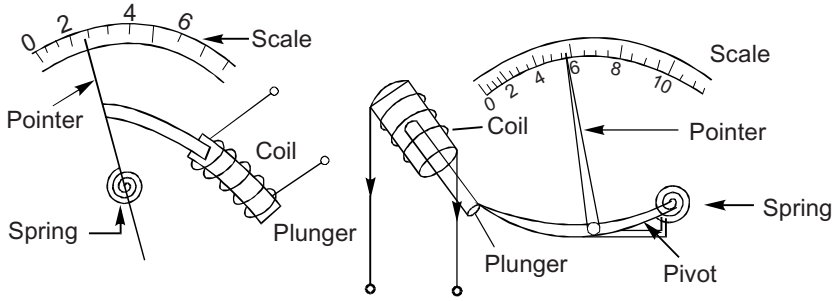


Fig. 23.5 Plunger Type Moving Iron Meter

Although the moving coil and the moving iron types of meters described so far are both based on the electromagnetic effects of current, the two meters use different types of scales. The scale used in the moving coil meter is *linear* whereas the moving iron type of meter makes use of a *non-linear* scale.

A *linear scale* is one in which divisions or markings are equidistant and the pointer moves over equal distances for equal changes of current. A linear scale is shown in Fig. 23.6(a). A linear scale is used by the moving coil type of meters because the deflection of the pointer is directly proportional to the current flowing through the coil.

A *non-linear* scale is one in which the divisions are not equidistant and the pointer does not move over equal distances for equal changes of current. A non-linear scale is shown in Fig. 23.6(b). Here the lower end of the scale is very congested and crowded, and the divisions are spread out and farther apart towards the upper end of the scale.

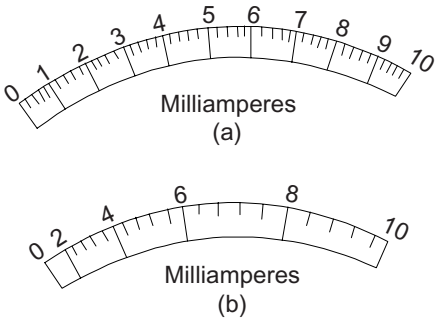


Fig. 23.6 (a) Linear scale (b) Non-linear Scale

In the case of moving iron type of current meters, the force of repulsion or attraction is proportional to the *square* of the current passing through the coil. When the current through the coil is doubled, the strength of the magnetic

poles also doubles and the force of attraction or repulsion becomes four times. Similarly, if the current becomes three times, the force of attraction or repulsion becomes nine times and so on. Thus the deflection of the pointer is proportional to the square of the current and the pointer will not move over the same distance on the scale for equal changes of current. The scale used is non-linear. Moving-iron meters are not very sensitive and are used in power circuits. These can be used both for AC and DC measurements.

23.3.3 Meter Movements Based on Heating Effects of Current

1. Hot Wire Type Meter Movement

This is based on the principle of expansion of a metal wire due to the heat produced by a current flowing through the wire. Current passes through a thin wire *AB*, tightly stretched between two points as in Fig. 23.7. From the centre of the thin wire is attached a tension wire which exerts a constant pull on the thin wire. A silk thread attached to the linear wire keeps the tension wire pulled sideways by a spring at the end. The silk thread passes over a roller so that any slight movement of the silk thread deflects the pointer attached to the thread. When the thin wire *AB* expands due to the heat produced by a current flowing through it, the thread gets pulled and the pointer moves over the scale. The expansion of the wire depends on the amount of heat produced which is proportional to the square of the current passing through the wire. The scale used by this meter will, therefore, be a non-linear scale. This meter movement is suitable both for DC and AC measurements.

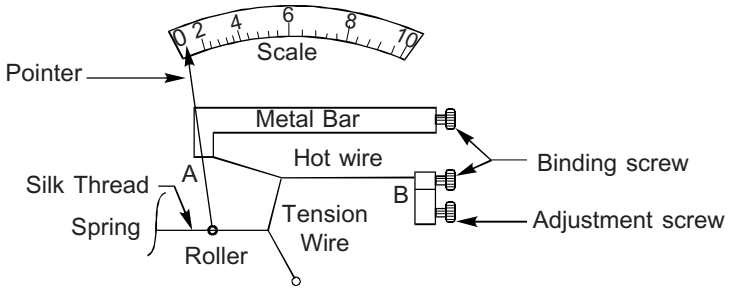


Fig. 23.7 Hot Wire Meter Movement

2. Thermocouple Type Meter Movement

As already shown a thermocouple is the junction of two dissimilar metals. When the thermojunction is heated, a DC voltage is developed between the open ends of the thermocouple. The DC voltage is proportional to the difference of temperature between the thermojunction and the open ends of the thermocouple. This DC voltage can be measured with a moving coil meter as shown in Fig. 23.8. We get the maximum voltage per degree rise of temperature when a thermojunction of bismuth and antimony is used.

When a current flows through the heating element, the voltage developed across the open ends of the thermocouple sends a current through the moving coil meter which shows a deflection proportional to the current flowing through the heating element. Since the heat developed is proportional to the square of the current (I^2R), the thermocouple meter uses a non-linear or square law scale. This meter is most suitable for measuring high frequency currents at radio frequencies although it is capable of measuring both DC and AC currents.

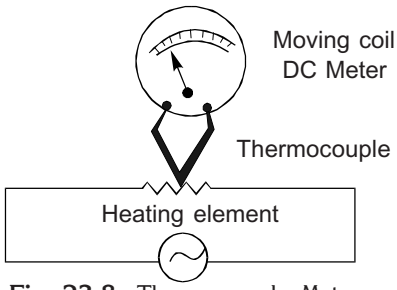


Fig. 23.8 Thermocouple Meter Movement

23.3.4 DC and AC Meters

Because of the basic difference between the nature of DC and AC voltages and currents, the meters used for the measurement of DC and AC quantities also operate on different principles. There are meters that will read only DC while others will read only AC. Of course, there are meters that will read both DC and AC. Of the meter movements described so far, the moving coil movement is meant to read DC currents only. This is a very sensitive and accurate type of meter movement and most meters meant for radio and TV work use this movement only. Even the AC meters make use of the moving coil movement by converting AC into DC by a rectifier and then feeding this DC into the moving coil movement. Such AC meters are called rectifier type meters.

The moving iron type of meter movements can measure both DC and AC but their use is confined mostly to low frequency AC circuits like the power supply lines. The hot wire meter movement and the thermocouple meter movement are also suitable for DC as well as AC applications, but the thermocouple meter movement is exclusively used for the measurement of high frequency (HF) currents whose frequencies extend from a few MHz to thousands of MHz. Such HF currents are obtainable in the antenna and feeder line circuits of radio and TV transmitters and receivers.

The methods for the measurement of current, voltage and resistance described below are based on the use of DC meters only because the AC measurements of corresponding quantities are only slight modifications of DC methods.

23.4 MEASUREMENT OF CURRENT

23.4.1 Ammeters

The instrument used for the measurement of current is called an *ammeter* or a *milliammeter* depending on whether the scale is calibrated in *amperes* or *milliamperes*. The current meter must be connected in series in the circuit in which

the current is to be measured. The current meter should have a very low (negligible) resistance compared to the total resistance in the circuit where the current is to be measured, so that the insertion of the meter in the circuit does not alter the value of the current flowing in the circuit. For connecting the meter in series, the circuit has to be broken at some point and the meter connected as shown in Fig. 23.9.

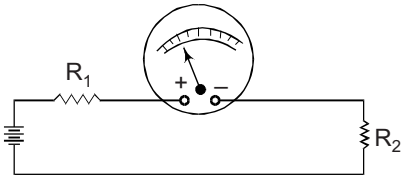


Fig. 23.9 Method of Connecting an Ammeter in Circuit

A DC current meter with a moving coil movement has its terminals marked for + and – polarities or painted with red for plus and black for minus. For the meter to read correctly, it should be so connected in the circuit that the high voltage point (+) is connected to the positive (red) terminal and low voltage point (–) is connected to the negative (black) terminal of the meter. Failure to do this might result in damage to the meter movement.

23.4.2 Meter Sensitivity

The DC resistance of a meter is the resistance of the coil and depends on the number of turns of the coil and the size of the wire used. This is called the internal resistance of the meter. The maximum deflection that the pointer can indicate on the scale is called *full scale deflection* (FD). The amount of current required to produce full scale deflection in the meter is called the *sensitivity* of the meter. If a current of 1 mA is required to produce full scale deflection, the meter has a sensitivity of 1 mA whereas a meter requiring only 100 μA for full scale deflection has a higher sensitivity of 100 μA. The lower the current required for full scale deflection, the higher is the sensitivity of the meter. The common types of meter movements have a sensitivity varying from 1 mA to 10 mA whereas in the case of more sensitive types, the sensitivity varies from 50 μA to 100 μA.

The maximum current that can be measured with the meter movement alone is the current required for full scale deflection which may be only a few milliamperes. Any current higher than the FD current will simply damage the meter. For extending the range of a meter, a low resistance called *shunt* is connected in parallel with the internal resistance of the coil. The value of the shunt is so chosen that only FD current passes through the meter movement and the rest of the current is bypassed by the shunt.

23.4.3 Calculating Shunt Resistance

Let I_m be the full scale deflection current for the meter movement whose internal resistance is R_m and I_s the amount of current that is to be bypassed by the shunt whose resistance is R_s (Fig. 23.10). Since R_m and R_s are in parallel, the voltage drop (IR) across these two resistances is the same and equal to, say, E volts.

Then $E = I_m \cdot R_m = I_s \cdot R_s$

or $$R_s = R_m \cdot \frac{I_m}{I_s} \qquad (23.1)$$

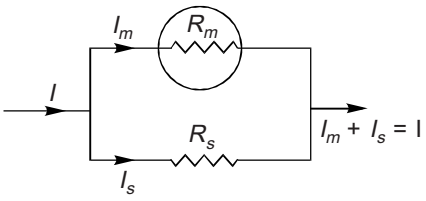


Fig. 23.10 Calculating Shunt Resistance

Suppose a meter movement with an FD current of 1 mA and internal resistance of 25 Ω is required to measure a current of 10 mA, then in Eq. 23.1,

$R_m = 25\ \Omega$, $I_m = 1$ mA and $I_s = 10$ mA – 1 mA = 9 mA

Thus, $$R_s = 25 \times \frac{1\,\text{mA}}{9\,\text{mA}} = 2.77\ \Omega$$

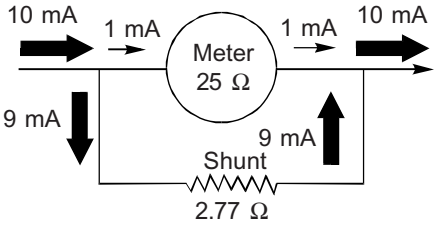


Fig. 23.11 Extending the Range of the Meter

So a shunt resistor of 2.77 Ω connected in parallel with the meter movement as in Fig. 23.11 will divert 9 mA through the shunt with only 1 mA flowing through the meter to show full scale deflection. With the shunt connected, each division of the scale will have to be multiplied by 10 to get the actual value of the current. If the pointer shows 0.5 mA on the scale, the actual value of the current is 0.5 × 10 = 5 mA.

If the range is to be extended to 100 mA, then 99 mA will pass through the shunt and 1 mA through the meter. As above, the shunt resistance is given by

$$R_s = 25 \times \frac{1}{99} = 0.25\ \Omega$$

An ammeter with a number of ranges is called a *multirange* ammeter. In multirange ammeters, the shunt resistances can be enclosed inside the case of the meter and brought into the circuit by a switching arrangement as shown in Fig. 23.12. These are called *internal-shunt* ammeters or milliammeter. Others that have the shunts connected from outside are called *external shunt* ammeters.

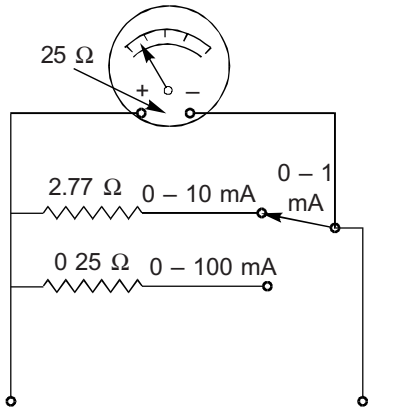


Fig. 23.12 A Multirange Milliammeter

Example A 50 μA meter movement has a resistance of 100 Ω. Calculate the shunt resistance for extending the range to 50 mA.

$$R_m = 100\ \Omega,\ I_m = 50\ \mu\text{A} = 0.05\ \text{mA}$$
$$I_s = 50 - 0.05 = 49.95\ \text{mA}$$
$$R_s = 100 \times \frac{0.05}{49.95} = 0.1\ \Omega \text{ approx.}$$

23.4.4 Universal Shunts or Ring Shunts

The universal or ring shunt method, also known as the Ayrton method of providing shunts for multirange current meters, makes use of a number of resistances in series and parallel arrangement to provide different current ranges in the meter. As shown in Fig. 23.13 a universal shunt $R_1 + R_2 + R_3$ of value equal to the shunt resistance for the minimum range is connected in parallel with the meter movement and suitable tappings are provided on the universal shunt for other ranges required for the meter. The method for calculating the values of R_1, R_2 and R_3 is explained in Fig. 23.13.

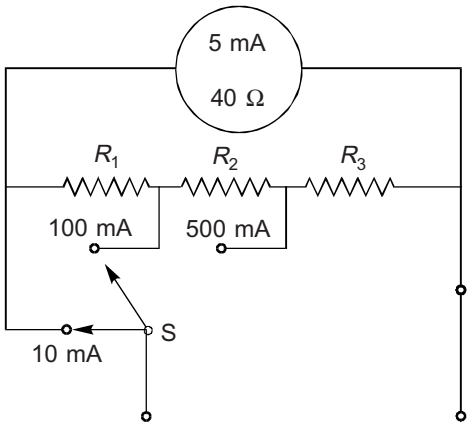


Fig. 23.13 Universal Shunt for Multirange Current Meters

The meter movement with 5 mA full scale deflection current and 40 Ω internal resistance is to be provided with a universal shunt suitable for current ranges of 10 mA, 100 mA and 500 mA. For the minimum range of 10 mA, 5 mA will flow through the meter and 5 mA will flow through the universal shunt consisting of $R_1 + R_2 + R_3$

Therefore $$R_1 + R_2 + R_3 = 40\ \Omega$$

To calculate the value of R_1 for the 100 mA range we can see that when the selector switch is in position 100 mA, the total current of 100 mA will have two parallel paths. The value of R_1 should be such that a current of 5 mA will flow through R_1 in series with the meter resistance of 40 Ω and the rest (100 – 5) mA will flow through $R_2 + R_3$ in series. The *IR* drop across the two parallel paths is the same. Therefore,

$$(R_1 + 40) \times 5\ \text{mA} = (R_2 + R_3) \times 95\ \text{mA}$$

since $$R_2 + R_3 = R_1 + R_2 + R_3 - R_1 = 40 - R_1$$

$\therefore$ $$(R_1 + 40) \times 5\ \text{mA} = (40 - R_1) \times 95\ \text{mA}$$
$$5R_1 + 200 = 3800 - 95R_1$$
$$100R_1 = 3600$$

or $$R_1 = 36\ \Omega$$

Therefore, $$R_2 + R_3 = 40 - 36 = 4\ \Omega$$

For the 500 mA range, 5 mA will flow through $R_1 + R_2$ in series with the meter resistance and the rest (500 – 5) mA or 495 mA will flow through R_3

$$\therefore \quad 5(R_1 + R_2 + 40) = R_3 \times 495$$

We know $R_1 = 36\ \Omega$ and $R_3 = 4 - R_2$

$$5(36 + R_2 + 40) = (4 - R_2)\ 495$$

Solving for R_2 we get

$$R_2 = 3.2\ \Omega$$

and R_3 will be $(4 - 3.2) = 0.8\ \Omega$

The universal shunt method has two main advantages over the parallel shunt method. Firstly, in the parallel shunt method, the contact resistance of the selector switch becomes comparable with the shunt resistance and this contact resistance coming in series with the shunt resistance can cause considerable error in reading. In the universal shunt method, the contact resistance is external to the shunt circuit and does not affect accuracy. Secondly, in the parallel shunt method there is the danger of full current passing through the meter and damaging the meter movement when the range switch is moved from one position to the other. No such danger exists in the universal shunt method because the meter is disconnected from the circuit while changing from one current range to another.

23.5 MEASUREMENT OF VOLTAGE

23.5.1 Voltmeters

The meter used for measuring the electromotive force or the potential difference between any two points in a circuit is called a *voltmeter*. A voltmeter is basically a current meter connected across the two points between which the potential difference is to be measured. A current flows through the meter which is proportional to the voltage across the meter and this can be measured on a scale calibrated to measure volts. Since the current meter has a low resistance and is connected in parallel with the circuit for voltage measurements, it is essential that the current flowing through the meter should not exceed the full scale deflection current to avoid damage to the meter movement. This is achieved by connecting a high resistance in series with the meter movement. This combination of a meter movement and a high series resistance forms the voltmeter. The series resistance is called a *multiplier* and its value depends on the voltage range and the internal resistance of the meter. The higher the value of the voltmeter resistance, compared to the circuit resistance, the less shunting effect it produces on the circuit and more accurate are the voltage measurements. A voltmeter is never put in series with a circuit because of its high resistance. The terminals of a DC voltmeter are also marked with + and – polarities and correct polarity must be observed while making voltage measurements with a DC voltmeter.

Calculating Multiplier Resistance

A basic meter movement can read a voltage equal to the voltage drop (IR) across its coil produced by the FD current. Thus, a meter movement with 1 mA

FD and 1000 Ω coil resistance can read a maximum voltage of $1000 \times 1 \times 10^{-3} = 1$ V. If voltages higher than 1 V are to be measured, the current through the coil must be limited to 1 mA by a series multiplier resistance of a suitable value. Suppose the voltage range is to be extended to 10 V. If R_{mult} is the resistance of the multiplier and R_m the resistance of the meter movement, then by Ohm's law,

$$R_{mult} + R_m = \frac{10}{0.001}$$

$$= 10{,}000\ \Omega \text{ since } 1 \text{ mA} = 0.001 \text{ A}$$

$$R_{mult} + R_m = 10{,}000$$

or $$R_{mult} = 10{,}000 - R_m = 10{,}000 - 1000 = 9000\ \Omega$$

When a resistance of 9000 Ω is connected in series with the meter movement as in Fig. 23.14, a voltage of 10 V applied across the combination will produce full-scale deflection. Either a new range can be added to the meter dial or the original 0 – 1 V range can be used by multiplying each reading by a factor of ten to get the readings for the new range.

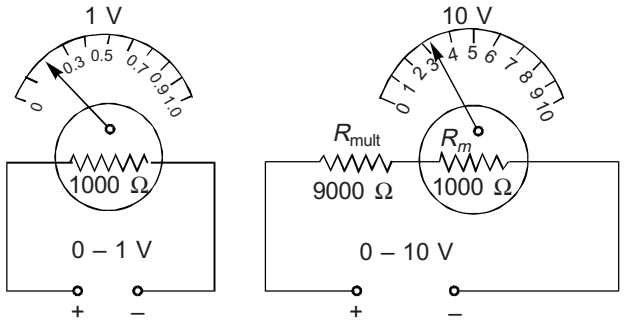


Fig. 23.14 Multipliers for Extending Voltmeter Range

In the same way it can be shown that for extending the range of the above voltmeter to 100 V the multiplier resistance required will be 99,000 Ω. In a multirange voltmeter, the multiplier resistances are connected inside the meter case and the required multiplier can be brought into circuit by a range switch as shown in Fig. 23.15. In another arrangement of multiplier resistances known as the series-multiplier arrangement, all the multiplier resistances are connected in series and their values are so chosen that the value of the multiplier resistance required for any particular position of the range switch is made up of the sum of all resistances connected in series before this range. Such an arrangement is shown in Fig. 23.16.

In this arrangement no external resistance is required for 0 – 1 V range. For 0 – 10 V range, the external series multiplier is 9000 Ω as in the previous case. For 0 – 100 V range, the series multiplier resistor is 9000 Ω + 90,000 Ω = 99,000 Ω and for the 0 – 1000 range, the multiplier resistance becomes 99000 + 900,000 = 999,000 Ω as before.

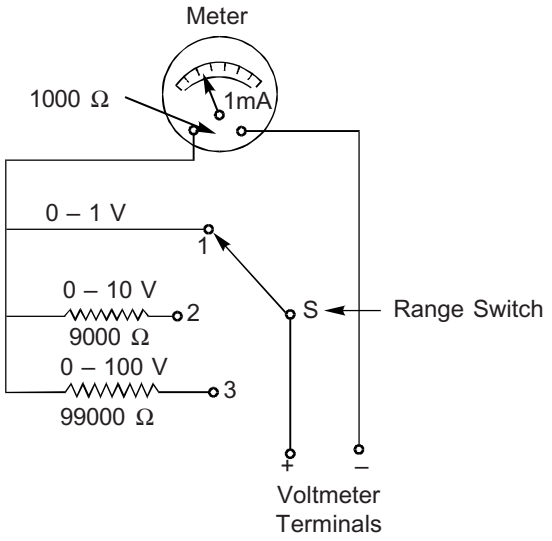


Fig. 23.15 Multirange Voltmeter

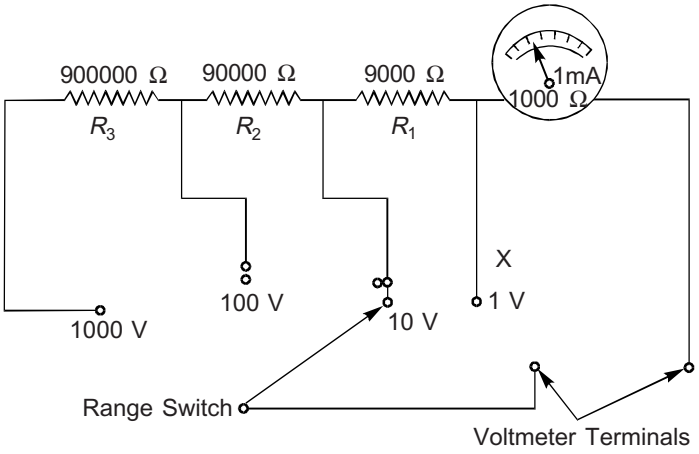


Fig. 23.16 Series Multiplier Arrangement

23.5.2 Ohms-Per-Volt Rating of Voltmeters

The Ohms-per-volt rating of a voltmeter is defined as the resistance for full scale deflection when 1 V is applied across the meter. Thus, a 1 mA meter movement will require 1 V/0.001 = 1000 Ω resistance per volt for full scale deflection. So its ohms/volt rating for the 0 – 1 V range is 1000 Ω/V. The ohms/volt rating is a very important characteristic of a voltmeter. It is an inherent characteristic of a voltmeter and is the same for all ranges in a multi-range voltmeter. Thus, in the case of the 1 mA, 1000 Ω meter movement mentioned above, for the 0 – 10 V range the total series resistance required is 10,000 Ω and in this range the ohms-per-volt rating is 10,000/10 V = 1000 Ω/V, which is the same as for 0 – 1 V range.

A 50 μA meter movement will require 1 V/50 $\times$ 10^{-6} = 20,000 Ω for full scale deflection and hence it has got a higher ohms-per-volt rating. The ohms-per-volt rating is also known as the *sensitivity* of a voltmeter. The higher the ohms/volt rating, the more accurate will be the readings obtained with the voltmeter because it produces less loading effect on the circuit as explained in Fig. 23.17.

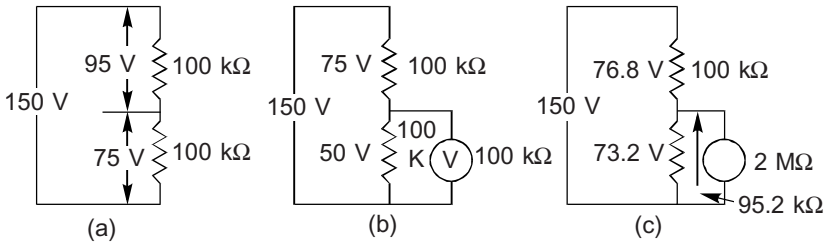


Fig. 23.17 A Higher Ohms/Volt Rating Gives More Accurate Results

In the circuit shown in Fig. 23.17(a) two series resistances of 100 kΩ each are connected across 150 V. Each resistance will have a voltage drop of 75 V across it, If a voltmeter with a 1000 Ω/V rating is used to measure the voltage across one of the resistors using the 0 – 100 V range, the total resistance of the voltmeter in this range is 100/0.001 = 100 kΩ and the two 100 kΩ resistances in parallel will be equivalent to 50 kΩ resistance. The voltage distribution across the network will change and the combination of voltmeter and the 100 kΩ resistance will now have a voltage drop of 50 V across it. The voltmeter will, therefore, read only 50 V against the actual value of 75 V which means an error of 33%. Now if a voltmeter with a 20,000 Ω/V rating (50 μA movement) is used to measure the same voltage, its total series resistance in the 0 – 100 V range be 2,000 kΩ which in parallel with the circuit resistance of 100 kΩ will be equivalent to a resistance of 95.2 kΩ which is not much different from the original value of 100 kΩ. The voltage across this combination will be 73.2 V and the error will be only 2.4%. Thus a voltmeter with a higher ohms/volt rating gives more accurate results than a voltmeter with a lower ohms/volt rating.

23.6 MEASUREMENT OF RESISTANCE

23.6.1 Ohmmeters

A meter used for the measurement of resistance is called an *ohmmeter*. Any meter movement in conjunction with a battery can be used to measure resistance. Consider the circuit in Fig. 23.18 in which a 4.5 V battery sends current through a 1 mA meter connected in series with one fixed resistance R_1 and a variable resistance R_2. The unknown resistance R_x is connected across a pair of test probes. The total R required to send a full scale deflection current of 1 mA through the meter is given by

$$R = \frac{4.5\text{ V}}{1\text{ mA}} = \frac{4.5}{0.001\text{ A}} = 4500\ \Omega$$

Of this, the meter itself provides 1000 Ω and the rest of 3500 Ω are to be provided by R_1 and R_2. R_1 can be a fixed resistance of 3000 Ω and the remaining resistance of (500 – R_x) can be provided by a variable 1 K resistor R_2. When R_x is zero, i.e. the two probes are shorted, the resistance R_2 provided by the variable resistor will be 500 Ω. Variable resistor R_2 is called the zero ohms adjustment, R_2 also compensates for any ageing of the battery cells.

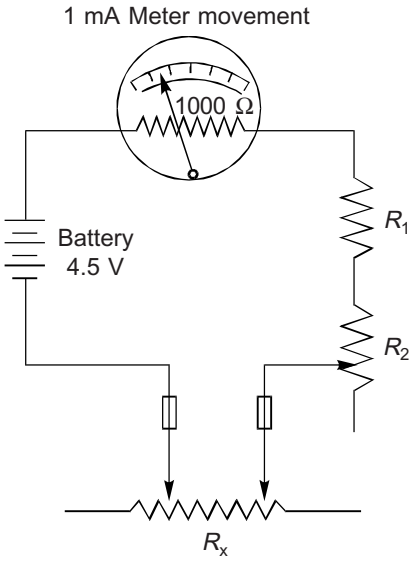


Fig. 23.18 Measurement of Resistance

When the two test probes are shorted, the unknown resistance R_x is zero and the meter shows full scale deflection. When the test probes are open, the unknown resistance R_x is infinity (∞) and the current value will be zero and the meter will show no deflection. Thus, full scale deflection means zero ohms and zero deflection means infinity ohms. It will thus be seen (Fig. 23.19) that the ohmmeter scale reads from right to left which is opposite in direction to the ammeter or voltmeter scales which read from left to right.

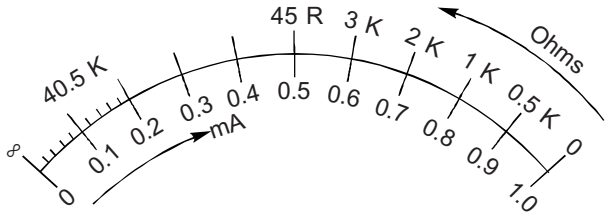


Fig 23.19 Ohmmeter Scale

The calibration of the ohmmeter scale between 0 Ω and ∞ Ω can be made as follows:

If R_x is 1000 Ω the circuit current will drop down from 1 mA to 4.5/(4500 + 1000) = 0.8 mA. 1 kΩ mark can be put against 0.8 mA current mark. If the value of R_x is made 2000 Ω, the current value will be 4.5/6500 = 0.7 mA approximately. The current will be 0.5 mA for R_x equal to 4.5 kΩ, and so on, till the other end of the scale is reached.

It will be seen that whereas the first 1000 Ω reduced the current value by 0.2 mA, the second 1000 Ω reduced it by only 0.1 mA. Equal amounts of resistance increase do not produce equal amounts of current decrease. The ohmmeter scale is, therefore, a non-linear scale which is very much crowded towards

the high resistance side. High resistance values cannot be read very accurately on this scale. A different scale has to be provided for higher values of resistance. This can be done by increasing the battery voltage to get the required value of full scale deflection current for higher values of R_x. Thus, if the battery voltage is made 45 V, the resistance range can be increased by a factor of 10. To increase the resistance range by a still higher factor of 100, a battery of 450 V will be required which is not practicable. An alternative to use of high voltage battery is the use of a more sensitive meter like a 50 μA meter which will require very low values of current for suitable deflection with high values of resistance and a low battery voltage. In commercial multiple range ohmmeters, either separate scales are provided for different resistance ranges with a separate socket for each range or a range switch is used to select different multiplying factors for different ranges to be read on the same scale. Thus with the range switch in position $R \times 10$ as in Fig. 23.20, if the pointer shows a resistance value of 3.3 kΩ on the scale, the actual value of the resistance is 33 kΩ.

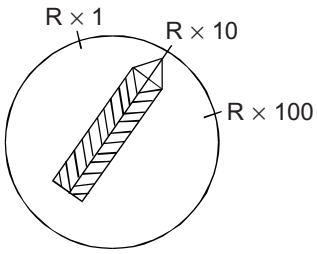


Fig. 23.20 Range Switch

23.6.2 Wheatstone Bridge

A Wheatstone bridge is a device for very accurate and precise measurement of resistances. Named after the British inventor Sir Charles Wheatstone, a Wheatstone bridge, in its simplest form, consists of four resistors arranged in a diamond-shaped figure as shown in Fig. 23.21.

A voltage source E is connected across points A and B and a sensitive galvanometer G with a centre-zero scale is connected between C and D. S_1 is a spring switch which is pressed for checking the reading in the galvanometer G. R_x is the unknown resistance to be measured and R_3 is a standard variable resistor which is calibrated in ohms. R_1 and R_2 are resistors whose ratio can be changed as desired. For measurement of the value of R_x, R_1 and R_2 are made equal and R_3 is adjusted till, on closing the switch S_1, the galvanometer shows no deflection. In this case no current flows through the galvanometer because points C and D are at the same potential. This is called the *null-method* of resistance measurement. When null point is achieved, R_x is equal to the value of R_3 as read from its calibration. For greater accuracy, the ratio of R_1 and R_2 can be made one to ten. With this ratio, if R_3 is 45 Ω, at null point, the actual value of R_x is 45/10 = 4.5 Ω. Still higher accuracy can be obtained by making $R_1/R_2 = 100$, and so on.

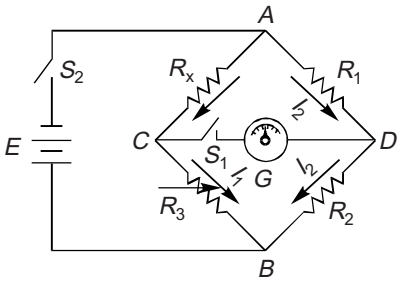


Fig. 23.21 A Wheatstone Bridge

To understand the working of a Wheatstone bridge, we can consider the entire bridge as consisting of two potential divider networks R_xR_3 and R_1R_2 connected across the same voltage source E. Points C and D will be at the same potential when the ratio of the voltage drops I_1R_x and I_1R_3 is the same as the ratio of the voltage drops $I_2 R_1$ and $I_2 R_2$. In other words, at null point

$$\frac{I_1R_X}{I_1R_3} = \frac{I_2R_1}{I_2R_2}$$

or
$$\frac{R_X}{R_3} = \frac{R_1}{R_2} \quad \text{or} \quad R_X = R_3 \cdot \frac{R_1}{R_2}$$

Thus with $R_1/R_2 = 1/100$, if the value of R_3 at null point is 452 Ω, the actual value of the resistance R_x is 4.52 Ω. The bridge method of measurement is a very useful method and with an AC source it can be used to measure capacitance and inductance in the same way as the measurement of resistance.

23.6.3 Megger

A megger is an instrument meant for testing the insulation of cables, motor windings and transformer windings. It is capable of measuring a very high resistance of the order of several thousand megaohms. Basically, a megger consists of a hand-operated generator and a moving coil meter contained in the same case. The generator is capable of generating voltages of the order of 100 V, 500 V, 1000 V, 2.5 kV or 5 kV depending upon the use for which it is required. The circuit arrangement of a megger is shown in Fig. 23.22.

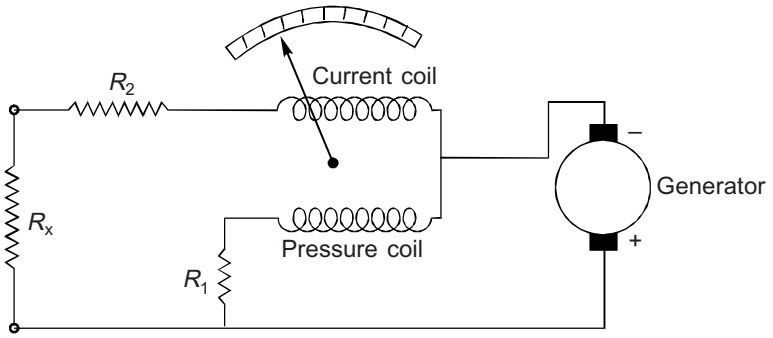


Fig. 23.22 Circuit of a Megger

Two coils called the current coil and the pressure coil are mounted at right angles on the same shaft. These coils are so wound and mounted that they tend to deflect the meter in opposite directions when carrying current. The equilibrium position will depend on the current in the two coils. The pressure coil is connected across the generator in series with a fixed high resistance R_1 and the current coil is connected in series with the unknown insulation resistance R_x through a fixed resistance R_2. When R_x is infinity, current will flow only through the pressure coil and the pointer will deflect to one extreme position

marked "infinity" (∞) on the scale. When the unknown resistance R_x is zero, i.e. the output terminals are shorted, the pointer will deflect to another position marked "zero" on the scale. The scale between ∞ and zero can be calibrated by inserting different known resistances in the position of R_x.

The generator is rotated by a handle outside the instrument. A constant speed clutch ensures that the generator speed does not exceed a maximum limit, no matter how fast the handle is rotated. Further, if the handle is suddenly stopped, a freewheel allows the generator to come to rest slowly.

23.7 MULTIMETERS

A multimeter, also known as VOM (Volt-Ohm-Milliammeter), is a combination of a voltmeter, ohmmeter and a current meter all contained in a single case. The same meter movement is used to measure all the three quantities—voltage, resistance and current. Various multipliers, shunts and batteries are all contained inside the case. Selector switches are provided on the front panel to select a particular function and the particular range for that function. A single selector switch often selects the function as well as the range for that function as in Fig. 23.23. In certain multimeters jacks or sockets are used, instead of selector switches, to select the function and the range. A single dial with scales calibrated to read volts, ohms and milliamperes with a number of ranges for each quantity is generally used. Most multimeters are designed to measure both DC and AC quantities with a selector switch to select DC or AC function of the multimeter.

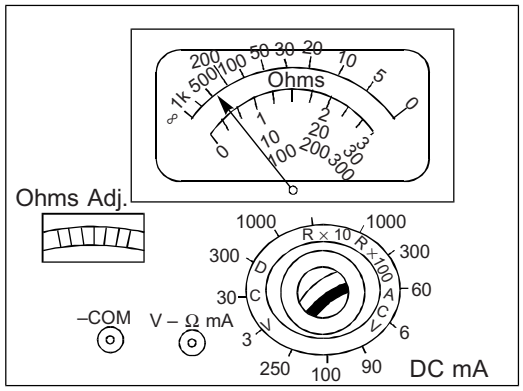


Fig. 23.23 A Multimeter

The basic circuit of a multimeter with three current ranges, three voltage ranges and two resistance ranges is given in Fig. 23.24.

23.7.1 Electronic Multimeters

Vacuum-Tube Voltmeters (VTVM)

VOM type of multimeters described so far have, at best, a sensitivity of 20,000 Ω/V and the accuracy of measurments is limited. This difficulty is overcome in

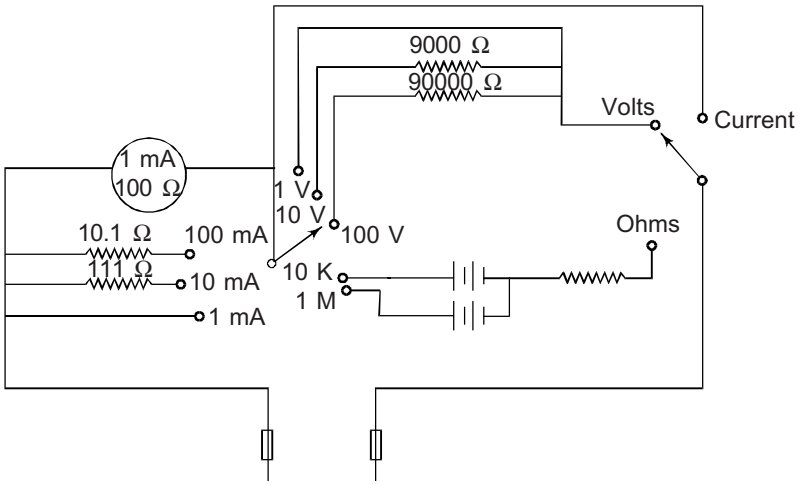


Fig. 23.24 Basic Circuit of a Multimeter

another type of multimeters called the *vacuum tube* voltmeter or VTVM which has a very high input resistance, of the order of 11 MΩ on all DC voltage ranges. As the name implies, the VTVM uses vacuum tubes or valves for its operation and needs an internal power supply. Transistors are sometimes used instead of vacuum tubes when batteries can be provided to make the unit portable. Figure 23.25 shows a normal VTVM with accessories.

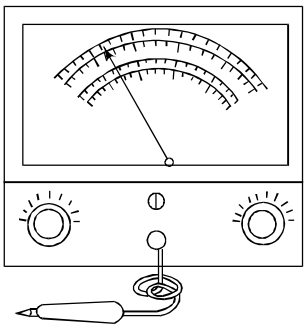


Fig. 23.25 A Vacuum Tube Voltmeter (VTVM)

High input resistance is the main advantage of a VTVM over an ordinary multimeter or VOM but the same high resistance makes it unsuitable for the measurement of currents, both DC and AC. It can measure AC and DC voltages and resistance but cannot measure currents. A VTVM can provide very high resistance ranges because the high voltage required for these high resistance ranges is provided by its internal power supply source.

When provided with a special RF probe, a VTVM can measure RF voltages up to 250 MHz or so. Such HF AC voltages cannot be measured by ordinary multimeters.

23.7.2 Digital Multimeters

Digital multimeters with a visual digital display (Fig. 23.26) are becoming quite popular. A digital multimeter (DMM) is characterised by high input impedance, better accuracy and resolution.

The digital multimeter converts an input analogue signal into its digital equivalent and displays it. The analogue signal input might be a DC voltage an AC voltage, a resistance or an AC or DC current. The operation of only one

function switch enables the user to get various facilities and ranges, etc. automatically indicated on the visual display.

23.8 AC MEASUREMENTS

In an AC circuit, the polarity of the voltage or current to be measured keeps on changing after every half cycle. An ordinary moving coil meter will not respond to AC changes because the pointer will tend to move in one direction during the positive half cycle and in the opposite direction during the negative half cycle and the net result will be no movement of the pointer at all. For getting a deflection of the meter in AC circuits, either some other type of meter movement has to be used or some different method must be adopted for the measurement of AC quantities.

The moving iron type of meter described earlier is capable of measuring both AC and DC currents but these meters are not very sensitive and their use is restricted to power circuits only.

The *thermal type* of meters are also capable of measuring both AC and DC currents because the heating effect of current is independent of polarity. Of the two types of thermal meters described earlier, the hot wire and the thermocouple type of meters are both capable of measuring AC and DC current. Both the types are suitable for RF measurements. However, the thermocouple meter makes use of moving coil meter movement and it is used exclusively for the measurement of (RF) currents. Thermal type meters have a non-linear scale.

Fig. 23.26 Digital Multimeter

Rectifier Type AC Meters

Of all types of AC meters, the most commonly used are the *rectifier type* instruments. In this meter the AC voltage is first rectified or converted into DC voltage with the help of a *diode* and the DC voltage so produced is measured with a moving coil meter which is calibrated to read rms values. Such a rectifier circuit is incorporated in a multimeter and comes into operation only when the AC ranges are selected either with the selector switch or with the AC jacks or sockets.

Various types of rectifiers are available for use with rectifier type AC meters. These include selenium, copper oxide, germanium (semiconductor) or vacuum tube rectifiers. A number of rectifier circuits can be used for converting AC into DC and these have been described in detail in an earlier chapter. However, a schematic diagram of the AC meter is given in Fig. 23.27.

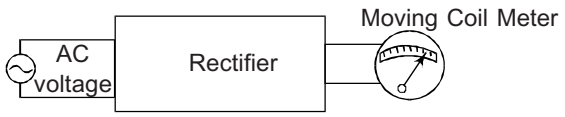


Fig. 23.27 Rectifier Type AC Meter

Rectification being non-linear at low voltages at 10 V or less, a separate scale is generally provided for 0 – 10 V AC. A rectifier type of instrument cannot measure AC currents because of its high input resistance.

23.9 MEASUREMENT OF POWER

The power P consumed in an electrical circuit is the product of the voltage E and the current I in the circuit. Thus $P = E \times I$ watts if E is in volts and I in amperes. If the voltage across the circuit can be measured with a voltmeter and the current with an ammeter as in Fig. 23.27, the power in watts will be the product of voltmeter reading and the ammeter reading.

If the value of the load resistance R_L is known is ohms, power in watts can also be calculated by use of the formula

$$P = I^2 R_L \quad \text{or} \quad P = \frac{E^2}{E_L}$$

Instead of using two meters as in Fig. 23.28, it is possible to read power directly in watts by means of a single meter called a *wattmeter*.

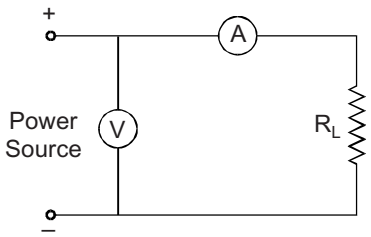


Fig. 23.28 Measurement of Power

23.9.1 Wattmeter

Figure 23.29 shows the circuit of a *wattmeter*. Two fixed coils L_1 and L_2 constituting an electromagnet take the place of the permanent magnet in a moving coil meter and these are placed in series with the circuit. The magnetic field produced is proportional to the current I flowing through these coils known as the *field coils*. A moving coil L_M, with a limiting resistance R is connected across the circuit and the magnetic field produced in this coil is proportional to the applied voltage E. The deflection produced in the moving coil is due to the reaction of the two magnetic fields and is proportional to the product $E \times I$ which is the power in the circuit. A pointer attached to the moving coil L_M reads the power on a scale calibrated in watts. In an AC circuit, the wattmeter will read the *true power* by taking into account the power factor of the circuit.

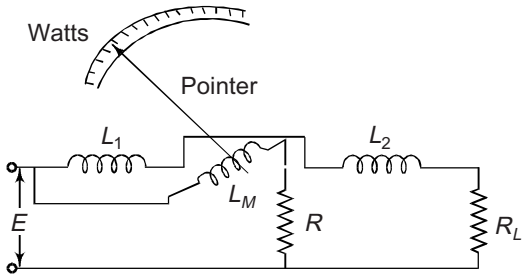


Fig. 23.29 Circuit of a Wattmeter

23.9.2 Watt-hour Meter

The type of wattmeter used for measuring the consumption of electrical energy in household circuits is called a watt-hour meter which is shown in Fig. 23.30. In this meter, the moving coil takes the form of the rotor of an electric motor whose speed of rotation will depend on the current through the armature of the rotor. The number of revolutions made by the motor is indicated by a gearing arrangement on four dials calibrated in thousands, hundreds, tens and units. The reading given by these dials is in kilowatt-hours which is the unit of energy consumption. An aluminium disc attached to the extension of the shaft of the rotor moves between the poles of two permanent magnets and this provides the necessary damping to the motion of the motor due to eddy currents produced in the aluminium disc.

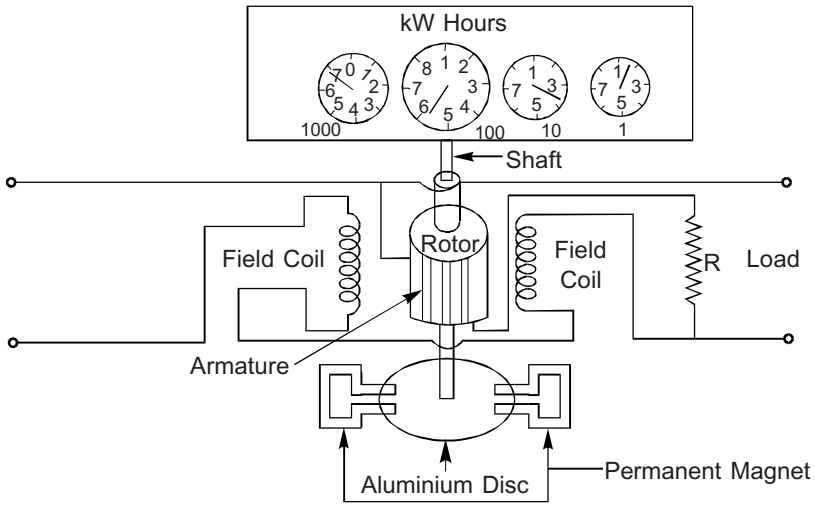


Fig. 23.30 Watt-hour Meter

23.10 OTHER TEST AND MEASURING EQUIPMENT

23.10.1 Signal Generators (Function Generators)

A signal generator is an equipment that supplies a standard voltage of known amplitude, frequency and waveform for test and measurement purposes.

Signal generators are classified according to the shape of the output waveform, e.g. sine wave oscillators, square wave generators, noise generators, sawtooth generators etc. These are also called function generators. They may be classified according to the range of frequencies they generate. For example, the audio generators, IF-RF generators, VHF-UHF generators, microwave generators etc. There are mostly sine wave generators and the output frequency is variable. Each type usually includes several ranges or bands.

Both AM and FM types of signal generators are available. AM signal generators are used in troubleshooting and aligning of radio receivers and for signal injection tests in television receivers.

23.10.2 Oscilloscope

A cathode ray oscilloscope (CRO) is one of the most useful and versatile test instruments not only for TV maintenance and servicing but also as a general purposes measuring and test equipment for many other electronic circuits.

The CRO produces a visual display of the waveform of time-varying voltages and currents. The instrument plots automatically a graph of voltage or current versus times of a wide range of frequencies and amplitude of AC voltages and currents. By suitable manipulation of the controls it is also possible to obtain a stationary display of one or more cycles of a periodic waveform which can either be photographed or studied at leisure.

A block diagram indicating the essential features of a CRO is given in Fig. 23.31. The functions of the various blocks are given below.

1. The CRT indicator is a normal CRT with electrostatic deflection and respective controls for brightness, focus and picture position.

2. The vertical amplifier (Y-amplifier) provides sufficient amplification for the input signal before applying the same to the vertical deflection plates. The input attenuator attenuates the high voltage signals so that they do not overload the amplifier.

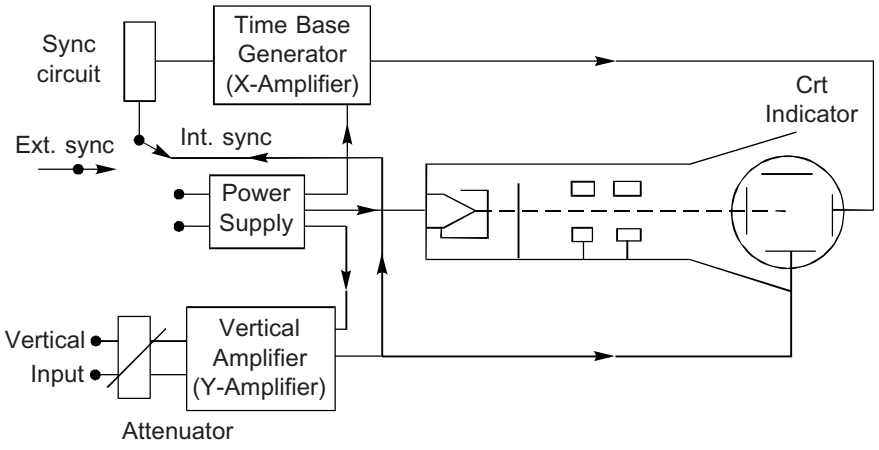


Fig. 23.31 Block Diagram of a Cathode Ray Oscilloscope

3. A time base generator produces a sawtooth voltage which varies linearly with time. When this sawtooth voltage is applied to the horizontal deflection (X-deflection) plates of the CRT, it produces a horizontal deflection of the electron beam. During the trace period, the rising voltage causes the electron beam to sweep horizontally across the screen of the CRT and during retrace interval, the falling voltage causes the beam to return rapidly to its initial position. The linear horizontal deflection of the beam due to the input signal produces on the CRT screen a graph of the input voltage versus time. Persistence of vision and phosphorescence of the CRT screen help in producing a clear visual display. The sawtooth generator has a variable frequency range which determines the frequencies that can be seen on the screen. There is a definite relation between the frequency of the horizontal sweep f_H, the frequency f_V of the input signal applied to the vertical plates and the number of complete cycles 'n' of input frequency that can be seen on the screen. This relation can be expressed as

$$\frac{f_V}{f_H} = n$$

Thus, if the sawtooth generator frequency is 50 Hz and a 50 Hz signal is applied to the input of the vertical amplifier, one complete cycle will be visible on the screen and with an input signal frequency of 100 Hz, two complete cycles will be seen and so on.

The usual range of time-base frequencies for a general purpose oscilloscope is from several cycles to 100 kHz.

4. Synchronisation of the time-base generator frequency and the vertical signal frequency is necessary for producing a stationary wave pattern on the screen. For this the time-base generator frequency has to be equal to or a submultiple of the signal frequency. This synchronisation or "locking-in" is provided by the sync, circuits. A sync. circuit is generally a simple RC circuit which feeds a part of the signal from the vertical amplifier to the time-base generator.

This locking-in or synchronisation of the sawtooth wave generator can be either internal as explained above or some external frequency can be used for this synchronisation by connecting the external source of frequency to the terminal marked, 'External sync'. In this case, the 'Internal External sync.' switch is changed over to the 'External sync.' position.

5. One of the most important characteristics of an oscilloscope is its frequency response by which is meant the frequency response of its vertical amplifier. Normal commercial type oscilloscopes have a low frequency response between 10 and 100 Hz and a high frequency response between 100 kHz and 5 MHz. Oscilloscopes for specialised work may have a high frequency limit of 10 MHz or even more. However, the frequency response of an oscilloscope need not extend much beyond the response of the circuits for which it is to be used because the cost of the equipment mounts up as the frequency response limits are raised.

It is necessary to mention that the low frequency response is as important as the high frequency response of the equipment for studying proper performance of TV circuits.

In TV maintenance and servicing, the oscilloscope is used for:

(a) Studying the waveforms at different points of the video and deflection circuits.
(b) Alignment of the RF/IF stages with a sweep oscillator.
(c) Measuring peak-to-peak AC voltage and periodicity of the waveforms under study.

All these observations provide useful clues to possible fault in the receiver circuits.

For measurement of peak-to-peak AC voltages, some of the oscilloscopes are provided with a calibrated vertical amplifier so that these voltages can be read directly. In the absence of such a calibration, a known voltage from 50 Hz mains can be used to calibrate the scale on the screen.

Peak-to-peak measurements of AC currents can also be made with the CRO by inserting a known resistance in series with the circuit and measuring the voltage drop across this known resistance. Application of Ohm's law enables the current to be calculated.

As a general-purpose testing and measuring equipment, the oscilloscope can also be used for frequency measurements by comparing the unknown frequency with a voltage source of known frequency. Two methods are available for such a comparison. In the first method, the voltage of unknown frequency is connected to the input of the vertical amplifier of the oscilloscope and a stationary trace of two complete cycles is obtained by adjusting the sweep frequency. If the unknown voltage is within the audio range, an audio oscillator is connected in place of the unknown frequency source and the frequency of the audio oscillator is adjusted till two cycles of the wave are again seen on the screen. The frequency of the audio oscillator and the unknown frequency are equal.

The second method is by the use of Lissajous figures which are produced on the screen when voltages of two different frequencies are applied to the horizontal and vertical inputs of the oscilloscope. The Lissajous pattern produced depends on the ratio of the frequencies applied to the horizontal and vertical deflection plates of the CRT as shown in Fig. 23.32.

The frequencies of the sources applied to the horizontal (f_H) and the vertical (f_V) deflection plates are in the same ratio as the number of times the pattern is intersected by the vertical and the horizontal coordinate axes respectively. Where a curve crosses itself on an axis, two intersections on the line are counted. Thus in Fig. 23.32(a), the curve produces two inter-sections with each of the coordinates and so $f_H/f_V = 2/2 = 1$, the two frequencies are equal. In Fig. 23.32(b) $f_H/f_V = 1/2$, the frequency applied to the horizontal deflection plates is one-half the frequency applied to the vertical deflection plates and so on. If one of the frequencies is known the other frequency can be calculated.

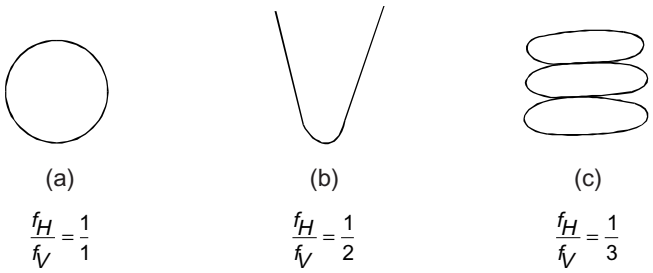


Fig. 23.32 Lissajous Figures

A CRO is a very sensitive and reliable test equipment provided the technician is quite conversant with the use and manipulation of its various controls. This needs both skill and experience. For satisfactory results, proper manipulation of the controls of an oscilloscope is as essential as the technical knowledge of the circuits of the TV receiver. A commercial type of CRO will normally have the controls shown in Fig. 23.33.

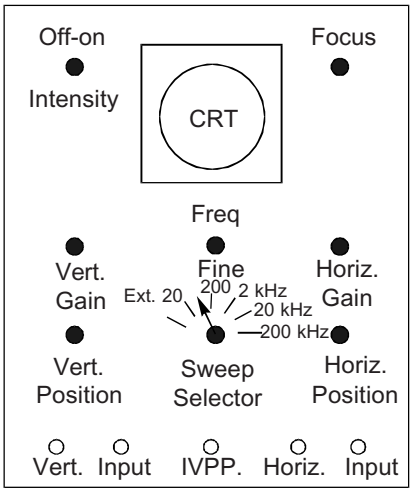


Fig. 23.33 Oscilloscope Controls

Visual graphs or traces of waveforms produced on an oscilloscope screen are called oscillograms. These oscillograms are quite helpful in checking and adjusting the frequency response of wideband circuits like the RF/IF circuits of a TV receiver. For this, another test equipment called the sweep generator is required.

Sweep Generator

A sweep generator is an RF oscillator whose frequency can be varied or swept over the alignment range electronically by means of a sweep voltage which could be a sine wave voltage at mains frequency of 50 Hz. For producing an oscillogram of the IF frequency response of a TV receiver, the IF output from the sweep generator is applied to the input of the IF amplifier stage and the output from the video detector is connected to the *vertical input* of the oscilloscope. The internal sweep of the oscilloscope is disconnected and a 50 Hz sine wave voltage from the sweep generator is applied to horizontal input of the oscilloscope. A phase control on the sweep generator is adjusted till a steady oscillogram is obtained on the screen of the oscilloscope. For indicating the correct frequencies on the response curve, crystal controlled oscillator or an RF signal generator is used as a marker for specific frequencies. Most sweep generators have a built-in marker which produces pips at known spot frequencies. For IF alignment, the coupling circuits and wave traps are adjusted till the

IF response curve on the oscilloscope fits the marker frequencies correctly. This method of alignment is much more accurate than the conventional method of measuring output voltage at frequencies over the alignment range.

A test instrument which combines all the service facilities of wobbulator or sweep generator, a CRO and marker generator is known as a *wobbuloscope*.

Methods of testing and alignment of TV receivers described above are more suited for well-equipped laboratories or workshops rather than for fieldwork or simple repair shops. A test instrument of great practical utility for a TV technician is a *pattern generator*.

Pattern Generator

A pattern generator is a special type of RF signal generator which is designed to produce video signals direct and RF signals amplitude modulated with video signals for testing and alignment of TV receivers. A pattern generator enables a service technician to check up the performance of a TV receiver even when the TV station is not on. It can be used to check up certain features like vertical and horizontal linearity which cannot be easily checked up during a regular TV transmission. Portable types of pattern generators make on the spot servicing of TV receivers a very convenient and efficient job.

Video patterns produced by a pattern generator generally consist of horizontal bars, vertical bars, grills and chessboard patterns. Some of them are shown in Fig. 23.34. A selector switch allows the selection of the pattern to be studied. Besides the video pattern and amplitude modulated RF signals, certain pattern generators, also provide FM signals spaced 5.5 MHz from the video carrier together with 1 kHz audio signals. FM signals at 5.5 MHz inter-carrier sound IF, are also available in most pattern generators for testing of the sound IF section.

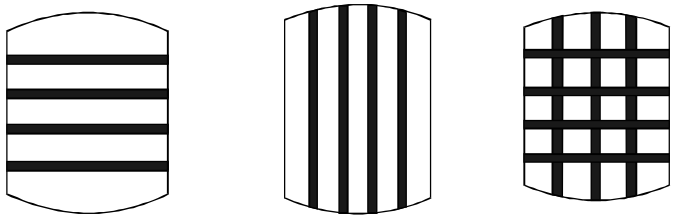


Fig. 23.34 Patterns Produced by a Pattern Generator

The circuit of a pattern generator comprises a stable 15625 Hz sine wave oscillator for the line frequency and the 50 Hz mains frequency controls the vertical frequency source. Blanking and sync. pulses are produced by multivibrators whose frequency can be varied to change the number of bars in a pattern. To avoid complicated circuitry and to reduce cost, most pattern generators produce patterns without interlaced scanning employed in a normal TV picture. The test picture produced by a pattern generator, therefore, consists of 50 pictures per second with about 312 lines per picture. As such the

picture from the test instrument is generally coarser than the picture from a TV station. Without interlaced scanning, the equalising pulses present in a normal TV signal are missing from the test pattern produced by the PG. However, for checking all other functions of a TV receiver except interlacing, the signal from a pattern generator is as suitable as a signal from a TV station.

Pattern generators are available for single channel operation with only 'high' and 'low' outputs. However, more sophisticated and expensive types provide RF signals with continuously variable output over a number of channels in the VHF/UHF range of frequencies (band I, III, IV and V). High output is required for picture IF and video detector stages but low output is suitable for testing the tuner and first IF amplifier stages.

A pattern generator can prove a very useful tool in the hands of a skilful TV technician not only for checking up the overall performance of a TV receiver but also for finer adjustments such as horizontal and vertical linearity, picture height and width and operation of AGC.

23.10.3 Television Test Patterns

Test patterns are optical charts which when reproduced on a TV screen provide useful information regarding the performance of the TV receiver by means of certain geometrical figures and forms depicted on the test pattern. A standard test pattern accompanied by a 1 kHz tone is normally transmitted by TV stations before the start of a regular transmission. This is done to enable the viewers and service technicians to evaluate the picture quality of their TV receivers and to make such adjustments as may be necessary to make the performance of the receiver conform as closely as possible to the standards indicated by the test pattern. One such test pattern is shown in Fig. 23.35.

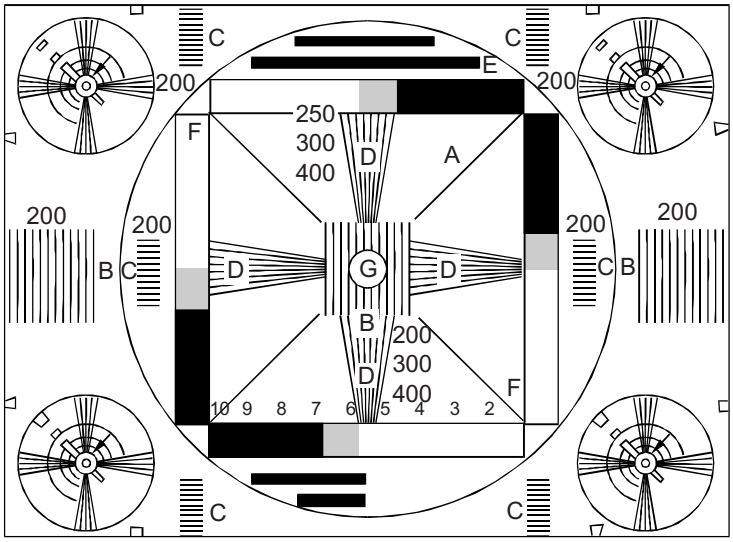


Fig. 23.35 Standard Test Pattern

Some of the important features that can be checked with the help of a test pattern are described below.

Linearity

Overall linearity is checked by the large white circle marked *A* at the centre of the pattern. For proper linearity and aspect ratio the circle should be a perfect geometrical circle and should be in the centre of the test pattern. If the circle is distorted at the top and bottom, the receiver lacks vertical linearity and distortion on the side of the circle indicates horizontal non-linearity. The horizontal and vertical linearity can be adjusted by their respective controls.

The three broad vertically striped patterns marked *B* at the centre and on the left and right hand sides of the picture help in adjustment of the horizontal linearity. When properly adjusted the horizontal dimensions and spacings of the stripes should be equal.

The vertical linearity is indicated by the narrow horizontal striped patterns marked *C* near the top and bottom edges of the picture. The width and dimensions of these stripes should be uniformly equal.

Both the striped patterns (*B* and *C*) mentioned above have a resolution of 200 lines.

Resolution

Resolution is the ability of the receiver to show distinctly and separately fine details of the picture, both in the vertical and horizontal direction.

The vertical and horizontal resolution is indicated by the resolution wedges *D*. The vertical resolution is indicated by the horizontal wedges and the horizontal resolution by vertical wedges. The figures on the tapered wedges indicate the number of resolution lines and corresponding bandwidth of the system. 400 lines correspond to a video bandwidth of about 5 MHz. Resolution wedges provided in the circles in the four corners of the pattern indicate the resolution in their respective regions.

Brightness and Contrast

Four vertically and horizontally arranged gradation bars (marked E) indicate brightness and contrast. A contrast ratio of 30 : 1 is provided by ten squares marked 1 to 10 on the gradation bars. When the brightness and contrast controls are properly adjusted distinct difference in shading should be noticeable for each square.

Interlacing

This is indicated by the diagonal lines (marked 'F') inside the white circle. With good interlacing the lines are thin and uniform but with unsatisfactory interlacing the lines are thick and broken in steps.

Focus

Concentric circles at the centre of the large circle (*G*) if clear and sharp indicate good focussing. The sharpness of the pattern as a whole also indicates the degree of focussing in the picture tube.

SUMMARY

Meters are instruments used for the measurement of current, voltage, resistance and power. Basically all meters are current meters with suitable modifications to measure the other three quantities.

A meter movement is a device for producing deflections of a pointer based on any of the effects of current. Two of the effects used for the measurement of current are the magnetic effect and the thermal or heating effect of current. Moving coil and moving iron meters are based on the magnetic effects of current. Of these the moving coil meter is the most common type of meter and is also known as a galvanometer. It can measure only DC currents. Meters based on the thermal effects of current are the hot wire meter and the thermocouple meter. Both these types are suitable for measurement of radio frequency (RF) currents.

An ammeter or a milliammeter is the meter used for the measurement of current. The amount of current required for full scale deflection is the sensitivity of a current meter. By itself an ammeter can measure only the current required for full scale deflection. For extending the range of the current meter low resistances known as shunts are connected in parallel with the coil of the meter.

A voltmeter is also a current meter connected across the two points between which the potential difference is to be measured. By itself a voltmeter can measure a voltage equal to the voltage drop (IR) produced across the coil by the full scale deflection current. The range of the voltmeter can be extended by connecting resistances called multipliers in series with the coil resistance. The higher the series resistance the more accurate is the voltmeter.

The meter used for the measurement of resistance is an ohmmeter. Any meter in conjunction with a battery can be used to measure resistance. The range of an ohmmeter can be extended either by using a battery with a higher emf or by using a meter with a higher *sensitivity*. A Wheatstone bridge is used for very accurate measurement of resistances. The meter used for the insulation test of cables, transformers and coils is called a megger.

A multimeter, also known as a VOM (Volt-Ohm-Milliammeter) is a combination of a voltmeter, ohmmeter and a current meter. It can measure both AC and DC quantities. An electronic multimeter is known as a VTVM (Vacuum Tube Voltmeter) and has got a very high input resistance. It measure only AC and DC voltages and resistances but cannot measure currents.

A moving coil meter cannot measure AC currents. For measurement of AC quantities either the moving iron type meters or the thermal type meters are used. The most common method of measuring AC is the rectifier method in which AC is converted to DC and then measured with a moving coil meter.

Power can be measured by measuring the voltage and current in a circuit and finding the product of the two. A single meter which indicates power in a circuit, is called a *wattmeter*. A *watt-hour meter* is used for measuring energy

consumed in household circuits in kilowatt-hours which is the unit of energy consumption.

Other test and measuring equipments used for electronic measurements are signal generators and oscilloscopes.

Review Questions

1. Name the meters used for the measurement of current, voltage, resistance and power?
2. What is a meter-movement? What are the two basic types of current meters?
3. Describe with a sketch a moving coil current meter. What are the other two names for this type of current meter?
4. What are shunts? What is the use of shunts in current meters?
5. (i) Calculate the value of the shunts required to extend the range of a moving coil meter to 100 mA when the values of the full scale deflection current and internal resistance of the meter are (a) 10 mA, 9 Ω, (b) 1 mA, 30 Ω, (c) 100 μA, 100 Ω.
   (*Ans.* (a) 1 Ω, (b) 0.3 Ω, (c) 0.1 Ω approx)
   (ii) Calculate the value of the universal shunt and the position of tappings for the current ranges of 2 mA, 10 mA, and 100 mA to be provided in a 1 mA 50 Ω meter.
   (*Ans.* 50 Ω, 40 Ω, 9 Ω and 1 Ω)
6. How can a moving coil meter be used to measure voltage?
   A moving coil meter has a coil resistance of 50 Ω and draws a full scale current of 1 mA. Calculate the value of the multiplier resistance if the meter is to read (a) 1 V full scale, (b) 100 V full scale.
   (*Ans.* (a) 950 Ω, (b) 99950 Ω)
7. A 50 μA meter has an internal resistance of 1000 Ω. What is the minimum voltage it can measure? Calculate the multiplier resistance required to extend the voltmeter range to (a) 10 V, (b) 30 V, and (c) 300 V.
   (*Ans.* 0.5 V, (a) 199,000 Ω, (b) 599,000 Ω, (c) 599,000 Ω)
8. In Example 7 calculate the Ω/V rating of the voltmeter and show that it is the same for all ranges.
9. Explain the following:
   (a) Why should a milliammeter have a low resistance and a voltmeter a high resistance?
   (b) Why can a moving coil meter measure only DC current but a hot wire meter measure both DC and AC currents?
10. Describe methods for the measurement of AC voltages.
11. Describe the principle of a Wheatstone bridge. What are the advantages of a Wheatstone bridge over an ordinary ohmmeter?

12. Write short notes on:
    (a) Megger
    (b) VTVM
    (c) Wattmeter.
13. Describe a signal generator and give its important uses.
14. Give the block diagram of an oscilloscope and describe its important applications.

chapter 24

Troubleshooting, Maintenance and Repair of Electronic Equipment

There has been a tremendous advance in the field of electronics and the development of electronic equipment particularly after the replacement of electron tubes and valves by transistors and other solid-state devices like IC with thousands of components mounted or printed on small chips. The advent of microprocessors of different configurations has not only added to the complexities of the design of modern electronic equipment but has also raised the problem of maintenance and repairs of these electronic equipment for trouble free working. The maintenance and repair of these sophisticated electronic equipment not only requires specialized test and measurement devices but also modern specialized skills, training and techniques on the part of those engaged in the maintenance and repair of these electronic equipment.

Some of the methods and techniques used for the maintenance and repair of such common electronic devices as radio receivers, PA systems, tape recorders, VCRs and televisions receivers will be discussed in this chapter.

24.1 CAUSES OF FAILURES

The causes of failures of electronic equipment can be divided into various categories. These may include:

(i) Careless handling during transport and storage.
(ii) Defective design or production deficiencies.
(iii) Poor environmental conditions like hostile weather, humidity and dusty atmosphere.
(iv) Operational errors and lack of operations skill.
(v) Fluctuations in mains voltage.
(vi) Aging of the equipment.
(vii) Lack of proper maintenance and care.

24.2 RELIABILITY FACTORS

Modern electronic equipment is used in some of the most important and sensitive devices like satellites, aircrafts, medical life support equipment and process control in industry. Reliability of such equipment has to be of very high order because failure of such equipment due to any cause whatsoever can have disastrous results.

The following factors decide the reliability of an electronic equipment.

24.2.1 Failure Rate

The reliability of any equipment or system is determined by the failure rate of various types of components used in the system. The smaller the percentage failure rate for each component, the better the reliability of the equipment.

24.2.2 MTTF (Mean Time To Fail)

This can be calculated from the failure rate of components used in the equipment. For example if an equipment uses only one transistor with failure rate of 2.5% per 1000 hours, then

$$\text{MTTF} = \frac{1}{2.5 \times 10^{-5}} = 40{,}000 \text{ hours} = 1666 \text{ days}$$

MTTF is normally applicable to small items like resistors, capacitors, diodes, transistors which can be thrown away and need no repairs.

24.2.3 MTBF (Mean Time Between Failures)

MTBF for a system is measured by listing it for a period T, during which N failures take place. Each time the fault is cleared and the equipment put back on test then

$$\text{MTBF} = \frac{T}{N}$$

This value does not include the time for repairs. The MTBF for an equipment is the total sum of the failure rate of components used. The failure rate for various components used is provided by the manufactures.

24.2.4 MTTR (Mean Time To Repair)

A system may have excellent reliability but if the time taken for repairs during a failure is large the system will have poor availability.

MTTR depends on a number of factors such as:

(i) Availability of service manuals and correct circuit diagrams and lay-out of components.
(ii) Clearly labelled components and plug in circuit boards and components.
(iii) Availability of spare parts and components.
(iv) Availability of suitable trained technical and operating staff.

The MTTR improves with time after the teething troubles are over and the staff gets familiar with the use of the equipment.

24.3 MAINTENANCE PROCEDURE

For troublefree operations and long life of any electronic equipment a regular and systematic procedure for maintenance of the equipment should be adopted and meticulously followed. Maintenance procedure can be divided into two parts viz Preventive Maintenance and Corrective Maintenance.

24.3.1 Preventive Maintenance

Preventive maintenance consists of dusting, cleaning and inspection of switches, relays, interlocks and other safety devices to avoid breakdown in service and damage to the equipment and injury to operational staff. Preventive maintenance is a regular and routine process which should be carried out in accordance with a maintenance schedule prepared in consultation with the manuals and instruction booklets supplied by the manufacturers of the equipment.

Proper log books should be maintenaned where meter readings and other observations made by the operational staff should be entered. Any abnormalities in the behaviour of the equipment should be instantly brought to the notice of the supervisory staff for necessary corrective action.

Thus, preventive maintenance can avoid sudden or frequent breakdowns in service to avoid annoyance to viewers of TV or listeners of Radio programmes.

Prevention is better than cure is an old saying which applies equally to human nature and electronic equipment.

24.3.2 Corrective Maintenance

Corrective maintenance consists of the replacement or repairs to any components that may fail during operation. To bring the equipment back to normal working condition a stock of spare parts and components should be readily available. To avoid any complications only the original spares supplied by the manufacturers should be used for replacing the defective components. Substitutes and equivalents should not be used for replacement unless recommended by the manufacturers.

While carrying out replacement and repairs the power supply to the equipment must be switched off and any switches, relays or interlocks provided with the equipment should not be shorted or bypassed when testing the equipment after repairs. This is essential to avoid damage to the equipment and injury to the operational staff.

24.4 COMPONENTS

Two types of components are used in building up electronic circuits and equipment. These are Passive Components and Active Components.

24.4.1 Passive Components

Passive components are those that control or modify the output of an electronic circuit without playing an active role in its performance. Examples of passive components are resistors, inductors (including transformers) and capacitors. Full details of properties, constructional details and applications of these components have already been discussed under relevant chapters of the book.

24.4.2 Active Components

Active components used in the modern electronic equipment use semiconductor devices and electron tubes and valves.

Semiconductor devices mainly used are:

1. **Diodes**, including crystal diodes, Zener diodes, varactor diodes, tunnel diodes, light emitting diodes, photodiodes.
2. **Thyristors** which include silicon controlled Rectifiers, Diacs, Triacs and UJTS.

   The applications and properties of these devices have also been discussed earlier in the text.
3. **Transistors** including bipolar transistors FETs, power transistors and photo transistors.
4. **Thermionic tubes** (electron tubes) which includes vacuum diodes, triodes, tetrodes, pentodes and cathode ray tubes. Thermionic tubes or valves have almost been completely replaced in electronic circuit except cathode ray tubes and these have already been described under electronic measuring equipment in under Chapter 23.

A brief description of valves is given in appendix.

24.5 FAULT LOCATION

Fault location is a systematic procedure based on the careful observations of any abnormal meter readings. Excessive sound or smell coming out of any component due to overheating, etc. should be carefully watched . For speedy and rapid fault location the circuit diagrams and maintenance manuals provided by the manufacturers should be used. Proper use of test and measuring instruments like multimeters (both analog and digital), oscilloscopes, signal generators and function generators, etc. also helps in fault location and subsequent repair.

24.6 TROUBLESHOOTING TECHNIQUES

Troubleshooting techniques used for fault finding and repair of electronic equipment will mainly depend on the type of electronic equipment under observation and the personal skill and ability of the technicians handling the equipment. Although there are no set rules and regulations for troubleshooting but some of the methods and techniques used in this type of work are described as follows:

24.6.1 Functional Area Approach

This method consists of drawing a functional diagram or a block diagram of the equipment and locating the fault to any particular block and then to any particular component in that block. This particular component is thus repaired or replaced.

For example Fig. 24.1 shows the block diagram of a TRF radio receiver indicating different blocks that constitute the functional areas for fault location.

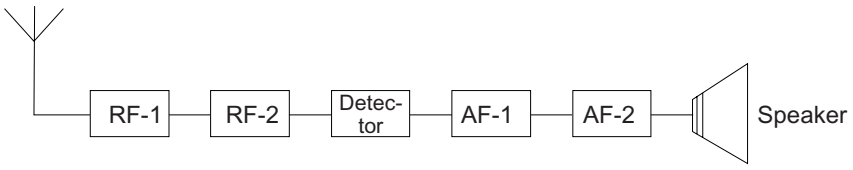


Fig. 24.1 Block Diagram of a TRF Receiver

24.6.2 Split Half Method

In this method of troubleshooting, the entire circuit is split into two halves and the output checked at the half-way point. The defect will be observed in either the first or second half. If this does not help, the second half is further divided into two halves and the process continued till the defective portion or component is isolated for necessary repair and replacement.

24.6.3 Divergent Paths

In divergent paths the output from one block feeds two or more blocks as in the case of a power supply system shown in Fig. 24.2.

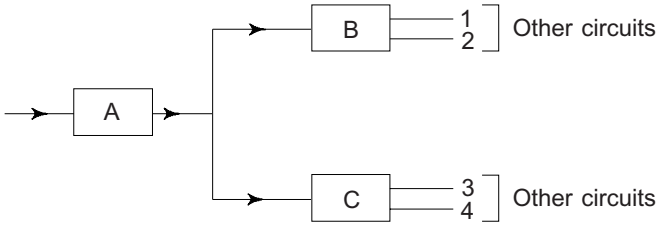


Fig. 24.2 Divergent Paths

In this case it is best to start by checking the common fault point. If one output is normal, check after the divergent point. However, if one output is abnormal, check before the common point.

24.6.4 Convergent Paths

In convergent paths, two or more inputs feed a circuit block as in Fig. 24.3.

To check the performance of such a system, all inputs at the point of convergence must by checked one by one. If all found to be correct the fault lies beyond the convergence point. However if any of the inputs is incorrect, then the fault lies in that particular input point.

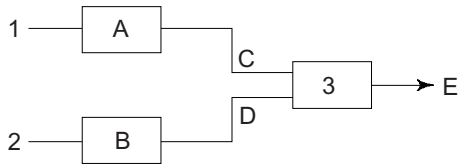


Fig. 24.3 Convergent Path

24.6.5 Feedback Paths

Feedback paths are normally provided in a circuit to improve or modify its performance. In modifying feedback it may be possible to break the feedback loop and convert the system into a straight linear data flow system. Each block can then be listed separately without the fault signal being fed around the feedback loop.

24.7 SEMICONDUCTOR DEVICES

Semiconductor devices mostly used in electronic circuit includes diodes, transistors (Bipolar) and unipopular transistors like FET, JFET and MOSFETs.

All these components have been fully discussed in Chapter 6 and their use in electric circuits and methods of testing, etc. have all been described earlier.

24.7.1 Linear Integrated Circuits

These circuits contain a large number of components integrated into specific circuits and printed on the surface of semiconductor materials by special printing and manufacturing processes described earlier.

The circuits printed on ICs consist of various types of amplifiers such as DC amplifiers and a special type of amplifiers called operational amplifiers (op-amps). All these aspects of linear ICs have been described in Chapter 7.

24.8 TROUBLESHOOTING DIGITAL CIRCUITS

The troubleshooting methods described so far apply to analogue equipment and circuits but the introduction of digital circuits in modern electronic equipment has added another dimension to the troubleshooting methods by the addition of techniques applicable to digital equipment and circuits only.

Digital circuits including logic gates and their functions have already been explained in Chapter 8. Troubleshooting methods and the special equipments required for digital troubleshooting will now be discussed.

24.9 TYPICAL FAULTS IN DIGITAL CIRCUITS

Digital Circuits mainly consist of ICs which are very carefully manufactured units and before subjecting the IC to any tests, a careful physical examination of the system is necessary to ensure that there are no dry joints, breaks in the PCB tracks or short circuits between any two tracks on the PCB. Failure of any components connected externally to the IC can also be the cause of the trouble.

While locating faults in digital circuits the following points should be given careful consideration.

(i) An open signal path in the external circuit to an IC produces results similar to that of an open output bond internal to the IC.

(ii) A short circuit between a circuit connection and Vcc or ground is indistinguishable from a short circuit that is internal to the IC.

The simplest method of troubleshooting digital circuits is by sequentially operating the gates and ICs within the system and thus comparing the resulting output with those which are normally present This can be done by applying suitable test signals and then checking the resulting operations by suitable displays. However, there are some special test equipment and service aids which are helpful in the troubleshooting process for digital circuits. A brief description of some of these aids is given as follows.

24.10 DIGITAL TEST EQUIPMENT

24.10.1 Logic Clip

The unit clips on to TTL, or DTL ICs and instantly displays the logic states of all 14 or 16 pins. Each of the clip's 16 light emitting diodes independently follows level changes at its associated pin a lighted diode corresponds to a high logic state. The clip contains its own gating logic for locating the ground and the +5 volt Vcc pins.

A logic clip is more convenient to use than analogue meters in many digital applications. The operation is automatic and there are no adjustments, switch settings or knobs to turn.

The display on a logic clip shows logic high (lamp on), logic low (lamp off) and pulse activity (lamp dim), brightness depends on duty cycle. If the system clock is replaced with a logic pulser, sequential logic devices are slowly stepped through an entire cycle to verify the operational shift registers, counters, flip flops and adders.

24.10.2 Logic Probe

This device is used by the technicians like a screw driver. It simplifies trouble-shooting by providing functional indications of in-circuit logic activity. The lamp indicator allows 360° viewing to clearly and quickly show the state of the circuit under test. The logic probe allows a TTL-C MOS-selectable operation. When switched to TTL position it operates from 4.5 to 15 volt dc power supply.

An important feature of logic probe is its ability to stretch pulses so that short, fast pulses are slowed down and lengthened at the display making them easy for the operator to see. For example 10 ns pulse is stretched to a 100 ms so that the user can see it.

An auxiliary unit to logic probe is the pulse memory unit which is capable of capturing and displaying transient pulses which are hard to see. When the probe tip detects such a pulse, it is stored in the memory and displayed until RESET is pressed.

Some logic probes have two LED display indicator: a green LED for logic 0 and a red LED for logic 1.

24.10.3 Logic Pulser

A logic pulser is used for in circuit testing of digital circuits and when supplemented by a logic probe, it helps in testing for circuit response for easily checking super gates, lines, busses and nodes. The pulser has a tristate output and until operated, its output remains in a high impedance state. This means that it pulses both HIGH and LOW meaning thereby that it has very high impedance when not operating and it can be attached to a node and left there.

Logic pulser 546A from Hewlett Packard provides six different push button selectable output patterns. It means the pulser provides versatile stimulus-response testing capability in both voltage and current applications for virtually any positive voltage logic family.

24.10.4 Logic Current Tracer

This fault locating device primarily locates low impedance faults in digital circuits by sniffing out current sources or sinks. Many troubleshooting problems such as wired-AND/OR configurations result in considerable waste of time in locating faults as several ICs may have to be removed before finding the bad one without any damage to the circuit board. The use of a current tracer helps to pin point the exact faulty point on a mode, even on a multilayer board.

A current tracer depends on its action on sensing the magnetic field by fast-rise-time current pulses in the circuit and displays steps, single pulses and pulse trains using a single one light indicator. The tracer uses a shielded indicative pick up and wide band high gain amplifier to provide the sensitivity needed to sense magnetic fields caused by current changes along PC board traces.

Knowing both current and voltage information helps determine possible fault on a node.

24.10.5 Logic Comparator

A logic comparator is used to check up the performance of a suspected defective IC by comparing the performance of the defective IC with that of a standard good IC. Comparator performs this function by comparing the output response of a reference IC and displaying subsequent errors in performance pin by pin. The IC output pin that does not correctly follow its inputs will produce an error indication even when the error is a short term (200 ns) dynamic fault.

Under most conditions, the logic comparator does not affect the circuit operation due to its loading effect. However, when analog components such as resistors, capacitors or transistors are used to control a timing or to buffer signals, the timing or drive capability of their circuits may be adversely affected.

24.11 SAFETY PRECAUTIONS AND FIRST-AID

Certain electronic equipments like the high power transmitters of All India Radio and Doordarshan operate at very high voltages of the order of thousands of volts (10,00 volts or so) and special safety precautions have to be observed by the maintenance and operational staff to avoid loss of life or time. The faults developing during transmission of programmes have to be located and cleared within the shortest possible time to avoid inconvenience to the listeners and viewers of programs.

Safety precautions to be observed and the first-aid measures to be provided in case of injury or accidents have been detailed in Appendix No— and these must be carefully read by the staff concerned.

24.12 TROUBLESHOOTING CHARTS (FLOW CHARTS)

There are no thumb rules or cut and dry methods for troubleshooting and maintenance of electronic equipment. However, based on a thorough knowledge of the functioning of the particular electronic equipment and a careful observation of the indications provided by various stages certain troubleshooting charts or tables sometimes prove quite helpful in quick location of the faulty stage or component. Some of these troubleshooting charts will now be considered.

24.13 SOME TYPICAL EXAMPLES OF TROUBLESHOOTING

Examples of troubleshooting that will be considered pertain mostly to entertainment electronic equipment like radio receivers (Both AM and FM), Stereo amplifiers, PA systems, VCR and television receivers. All these equipments have already been discussed earlier in this text and only the troubleshooting aspects of these equipments will now be discussed and troubleshooting charts for these equipments will be provided where necessary .

24.13.1 Troubleshooting Radio Receivers

The present day radio receivers are mostly transistor radio receivers which may be of the AM, FM or AM/FM varieties. They may cover only the broadcast band band(s), or they may include one or more short wave bands. An occasional broadcast receiver may also provide a long-wave band. In some receivers a TV band (sound only) is also included. However, there are basic troubleshooting principles that apply to any of the above designs.

Most symptoms in a radio receiver can be classfied under “dead” receiver, weak output, distorted sound, incorrect dial calibration, tuning drift, poor selectivity, noisy output, intermittent operations, motor boating, impaired AVC action or mechanical defect like a broken dial cord or a damaged speaker cone.

Some of the important fault locations and fault rectification methods are discussed as below:

24.13.2 Receiver "Dead" on Only One Band

If an AM/FM receiver is dead on one band say AM band only and operates normally on other bands, then only the AM band should be analysed carefully. Again if a multiband receiver operates normally on broadcast band but is "dead" on one or more of its short wave band, then attention should be focussed on the "dead" band only. If a receiver operates normally on its battery power supply but does not operate properly on its AC power supply, it is clear that the defect is in power supply only.

24.13.3 Weak Output

Weak output can be caused by a fault in any receiver section. In a majority of cases, there is only one fault to be pin pointed. Unless an obvious defect is present, such as broken antenna lead, or a faulty volume control, sectionalization must be made by means of signal tracing or signal injection tests. This procedure will localize the defect to the input circuit, RF stages, local oscillator, mixer, IF section detector, audio section or the power supply. An oscilloscope is the most informative signal tracing instrument. It should have a high sensitivity vertical amplifier and a bandwidth of at least 15 MHz. To avoid undue circuit loading and de-tuning of resonant coils, a low capacitance probe should be utilized.

24.13.4 Distorted Output

Distortion analysis requires the use of an AM signal generator with an oscilloscope as an indicator. Conventional AM generators provide a single audio tone at 400 Hz (or 1 kHz) and unless this tone signal is substantially distorted evaluation of distortion by ear is difficult. On the other hand an oscilloscope will clearly show even a 5% dislocation in the signal.

Overload distribution is likely to occur in the high level output stage of the receiver. Frequency distortion is likely to occur in the front end or IF sections. For example an open by pass or decoupling capacitor in the front end or IF stage can cause premature feedback with the result that timing becomes very critical and signal side bands are 'cut'. In turn, either the low, medium or high audio frequencies are increased abnormally, and the other audio frequencies are attenuated or rejected. This trouble can also be caused by replacing a conventional loop antenna with a high-Q-ferrite rod antenna.

24.13.5 No Output

A no output symptom may or may not be accompanied by noise from the speaker when the volume control is tuned to maximum. If a normal noise is present it is logical to conclude that the signal sections are operating at normal gain, that is, most of the noise voltage is contributed by the input sector of the receivers and this energy is amplified by the IF and audio sections. On the other hand, a negligible proportion of noise voltage is contributed by the output section of the receivers. When a normal noise level is present, the trouble will

not be found in the AVC section or in the power supply. Instead the suspicion falls on the local oscillator. Initial tests are made to determine whether the local oscillator is operating and whether it is operating on correct frequency. In case the oscillator is operating normally and the noise level is normal, suspicion next falls on the RF input circuit. Systematic troubleshooting of a no output system involves the same basic procedure as in the case of a weak output system which has been explained earlier.

If a normal noise level is not present, it must be determined whether the fault lies just before or after the detector stage.

For this, turn the volume control and listen to the AC hum. If power supply hum is heard, the speaker and audio amplifier are working. If this level of AC hum varies with volume control setting, the audio amplifier stage is working and the fault must be in the detector or IF stages. If the diode circuit is normal, the fault most likely to be found in one of the IF stages.

24.13.6 Poor Selectivity

Poor selectivity denotes failure of a receiver to separate stations that are operating on different frequencies, assuming that the stations could be separated if the receiver were working normally. For good selectivity the tuned circuits must operate as a team to pass the desired frequency and to reject other signal frequencies. Poor selectivity can result from improper tracking and can be corrected by realignment of the front end. Another cause for poor selectivity is misalignment of the IF signal channel. All tuned circuits in the IF channel must resonate to 455 kHz. Misaligned stages can be pin pointed by a systematic alignment procedure. In this procedure the radio receiver is aligned by the peak response method. In this method alignment adjustments are made with respect to individual specified frequencies supplied by an AM signal generator.

24.13.7 FM Receiver Troubleshooting

Troubleshooting FM receiver involves the same basic principles as explained for troubleshooting AM radio receiver. There are, however certain technical distinctions to be observed in this case. For example, the signal channel up to audio amplifier section processes a frequency modulated signal. Therefore the signal injection tests require the use of an FM signal generator. An IF channel for an FM receiver operates on 10.7 MHz instead of 455 kHz as in the case of an AM receiver. Also, the IF channel includes both 10.7 MHz and 455 kHz IF transformers as this is a part of a combination AM/FM receiver.

24.14 STEREO AMPLIFIER SYSTEM (FLOW CHART)

A stereo amplifier system consists of two independent channels each having its own microphone, loudspeaker and other balancing components. Common faults that develop in a stereo system are no output, distortion, noise, cross-talk and

unbalanced output. The flow chart given in Fig. 24.4 gives a step by step method of testing and rectifying faults in a stereo system.

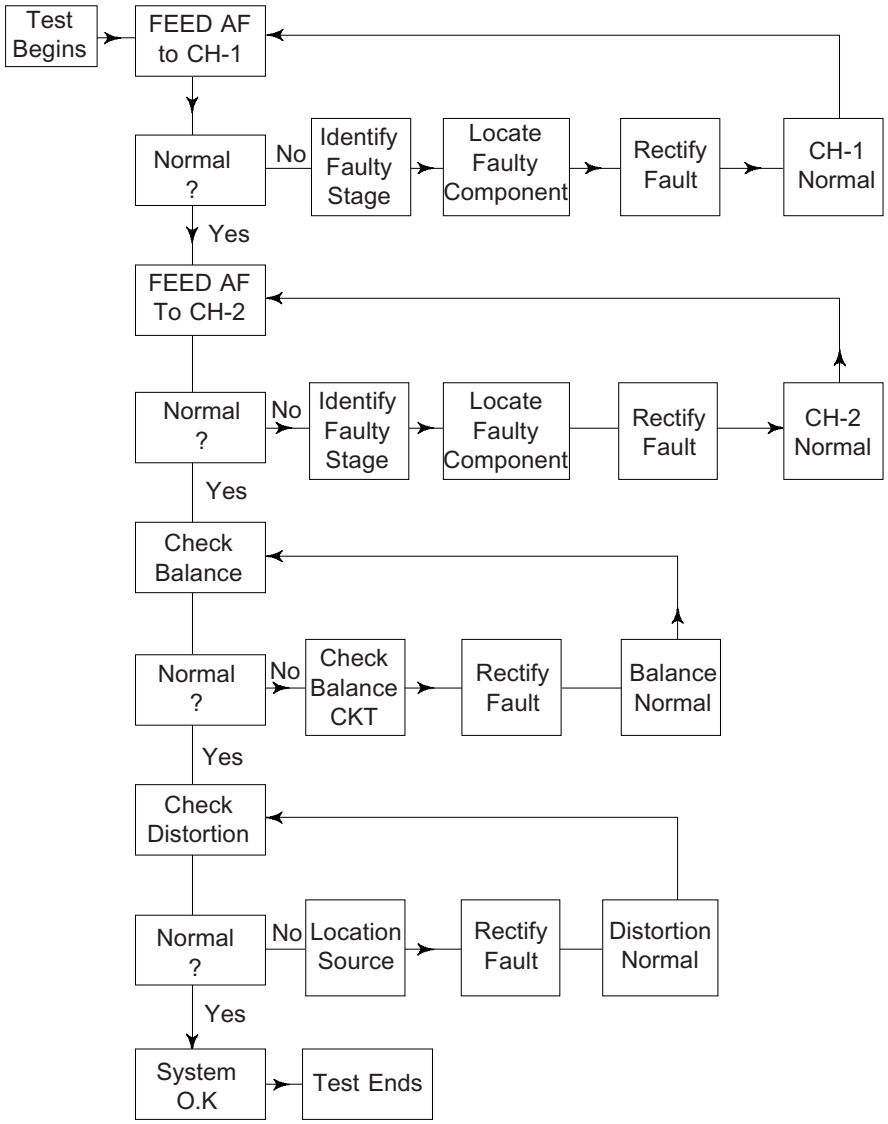


Fig. 24.4 Flow Chart for Troubleshooting in a Stereo System.

24.15 PUBLIC ADDRESS (PA) SYSTEM

As shown in Fig. 24.5, a PA system consists of microphones, a mixer stage, preamplifier, processing circuits, voltage amplifier, power amplifier, loud-speaker/speakers and power supply unit.

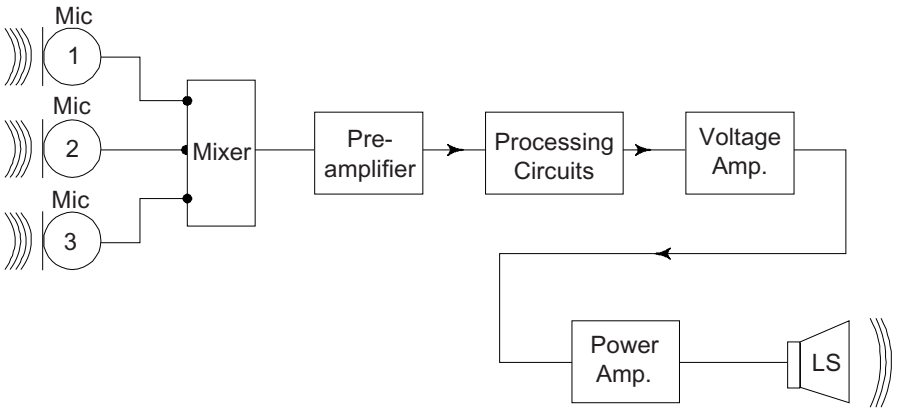


Fig. 24.5 PA System (Block-Diag.)

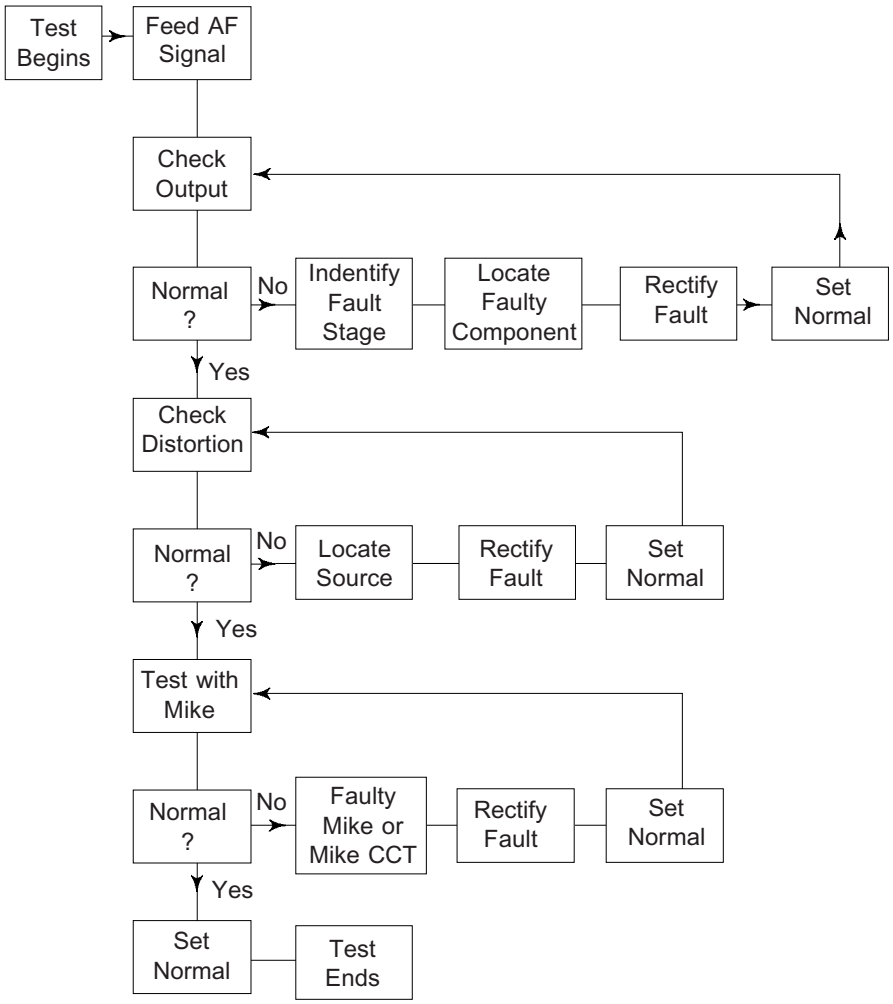


Fig. 24.6 Flow Chart for Troubleshooting in a PA System

Some of the common faults observed in a PA system are, no sound output from the speaker, no hissing sound or noise from speaker, hissing sound present but no sound output, power supply transformer getting over heated, no sound output in one of the speaker, output low, excessive hum and excessive distortion. A flow chart for troubleshooting in a PA system is given in Fig. 24.6 on previous page.

24.16 TROUBLESHOOTING IN VCRs

Video Cassette Recording Machines or VCRs play a very important part in the present day Television broadcasts. Hence, the troubleshooting and fault finding procedure for these machines has been described in detail to enable the technicians to detect the faults quickly and render effective and reliable servicing and repairs.

24.16.1 Localizing Faults in Recording and Playback Systems in VCR Machines

The best way to localize a fault in a VCR is to make use of the standard alignment cassette supplied by most manufacturers. A machine that gives satisfactory reproduction on the standard alignment cassette, but does not satisfactorily playback its own recordings has a defective record processing system. In the absence of a standard alignment cassette, the alternative method of testing a suspected machine is to record a cassette on a good compatible machine and playing it back on the suspected machine to decide whether the fault lies in the recording system or the playback system of the machine. A good theoretical knowledge of the circuitry involved, combined with a careful monitoring of the played back picture, helps localize the fault. After the fault has been localized to a particular section of the machine, corrective measures like alignment of the circuit concerned or replacement/repair of the defective component can be decided upon.

All modern VCRs are colour compatiable and capable of recording or playback of both monochrome and colour signals.

For overall satisfactory results, the chrominance as well as the luminance circuit must perform their prescribed functions correctly. A defect in the luminance processing circuits will adversely affect the performance as much as a defect in the chrominance circuits.

In spite of their seemingly complicated nature, the colour procesing circuit in VCR remain mostly trouble-free. This is due to the fact that chrominance processing is mainly based on the heterodyning process which involves only changes in frequency without any demodulation of the colour signals. Most troubles in colour circuits can, therefore, be ascribed to the misalignment of filter circuits or frequency deviations of the stabilized reference oscillator.

After localizing the fault, the corrective methods will mainly consist of alignment of filter circuits and measuring and adjusting the frequency of the reference oscillator. This will necessitate the use of a good oscilloscope having a minimum bandwidth of 10 MHz and a freqeuncy counter capable of measuring frequencies up to 5 MHz. As stated earlier, a standard alignment cassette with NTSC/PAL colour bars is one of the most useful troubleshooting aids for VCR machines.

Since the signal processing circuits for luminance and chrominance signals have been described separately, the troubleshooting charts (Tables 24.1 and 24.2) for these circuits are also given below separately to facilitate fault finding and rectification of these faults.

Table 24.1 Troubleshooting Chart for Luminance Circuits

Fault symptoms *1*	*Likely defects* *2*	*Remedial measures* *3*
1. Complete loss of picture, with noise.	Accumulation of oxide particle on video heads	Clean video heads as explained in this chapter. Replace video heads if cleaning does not help.
2. Weak or absent video without nose.	Fault lies after the pre-amp stage.	Trace the circuit with an oscilloscope.
3. Partly snowy picture, with 25 Hz flicker.	Failure of one of the two heads.	Complete replacement of the scanning head-assembly, if cleaning does not help.
4. Noisy picture.	(i) Mistracking of heads. (ii) Tape too high or too low at some point of the scanner.	(i) Adjust tracking control. (ii) Mechanical error in the tape path to be checked and corrected.
5. Over-modulation, noise and flicker.	Pre-amp not reproducing high frequency end of FM signal.	One of the preamps not giving proper response. Look for defective components after response tests as per manufacturers instructions.
6. Poor picture quality or lack of sharpness.	(i) Poor response of the pre-amps. (ii) Misadjusted noise canceller.	(i) Check the response with a sweep generator. (ii) Adjust the noise canceller.
7. Beat patterns or herring-bone interference in the luminance signal.	Carrier leak, due to an imbalance in the demodulator.	Check for imbalances with playback of the manufacturer's tape and make necessary balance adjustments.

(*Contd.*)

(*Contd.*)

Fault symptoms 1	*Likely defects* 2	*Remedial measures* 3
8. Normal playback with factory tape, but noisy picture on playback of its own recordings.	Weak record current in the heads.	Measure record current and if necessary, replace open emitter bypass capacitor in the output record amplifier.
9. (i) Weak, low contrast picture. (ii) Over modulation, noise (iii) Carrier leak.	(i) Low deviation. (ii) Deviation too high. (iii) Modulator not balanced.	Adjust deviation to produce the same video level as with the standard alignment tape.
10. Peak white parts of the picture lack detail. Faces appear featureless.	Faulty or misadjusted white-clip stage.	Check waveform at the input of modulator with an oscilloscope or use off-air video and adjust white-clip with trial and error method.

Table 24.2 Troubleshooting Chart for Chrominance Circuits

Fault symptoms	*Likely defects*	*Remedial measures*
1. No colour/weak colour.	CW frequency (4.992 MHz) to the up-convertor missing due to absence of 562.5 kHz or 4.43 MHz to Mixer 1.	Check the frequency of the crystal oscillator and the working of the AFC circuit and make necessary adjustments.
2. Loss of colour lock.	Failure of AFC system.	Check to ensure that horizontal sync pulses and the H-rate pulses from the divider counter reach the AFC system.
3. Shifting of hue between recorded and played back signals.	Shift in the frequency of the 4.43 MHz reference oscillator.	Check and adjust, if necessary the frequency of the reference oscillator with a frequency counter.
4. Colour flicker in playback.	Imbalance of playback chroma signals between video heads.	Check and adjust the balance at the output of preamps.
5. Herring-bone pattern on the coloured part of the picture.	(i) Leakage of 4.992 MHz frequency from the up-convertor into the video output. (ii) Interference from a local broadcast AM station operating near 562 kHz.	(i) Adjust balance control of upconvertor to minimize the output of HF-CW signal. (ii) Provide additional screening for the scanning heads and playback pre-amps.

Technical Staff

Suitably trained and skilful technician and maintenance staff should be employed to get the desired trouble-free service from the electronic equipment in use.

24.17 TROUBLESHOOTING IN TELEVISION RECEIVERS

Maintenance and repair of television receivers has already been discussed in Chapter 19 and detailed troubleshooting charts have been provided there both for the B/W and Colour Television receivers.

24.18 SERVICING, MAINTENANCE AND REPAIR OF ELECTRONIC EQUIPMENT

General

Besides the different methods and techniques described earlier in this chapter there are some other important steps to be taken for long and trouble-free life of an electronic equipment. These steps which are more of an administrative than technical nature are described below:

1. *Maintenance Schedules*
Maintenance schedules which may be of daily, monthly or annual nature should be prepared and these should be meticulously followed.

2. *Log Books*
Log books should be maintained and entries made as and when any maintenance item is completed.

3. Maintenance Manuals and circuit diagrams supplied by the manufacturers should be readily available and should be frequently consulted. Any deviations from the maintenance suggested by the manufacturers should be made only in consultation with the manufacturers. This has already been briefly mentioned in Section 24.2.4 under MTTR.

SUMMARY

Rapid advances in the electronic field have resulted in the introduction of electronic equipments using solid state devices like transistor ICs and microprocessor. Problems arising out of the maintenance and repair of these equipments like causes of failures and reliability factors have been discussed. Items like MTTF, MTBF and MTTR have been defined. Both preventive maintenance and corrective maintenance together with the use of active and passive components have been defined. Trouble shooting techniques like Functional Area approach, Split half method, Divergent paths, Convergent paths and Feedback paths have been discussed. Troubleshooting techniques for Digital circuits and the digital test equipments like Logic Chip, Logic Probe, Logic Current Tracer and Logic Comparator have been described. Typical examples

of troubleshooting techniques with respect to radio receivers, stereo systems, PA systems, VCRs and television receivers with their troubleshooting charts and Flow charts have been included in this chapter.

Review Questions

1. Describe the main causes of failure of electronic equipment.
2. What factors decide the reliability of electronic equipment. Give brief details.
3. Mention the typical faults that develop in Digital circuits.
4. Name the test equipment used for troubleshooting of digital circuits and describe briefly the functioning of these equipments.
5. Describe the troubleshooting processes in radio receivers.
6. Give brief details of troubleshooting procedure in a VCR.
7. Give a troubleshooting chart for the Luminance circuits of a VCR.
8. Give a troubleshooting chart for a Colour Television.
9. What are flow charts? Give a flow chart for stereo amplifier system.

Appendix A

Safety and First Aid

A.1 SAFETY RULES AND PRECAUTIONS

Electronic technicians and mechanics have to work with electricity, electronic devices, motors and other rotating machinery. They are often required to use hand and power tools for constructing models of new devices or setting up new experiments. They use test instruments to measure the electric characteristics of components, devices and electronic systems. All these tasks involve certain risks and hazards if the technicians are not careful in their work and habits. It is essential that students, mechanics and technicians learn the principles of safety at the very start of their careers and that they practice these principles throughout their career.

A summary of the important safety rules and precautions is given below:

1. Plan the job properly, setting out on the work bench in a neat and orderly fashion all tools, equipment and materials. All extraneous items should be removed and cables should be firmly secured.
2. When working on or near rotating machinery, loose clothing should be avoided and ties, etc. tucked away.
3. Line (Power) Voltages should be isolated from ground by means of an isolation transformer.
4. Power Line Voltages are dangerous and can kill so these should not be contacted with hands or any other part of the body.
5. Line cords used must be insulated. If the insulation on any cord is broken or cracked, these cords must be rejected and not used for any measurement.
6. Avoid direct contact with any voltage source.
7. Measure voltages with one hand in your pocket.
8. Wear rubber soled shoes or stand on a rubber mat while working on your experiments.
9. Make sure that your hands are dry and you are not standing on a wet floor when making tests and measurements on a live circuit.

10. Shut off power before connecting test instruments in a live circuit.
11. Make sure line cords of power tools and non-isolated equipment use safety plugs, i.e. proper 3-pin plugs. Do not defeat the safety features of these plugs by using ungrounded adapters.
12. Do not defeat any safety device, such as fuse or circuit breaker by shorting across it or using a higher amperage from than that specified by the manufacturer.
13. Safety devices are intended to protect you and your equipment.
14. Handle tools properly and with care and do not play any and practical jokes in the laboratory.
15. When using power tools, secure your work in a vise or jig.
16. Wear gloves and goggles when necessary.
17. Exercise good judgement and commonsense and make your life safe in the workshop or laboratory.

A.2 FIRST AID

First Aid is the initial assistance or treatment given to a casualty for any injury or sudden illness before the arrival of an ambulance, doctor or any other qualified person.

The first aid methods described in this chapter are those applicable mainly to casualties resulting from electrical accidents.

In the case of an electrical accident the following steps should be taken immediately:

1. Shut off power immediately at the mains if easily accessible or move out the plug or wrench the cable free.
2. Report the accident at once to your instructor.
3. Call for medical help immediately.
4. It may be necessary to render some first aid to the injured before the arrival of proper medical aid.
   The following first aid suggestions are set forth for the guidance of person rendering first aid.
   (i) The injured person should be kept lying down until medical help arrives.
   (ii) He should be kept warm to prevent shock.
   (iii) Do not attempt to give water or other liquids if the injured person is unconscious.
   (iv) Be sure nothing causes further injury. Keep the injured person cheerful until medical help arrives.
5. Artificial Respiration – Resuscitation
   Severe electrical shock may cause stoppage of breathing. The result of an electrical shock on the body is to affect the nervous system and the muscles contract automatically. Those controlling the heart and breathing action are stopped, the patient appears dead.

The electrical resistance of the human body is approximately 30,000 ohms when the skin is perfectly dry but it may be as low as 200 1/N 300 ohms if the skin is wet. In this case a voltage as low as 100 has proved fatal.

In the case of severe electrical shock, the process of artificial respiration or resunsitation must start immediately by a member of the staff *who is trained in artificial respiration*.

The following are the two recommended methods of artificial respiration –

(a) Mouth to Mouth Ventilation

(i) With the casually lying flat on his back, first remove any obvious obstructions including broken dentures from the mouth.

(ii) Open the airway by tilting the head and lifting the chin.

(iii) Close the victims nose by pinching it with your fingers and thumb. Take a full breath and place your lips around his mouth making a good seal.

(iv) Blow into the patient's mouth until you see the chest rise. Take about two seconds for full inflation.

(v) Remove your mouth and allow the chest to fall fully. Deliver subsequent breaths in the same manner.
Repeat till normal breathing is restored.

(b) Schaeffer Method

(i) When an accident occurs, remove the body at once from the circuit by breaking contact with the live conductors. This may be accomplished by using a dry stick of wood, which is a non-conductor, to roll the body over to one side, or brush aside a wire, if that is conveying the current. When a stick is not at hand, any piece of dry clothing may be utilized to protect the hand in seizing the body of the victim unless rubber gloves are available. If the body is in contact with the earth, the coat tails of the victim or any loose or detached piece of clothing should be siezed with care or caution to draw him away from the conductor.

Do not touch the flesh of the victim with your hands till the contact with the live wire is broken.

After this has been accomplished proceed as follows:

(ii) Lay the man on the ground, face downwards, Turn his head on one side. No time should be lost by removing or loosening clothes, begin artifical respiration at once by the following method:
Tell one of the by standers to prepare some sort of a pad like a folded coat and slip it under the patients body just above the waist but do not wait for this. You will have probably performed several movements of respiration before the pad is ready and thus gain all valuable time.

Kneel by the patients side or across his body facing his head. Spread your hands out flat on his back at his lowest ribs, one on each side, the thumbs being close to and parallel with the spine. Press gradually and

slowly for about 3 seconds by leaning forward on to your hands. Use no violence. Relax the pressure by falling back into your original upright. Kneeling position for two seconds without lifting your hands from the patient.

The process of artificial respiration consists in repeating the swaying motion backward and forwards for about 12 to 15 times a minute. The efforts to restore breathing must be carried out with perserverence as in some cases it has been restored after a long period of apparent death. Keep the patient warm.

(c) Some More Precautions

(i) The dashing of cold water into the face will sometime produce a gasp and start breathing, which should then be continued as directed above.

(ii) If this is not successful, the spine should be rubbed vigorously with a piece of ice.

(iii) If the patient has been burned, oil should not be used. Sterlized cotton should be applied. In the case of burns resulting from accidental shock, when the respiratory system has not been affected, a patient may appear perfectly well once his burns have been dressed.

(iv) In the case of an electric shock, it is important that the patient should be kept under watch for at least one day to avoid any unexpected breathing problems.

(d) Other Suggestions

1. It is important that one or two senior members of the technical staff are trained for First Aid Methods through local Red Cross Society, who provide such training by arranging special courses in First Aid.

   Red Cross Society also arranges practical demonstrations in first aid if specially requested.
2. A first aid kit containing required medicines and equipment should be kept ready at a convenient place in the workshop. This kit should be replenished from time to time.
3. Telephone numbers of the nearest doctor/Hospital (Health Centre) and the Ambulance should be prominently displayed in the laboratory or the workshop.

Appendix B

Internet

For the last few years we are living in information age, i.e. people are more aware of their surroundings and world. There is an ever increasing need for sharing information among the people. To bridge the gap people have taken advantage of a wonderful versatile machine "*computers*" by connecting them with each other.

When the computers are connected in individual office through cables it is known as Local Area Networking (LAN) when many such individual networks connect with each other they form a larger network. Internet is essentially the same – ***a worldwide network of networks***. With so many computers connected to each other across the globe, there are lot of things that people can do. For example, access information, display information, communicate with each other, etc.

Files, messages and all kinds of data zoom through these cables at lightning speed. And for this reason the Internet has come to be known as Information ***Superhighway***.

People in all sectors of society have a need to get specific information, and often very quickly. The internet meets this demand, providing people access to a wealth of information in the form of web pages. These digital pages are stored on special host computers and can be viewed with the help of a Web Browser (a program on computer for seeing digital pages on internet, e.g. Internet Explorer or Netscape Navigator). Web pages may have some "links" which when clicked bring you to another related page that might be located on same computer or even on a computer in another country. The sum of all these web pages is known as the ***World Wide Web***.

In a sense, the World Wide Web is like a massive, global, digital information resource. By simply clicking your mouse, you can browse through different websites (*collections of web pages*) and explore information on every subject imaginable.

There are literally millions of websites and web pages that are freely available to anyone with a computer and Internet connection.

Internet connection can be taken by buying a CD from any reputed ISP (Internet Service Provider) for a stipulated period of time. ISP like MTNL have started free Internet connection facility if you have a telephone connection with you.

For the connection one needs a computer, a MODEM and a telephone line. As the data has to pass through telephone lines hence to convert the digital data of computer to the telephonic signals and vice versa we need MODEM.

Internet through Cable TV network is also becoming popular. Now we can connect our television to internet by buying a kit for the purpose which includes a keyboard, a mouse and a converting device like MODEM.

Internet uses TCP/IP based communication architecture to interconnect different LANs. The TCP/IP architecture consists of five core protocols (set of rules). This family of protocols is known as "*Internet protocol suite*". TCP and IP are two most common protocols of this suite.

IP (Internet Protocol)

This is a protocol used by the Internet for transferring messages from one machine to another. The messages are sent in the form of packets of digital codes. This protocol identifies each network and each computer on network by a unique fixed address. This address is used by IP to determine the source and destination of a packet.

TCP (Transmission Control Protocol)

The TCP provides the logic for ensuring the reliable delivery of data to be exchanged between the host systems. It collects the elated packets of a message or of blocks of data and places them in proper order. It also checks the validity of the packets. Thus it keeps the tracks of the blocks of data to ensure reliable delivery to the appropriate application.

B.1 APPLICATIONS OF INTERNET

E-mail

It has revolutionized the way people communicate. Within a matter of minutes information can be transferred from one corner of the world to another through electronic mail. We can not only send text messages but also graphics and files as attachments.

Chat

Another method of communication through internet is Chat. Chat is live discussion through typing. This enables people anywhere in the world to talk online through text.

Message Boards

Message Boards are special web pages that allow people to display their written messages publicly. People can write messages and even respond to a message, where their reply is also made available for others to read. Message

Boards are different from E-mail. E-mail is a private method for communication but messages boards. Are like a conversation that many people watch or listen to.

Search Engines

Search Engines facilitate locating websites through word searches. Examples of search engines are Infoseek (http://www.infoseek.com)
Exite (http.//www.excite.com) and
Alta Vista (http://www.altavista.digital.com)

B.2 SOME COMMON TERMS

URL

Uniform Resource Loader (address of a Website e.g.
http://www.Encyclopaedia.com)

Protocols

Set of rules defined for transferring data from one computer to another across the world. For example, SMTP (Simple Mail Transfer Protocol) for E-mail (Electronic Mail).

Bandwidth

The capacity of a medium to transmit a signal. On internet more the bandwidth the tranfer of data is more fast.

Surfing

Using the Internet Browser program to access various websites.

Gateway

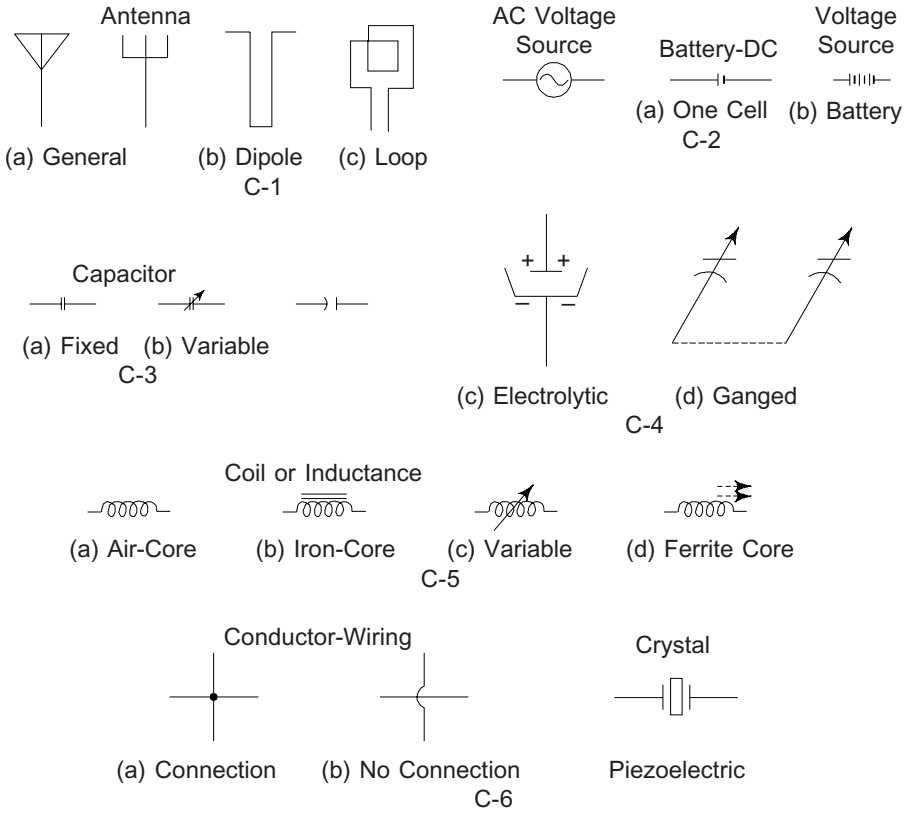
Device that connects dissimiliar networks.

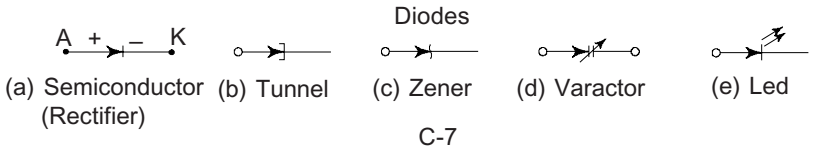
Node

A computer attached to network.

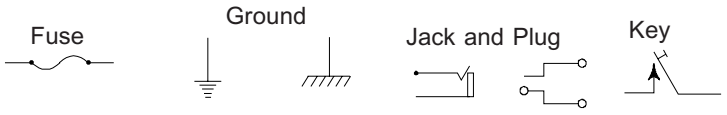
Appendix C

Common Schematic Symbols

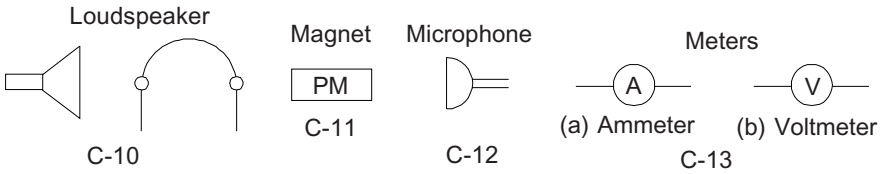




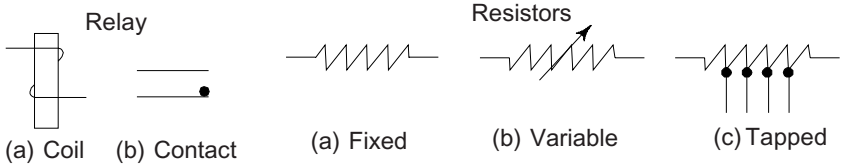
C-7



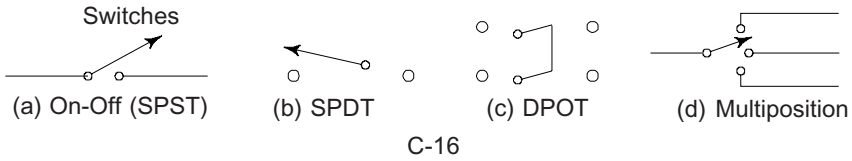
C-8 C-9



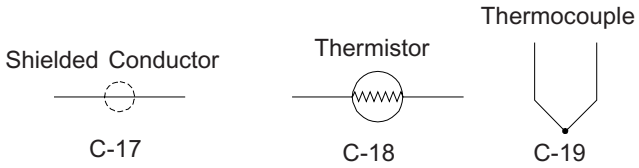
C-10 C-11 C-12 C-13



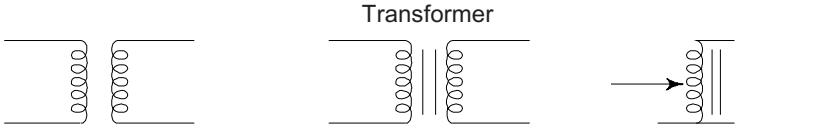
C-14 C-15



C-16

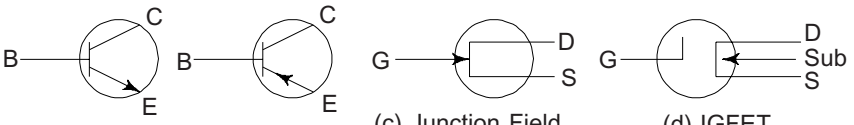


C-17 C-18 C-19



C-20

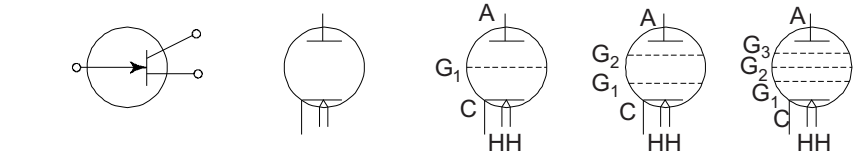
Transistors



(a) NPN (b) PNP (c) Junction Field Effect Transistor (JFET), N Channel (d) IGFET N Channel

C-21

Vacuum tubes



(e) Silicon Controlled Rectifier, SCR (a) Diode (Indirectly Heated) (b) Triode (c) Tetrode (d) Pentode

C-22

Appendix D

Soldering

D.1 WHAT IS SOLDERING?

Soldering is the process of joining together two metallic conductors at a relatively low temperature of 500 to 600 °F. The joint where the two metal conductors are to be joined or fused is heated with a device called a *soldering iron* and then an alloy of tin and lead called *solder* is applied which melts and covers the joint. The solder cools and solidifies quickly to ensure a good and durable connection between the joined metals. Covering the joint with solder also prevents oxidation.

D.2 TOOLS AND MATERIALS REQUIRED FOR SOLDERING

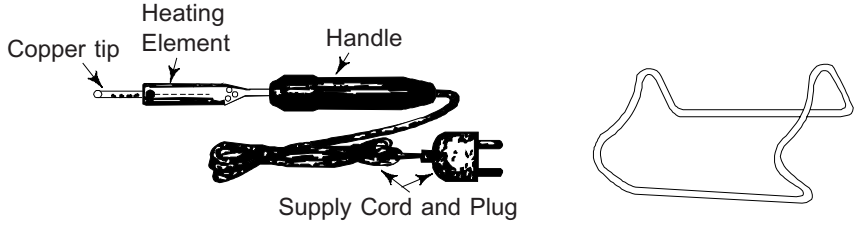
Soldering Iron

Figure D.1(a) shows the details of an electric soldering iron used in radio and TV work. It consists of:

(i) Supply cord and plug, (ii) a wooden or plastic handle,

(ii) Metallic shank containing the heating element, and (iv) a copper tip.

All these parts are replaceable. Electric soldering irons are normally mains operated, although battery operated soldering irons are also available. Solder-



(a) Soldering Iron (b) Soldering Iron Support

Fig. D.1

ing irons are rated in watts. For electrical and electronic work the rating of a soldering iron varies from 25 to 100 W. A 25 to 35 W soldering iron is commonly used for radio and TV work. However, soldering irons rated at 250 W or more are available for heavy types of soldering work.

To avoid damage to work tables and other surfaces, a heated soldering iron should be placed on its holder or stand of the type shown in Fig. D.1(b).

The other tools required for soldering work are shown in Fig. D.2. These tools are required in addition to normal workshop tools.

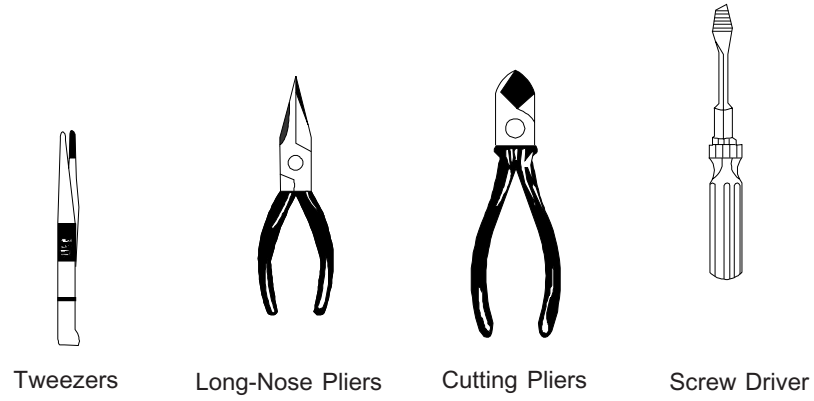


Fig. D.2

Solder

Solder is an alloy of tin and lead whose melting point depends on the ratio of tin and lead in the alloy. The three grades of solder available contain the percentage of tin to lead respectively as 40–60, 50–50 and 60–40. The 60–40 (60 tin-40 lead) solder melts easily and makes soldering easy and safe without any damage to the surface or components being soldered. This type of solder is most suitable for electronic work. Solder is usually available in the form of wires of different sizes.

Solder Flux

A rosin solder flux or paste is generally used for soldering. The flux easily melts and cleans the surface of any oxide films to ensure a smooth and clean soldered joint. However, solder wires with flux filled in the hollow core of the wire are also available. Use of flux-core solder wires makes the external application of flux unnecessary. Acid flux should *not* be used particularly when soldering copper wires.

D.3 HOW TO SOLDER

Good soldering practice is very important for assembling any electronic circuit. A poorly soldered joint or connection in electronic circuits is the cause of most

service problems. Given below are some important steps to be followed in good and correct soldering practice:

1. Use the correct type of soldering iron and solder. A 25-30 W soldering iron with a pencil tip and a 60–40 solder wire with a flux-core are most suitable for radio and TV work. Avoid the use of excessive flux.
2. The tip of a new soldering iron should be cleaned with a file and tinned (covered with solder) before use. Keep the soldering iron hot during the working period and let it rest on its stand when not in use.
3. All component leads and wires should be thoroughly cleaned to remove dust and rust before soldering. Application of flux is not a substitute for cleaning.
4. Apply enough heat to the joint so that the solder metal flows freely over the joint. Remove first the solder wire and then the soldering iron and do not move the joint till the solder has set properly.
   Remember, a well soldered joint will look smooth, bright and shining.
5. Avoid overheating of components or printed circuit boards (PCB); overheating may result in damage to components or PCBs. It is advisable to hold the component lead with a long-nose pliers while soldering to dissipate heat.
6. Do not use too much solder to avoid short circuits between conducting paths on a PCB.
7. Remember larger metal surfaces like legs of selector switches and legs of IFTs take a longer time to heat.
8. Keep the soldering iron tip clean by wiping it from time to time with a damp sponge or cleaning with a file.

For details of soldering on PCBs see Appendix E.

Appendix E

Printed Circuit Boards (PCB)

E.1 PRINTED CIRCUIT BOARDS (PCB)

A printed circuit board popularly known as PCB is a piece of plastic insulating board, on one side of which a complete layout diagram of an electronic circuit consisting of copper or silver conducting paths is printed by a special photo engraving process. On the other side of the PCB are mounted electronic components like resistors, capacitors, coils, transformers, tubes, transistors, diodes and ICs. Suitable holes are punched in the PCB for mounting the components which are connected to the conducting paths by soldering. The plastic board is made of a translucent material and with a bright light on one side, the circuit on the opposite side can be easily traced.

Printed circuit boards are extensively used for assembling electronic circuits particularly those using semiconductor devices like transistors and ICs. Use of PCBs makes the circuit assembly more compact, uniform and stable than the one using hand wiring. PCB assembly also lends itself efficiently to mass production methods.

Figure E.1 (a) shows the component side of a PCB on which the circuit of a 3-band transistor radio of Fig. E.1(b) has been assembled.

Assembling PCB circuits, their maintenance and repair requires special techniques for which some useful hints are given below.

E.2 COMPONENT MOUNTING

1. Before mounting any components, examine the PCB carefully for any cracks, breaks or other defects in the conducting paths.
2. The leads of components like resistors, and capacitors should be fully inserted into the mounting holes taking care to mount the components so that any information written on the components is clearly visible.

CAPACITORS (Values printed or colour coded)

Symbol

C – 1A and C-1B 3365pF PVC gang
C – 2 15pF Ceramic
C – 3 33pF Ceramic
C – 4 33pF Ceramic
C – 5 2-22pF Trimmer
C – 6 2-22pF Trimmer
C – 7 2-22pF Trimmer
C – 8 0.047 μF Polyester
C – 9 3kpF Pin-up
C – 10 3K or 3000pF Styroflex
C – 11 4K7 Pin-Up
C – 12 2-22pF Trimmer
C – 13 2-22pF Trimmer
C – 14 2-22pF Trimmer
C – 15 15pF Ceramic
C – 16 47pF Ceramic
C – 17 33pF Ceramic
C – 18 330pF Ceramic
C – 19 3KpF or 3000pF Styroflex
C – 20 3KpF or 3000pF Styroflex
C – 21 2KpF or 2700pF Styroflex
C – 22 10μFD/6V or 16V Electrolytic
C – 23 0.047 μF Polyester
C – 24 0.047 μF Polyester
C – 25 0.047 μF Polyster
C – 26 10KpF Pin-Up or 0.01 μFD
C – 27 10KpF Pin-Up or 0.01 μFD
C – 28 C-28 200μFD/6V or 220μFD/10 V Electrolytic
C – 29 1μFD/6V or 80V Electrolytic
C – 30 15 kpF (0.015μF) Polyester or 10kpF Pin-Up
C – 31 200μFD/6V or 220μFD/10V Electrolytic
C – 32 220μFD/6V or 220μF/10kV Electrolytic
C – 33 5μFD/6V or 64V Electrolytic

TRANSISTORS AND DIODE

TR – 1 BF 194B
TR – 2 BF 195C
TR – 3 BF 195D
TR – 4 BC 148F
TR – 5 AC 128 } Matched Pair
TR – 6 AC } with Heat Sink
D – 1 OA 79

COILS AND IFTs

L – 1 MW Antenna Coil (on paper former with 3 leads)
L – 2 SW1 Antenna Coil (on paper former Tinned wire)
L – 3 SW2 Antenna Coil (Square black former with white dot and with blue core)
L – 4 MW Osc. Coil-Red
L – 5 SW1 Osc. Coil-Pink
L – 6 SW2 Osc. Coil-Green
L – 7 1st IFT- Yellow
L – 8 2nd IFT – White
L – 9 3rd IFT –Black
T – 1 6V Driver Transformer

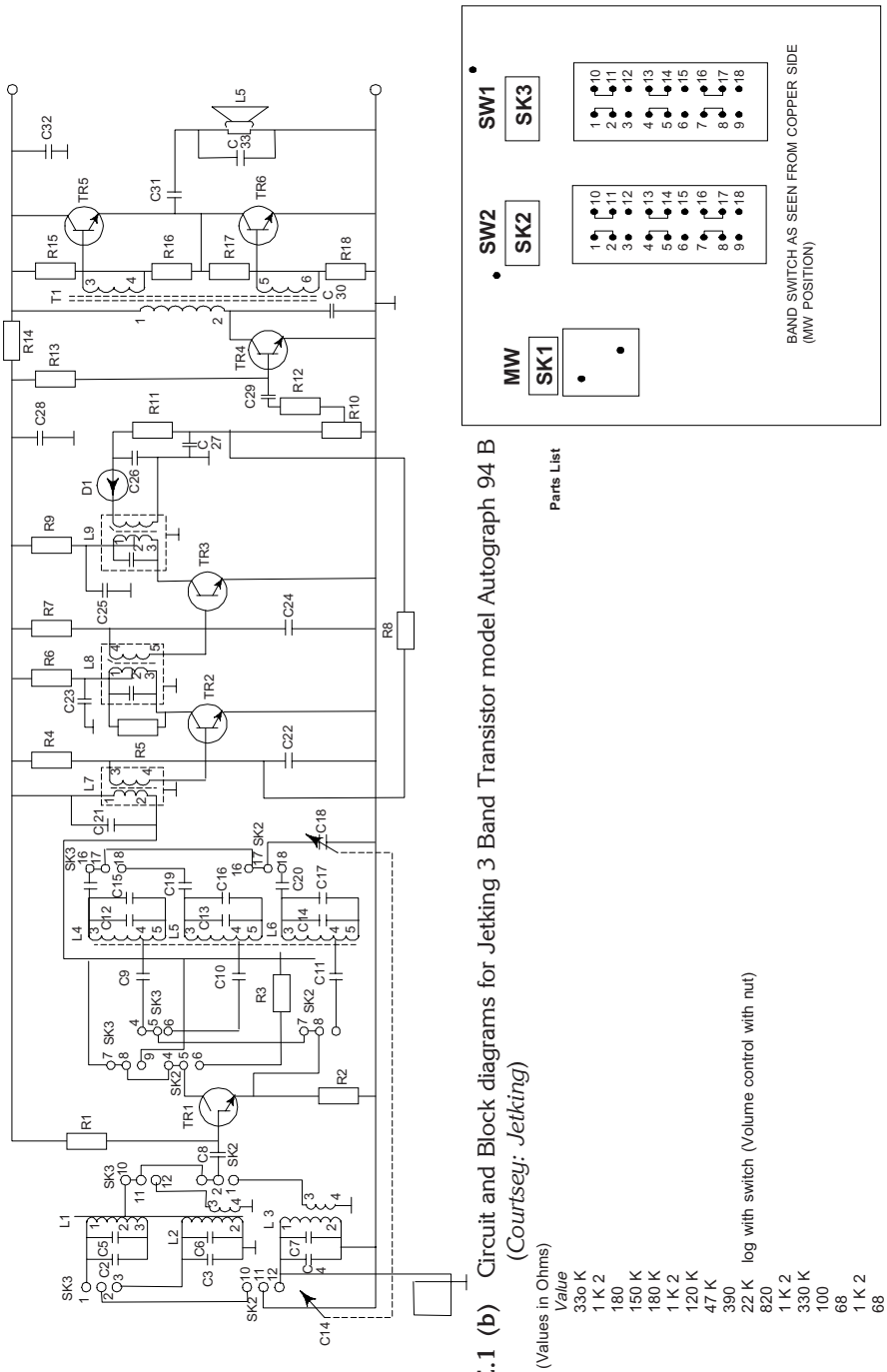


Fig. E.1 (b) Circuit and Block diagrams for Jetking 3 Band Transistor model Autograph 94 B
(*Courtsey: Jetking*)

Parts List

Resistors (Values in Ohms)

Symbol	*Value*	
R – 1	33o K	
R – 2	1 K 2	
R – 3	180	
R – 4	150 K	
R – 5	180 K	
R – 6	1 K 2	
R – 7	120 K	
R – 8	47 K	
R – 9	390	
R – 10	22 K	log with switch (Volume control with nut)
R – 11	820	
R – 12	1 K 2	
R – 13	330 K	
R – 14	100	
R – 15	68	
R – 16	1 K 2	
R – 17	68	
R – 18	1 K 2	

3. Carefully cut the leads of the components so that about 3 mm of the end extends beyond the wiring side of the PCB. The ends of the leads are then bent at right angles to make a firm contact with the surface to which it is to be soldered.
4. In the case of semiconductor devices like transistors and diodes the length of the leads extending above the components side of the PCB should be about 1 cm. This will not only enable a heat sink to be applied to each lead while soldering but will also be useful for measuring voltages across these leads.
5. Certain components like transformers, potentiometers and variable capacitors which are meant for use with PCBs are provided with pin-type terminals which can be simply inserted into the holes in the PCB and soldered.

E.3 SOLDERING

PCB soldering requires a proper soldering technique as explained below:

1. A light duty soldering iron of 25 to 30 W rating should be used to prevent damage to the printed circuit wiring by excessive heating.
2. Do not use excess solder to avoid solder flowing to adjacent conducting paths to form bridges which cause short circuits. Thin solder wires of 60/40 composition should be used.
3. For soldering of ICs the soldering iron should have a thin tip because of the close spacing of the pins. Number 12 SWG wire wrapped round the iron tip can also be used for soldering.
4. When soldering a semiconductor device, hold the lead to be soldered with a pair of long-nose pliers or use a crocodile clip to serve as a heat sink to prevent damage to the component.

E.4 REPAIRING

Figure E.2(a) indicates the correct method of soldering on a PCB whereas Fig. E.2(b) shows poor soldering practice. Repair jobs on PCBs include replacement of defective components, refixing of loose joints and repairing of cracks and breaks in printed conductors.

1. Defective components like fixed resistors and capacitors can be easily replaced by cutting the leads near the body of the component with a pair of cutting pliers and soldering the new component to the old leads. This method is particularly useful when the wiring side of the PCB is not accessible. If, however, the wiring side of the PCB is accessible, the leads of the defective components can be slowly pulled out of the mounting holes by desoldering. The solder in the holes is then melted and the leads of the replacement component are inserted into the holes and soldered in place.

2. Loose solder joints in a PCB can be refixed by carefully reheating the joint without using additional solder.
3. A thin crack in the printed circuit can be easily repaired by filling in with solder or by soldering a piece of bare copper wire over the crack. However, when a larger section of the printed wiring is damaged, the gap can be bridged with a length of hookup wire soldered at two convenient terminal points.

E.5 MODULES

Modules are small PCBs which contain the components of a specific unit or the components of a combination of circuits. A complicated circuit like that of a TV receiver may consist of a number of modules which have pin connections that plug with the main chassis or are equipped with plated terminals that can be inserted into clip-type sockets. Modules can be easily removed from the circuit for inspection and testing purposes. Any defective module can be conveniently repaired or replaced without disturbing the rest of the circuit, the only disadvantage being that oxidation of the contacts may sometimes cause intermittent connections.

Appendix F

Vaccum Tubes

F.1 THERMIONIC EMISSION

A conductor contains mobile electrons which are continually moving about from atom to atom inside the conductor. If the conductor is heated, the velocity of the electrons increases. In certain substances, when the temperature is raised sufficiently, the electrons are agitated to such an extent that they leave the surface of the material. This "evaporation" of electrons from the surface of a material is similar to the turning of water into water vapour when heated. The phenomenon of ejection or emission of electrons by heat is called *thermionic emission* or electron emission.

The conductor that emits electrons is called the *cathode*. If another conductor or a metallic plate is placed near the cathode and connected to the positive terminal of a battery as in Fig. F.1, the negatively charged electrons emitted by the cathode are attracted by the positively charged metal plate and an electric current is established between the cathode and the second metal element maintained at positive potential with respect to the cathode. The element or electrode which collects electrons given out by the cathode is called the plate or anode. Thus, the cathode is the emitter of electrons and the plate or anode is the collector of electrons.

The process of emission and collection of electrons mentioned above is best performed in a vaccum. Accordingly, the cathode and plate are enclosed in a glass, metal or ceramic enclosure from which air has been pumped out to create a vacuum. The most convenient way of heating the cathode is by passing an electric current through it from a battery as shown in Fig. F.1.

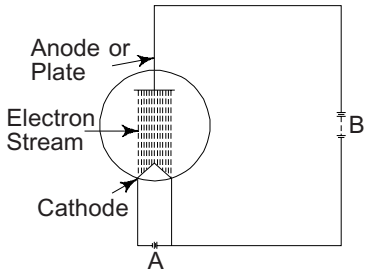


Fig. F.1 Vaccum diode

This arrangement of the cathode and plate enclosed in an evacuated glass, metal or ceramic enclosure constitutes the simplest type of *vacuum tube* called the *diode* and has two electrodes. Vacuum tubes with three, four and five electrodes have also been constructed and are called *triode, tetrode* and *pentode*, respectively. Since the electrons will always flow form the cathode to the plate and not in the reverse direction, the diode forms a unidirectional device and was originally called a *valve.*

A vaccum between the cathode and anode is necessary for two reasons:

(i) At the very high temperatures to which the cathode is heated, it gets oxidised if air is present and is burnt out. (ii) The electrons moving towards the anode acquire sufficient velocity due to the positive potential on the anode and knock off electrons from atoms of the gas and produce positive ions. These heavy positive ions are attracted by the negative potential on the cathode and bombard the cathode with such a tremendous force that the cathode completely disintegrates under this bombardment.

F.2 OTHER TYPES OF ELECTRONIC EMISSIONS

Besides the thermionic emissions known as the primary emissions, there are other types of electronic emissions. These are:

Secondary Emission

Electrons are also sometimes driven off the surface of a metal when bombarded by high speed primary electrons. This type of emission is called *secondary emission.* In vacuum tubes secondary emission is an undesirable phenomenon and creates design problems.

Photo-electric Emission

Electrons can also be driven off the surface of certain metals when light rays strike these metals. This type of emission is called *photo-electric emission*. Photo-electric tubes based on photo-electric emission are not used in radio and TV work but they are extensively used for sound reproduction in the motion picture industry.

Field Emission

Electrons can be pulled out of the cold emitter surface by the application of a strong electric field outside the emitter surface. The stronger the electric field, the greater is the field emission from the emitter surface.

Thermionic emission is the type of emission that is most commonly used in vacuum tubes.

F.3 CATHODE

A cathode is an emitter of electrons. It emits electrons when heated electrically. There are two methods of heating a cathode. When the current passes directly through the cathode wire called the *filament*, the cathode is said to be directly heated (Fig. F.2 (a)). In the indirectly heated cathode, the electric current is

applied to a separate heating element located inside a cylindrical sleeve that is coated with the emitting material (Fig. F.2 (b)).

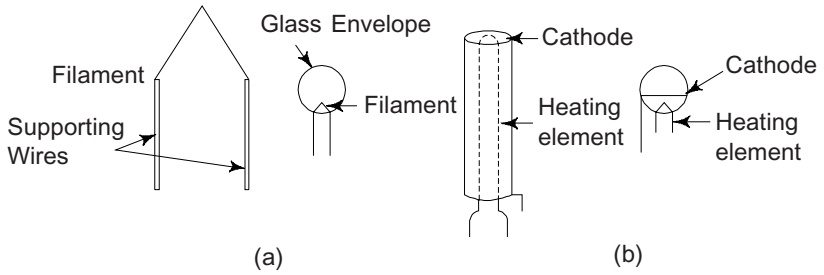


Fig. F.2 Types of Cathodes (a) Directly Heated Filament and its Symbol (b) Indirectly Heated Cathode and its Symbol

The materials most commonly used as emitters in vacuum tubes are tungsten, thoriated-tungsten and oxides of barium and strontium. Pure tungsten metal can be heated to very high temperatures and is used for the filaments of large transmitting tubes. Thoriated tungsten, made by adding thorium and carbon to tungsten, can emit electrons at comparatively lower temperatures and is used in small transmitting tubes. Oxide-coated cathodes can emit electrons at much lower temperatures and all small receiving type vacuum tubes employ oxide- coated filaments.

F.4 CLASSIFICATION OF VACUUM TUBES

Vacuum tubes are classified according to the number of electrodes used in the construction of the tube. A vaccum tube with only two electrodes – a cathode and an anode – is called a *diode.* When another electrode called the grid is introduced between the cathode and the plate, the three element tube becomes a *triode*. Similarly, a *tetrode* has four electrodes including two grids and a *pentode* is a five element tube with three grids interposed between the cathode and the anode. Figure F.3, shows the circuit symbols of various types of tubes

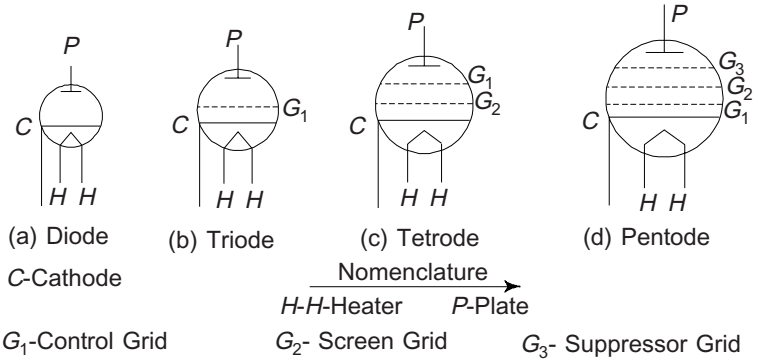


Fig. F.3 Circuit Symbols of Vacuum Tubes (a) diode (b) triode (c) tetrode (d) pentode

mentioned above. These are the main types of tubes. In all these tubes, the cathode is the emitter of electrons and the anode or the plate is the collector of electrons. The grids modify the flow of electrons from the cathode to the anode thereby enabling these tubes to function as rectifiers, amplifiers and oscillators.

Index