Basic Electronics

Second Edition

About the Authors

D P Kothari obtained his BE (Electrical) in 1967, ME (Power Systems) in 1969 and PhD in 1975 from BITS, Pilani, Rajasthan. He was involved in teaching and development of several courses at BITS Pilani, from 1969 to 1977. Currently, he is the Dean (R&D) at J D College of Engineering & Management, Nagpur. Earlier, he served as Vice Chancellor at VIT, Vellore. He was also the Director in-charge and Deputy Director (Administration) as well as Head at the Centre of Energy Studies at Indian Institute of Technology, Delhi. He also served as Principal at VRCE, Nagpur. During 1982–83 and 1989, he was a visiting professor at the Royal Melbourne Institute of Technology, Melbourne, Australia. He taught at Melbourne University for one semester in 1989. He was also NSF Fellow at Perdue University, USA, in 1992.

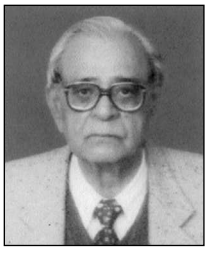
Dr Kothari, who is a recipient of the most Active Researcher Award, has published and presented 812 research papers in various national and international journals, conferences, guided 50 PhD scholars and 68 M.Tech. students, and authored 50 books in various allied areas. He has delivered several keynote addresses and invited lectures in both national and international conferences. He has also delivered 42 video lectures on YouTube with maximum of 40,000 hits!

Dr Kothari is a Fellow of the National Academy of Engineering (FNAE), Fellow of Indian National Academy of Science (FNASc), Fellow of Institution of Engineers (FIE), Fellow IEEE, Hon. Fellow ISTE and Fellow IETE.

His many awards include the National Khosla Award for Lifetime Achievements in Engineering from IIT, Roorkee, in 2005. The University Grants Commission (UGC), Government of India, has bestowed the UGC National Swami Pranavandana Saraswati Award for his outstanding scholarly contributions in the field of education, in 2005.

He is also a recipient of the Lifetime Achievement Award conferred by the World Management Congress, New Delhi, for his contribution in the areas of educational planning and administration, in 2009. Recently, he received Excellent Academic Award at IIT Guwahati by NPSC, 2014.

In 2016, he received six Life Time Achievement awards by various agencies and Living Legend Award at Chennai Conference.



I J Nagrath was an Adjunct Professor at BITS Pilani, from where he retired in July 1999 as Professor of Electrical Engineering and Deputy Director. He was actively engaged in writing books related to his long teaching and research experience.

Prof. Nagrath obtained his BE with Honours in Electrical Engineering from Birla Engineering College in 1951, and MS from the University of Wisconsin, USA, in 1956.

He had co-authored several successful books, which include *Electric Machines*, 4/e, *Modern Power System Analysis*, 4/e, *Power System Engineering*, 2/e, *Electrical Machines*, 2/e (Sigma Series) and *Basic Electrical Engineering*, 3/e (all published by MHEI); and also co-authored *Control System Engineering* and authored *Electronics: Analog and Digital*. He had also published several research papers in prestigious national and international journals.

Basic Electronics

Second Edition

D P Kothari
Dean (R&D), J D College of Engineering & Management, Nagpur
Fellow-INAE, Fellow-INASc, Fellow-IEEE, Hon. Fellow-ISTE, FIE (India), Fellow-IETE, MCSI
Former Director Research, MVSR Engineering College, Hyderabad
Former Director General, J B Group of Educational Institutions, Hyderabad
Director General, Raisoni Group of Institutions, Nagpur
Former Director General, Vindhya Institute of Technology and Science, Indore
Former Vice Chancellor, VIT University, Vellore
Former Director-Incharge, IIT Delhi
Former Principal, VRCE, Nagpur
Visiting Fellow, RMIT, Melbourne, Australia

I J Nagrath
Adjunct Professor and Former Deputy Director
Birla Institute of Technology and Science
Pilani

McGraw Hill Education (India) Private Limited
CHENNAI

McGraw Hill Education Offices

Chennai 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

McGraw Hill Education (India) Private Limited
Published by McGraw Hill Education (India) Private Limited
444/1, Sri Ekambara Naicker Industrial Estate, Alapakkam, Porur, Chennai 600 116

Basic Electronics, 2/e

1 2 3 4 5 6 7 8 9 D103074 22 21 20 19 18

Printed and bound in India.

Print Edition
ISBN (13): 978-93-5260-646-7
ISBN (10): 93-5260-646-9

e-Book Edition
ISBN (13): 978-93-5260-647-4
ISBN (10): 93-5260-647-7

Managing Director: *Kaushik Bellani*

Director—Science & Engineering Portfolio: *Vibha Mahajan*
Senior Manager Portfolio—Science & Engineering: *Hemant Jha*
Portfolio Manager—Science & Engineering: *Navneet Kumar*

Senior Manager—Content Development: *Shalini Jha*

Production Head: *Satinder S Baveja*
Copy Editor: *Taranpreet Kaur*
Assistant Manager—Production: *Anuj K Shriwastava*

General Manager—Production: *Rajender P Ghansela*
Manager—Production: *Reji Kumar*

Typeset at APS Compugraphics, 4G, PKT 2, Mayur Vihar Phase-III, Delhi 96, and printed at

Cover Printer:

Visit us at: www.mheducation.co.in

Contents

Preface

The updated second edition is the outcome of excellent response to the first edition of *Basic Electronics* released in 2014. For an engineer in the area of Electronics and Communication it is very important to have strong foundation on fundamental concepts, techniques and devices in Electronics Engineering. Therefore, our aim with the book is to make that learning easy for students. In attempting to do so, the book covers the basic concepts of Electronics, Communication Engineering and Computers in a single volume.

Since the appearance of the first edition in 2014, latest advances in the field of electronics have been included in this new edition. Certain portions have been rewritten for more clarity and better understanding. The book offers:

- Full coverage of course on Basic Electronics including topics like Basic Computers and Communication Engineering.
- Chapters are enriched with excellent pedagogy :
  - Highly illustrative with over 415 diagrams
  - 140+ Solved Examples including questions from previous year university question papers
  - 240+ Review Questions to test the conceptual understanding
  - 165+ Numerical Problems to practice and prepare well for exams
  - 200+ Multiple-Choice Questions

Web Supplements

This book comes with excellent PowerPoint lecture slides as supplements which can be accessed at: *http://www.mhhe.com/kothari/be2*

Acknowledgments

I wish to place on record my gratitude to Prof (Dr) Priti Singh, Shruti Karkra, Dr Karamjit Kaur and Anil Kumar, for their contributions in the making of this updated edition.

I am extremely thankful to the team members at McGraw Hill Education (India), for their interest, cooperation and help in making the second edition a reality within record time.

D P KOTHARI
I J NAGRATH

Publisher's Note

McGraw-Hill Education (India) invites suggestions and comments from you, all of which can be sent to info.india@mheducation.com (kindly mention the title and author name in the subject line). Piracy-related issues may also be reported.

CHAPTER 1

Semiconductors

GOALS AND OBJECTIVES

- Introduction to semiconductors and insulators
- Conduction in solids and metals
- Explanation of doped semiconductors
- Definition of diffusion

1.1 CONDUCTION IN SOLIDS

Solid materials, from the point of view of current carrying capability are of three types—*conductors*, *insulators* and *semiconductors*. Conductors have an abundance of free electrons, which act as the charge carriers, which means that they have high conductivity. Insulators, on the other hand, have hardly any free electrons and therefore practically cannot conduct any current, i.e. these are poor conductors. Semiconductors, as such, have conductivity intermediate between conductors and insulators and have the special feature of having two types of charge carriers—electrons and holes (which are absence of electrons in covalent bonds and act equivalently as positive charge carriers). Further, semiconductors have the peculiar property such that by doping with certain elements, their conductivity gets increased by several orders of magnitude (up to about six). It is this property, in fact, which is exploited in constructing electronic devices and integrated circuits from semiconductor materials.

1.1.1 Conduction in Metals

Typically, in metals (copper, aluminium, silver, etc.), atoms are arranged in systematic array in the form of crystals. The electrons of the outer orbit of metal atoms being loosely bound to the nuclei get detached and become free because of attractions from neighbouring nuclei. The metal crystal thus exists in the form of a lattice of positive ions with fixed (but vibrating) locations surrounded by randomly moving free electrons such that any macrosize metal piece acts electrically neutral as shown in Fig. 1.1. Electrons move about rapidly and randomly (in Brownian motion) losing kinetic energy upon elastic collision with positive ions and gaining energy (accelerating) between collisions because of attractive forces of adjoining ions with net zero movement as shown in Fig.1.2(a).

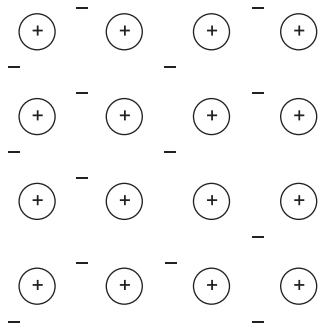
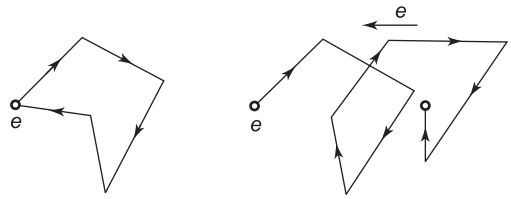


Fig. 1.1 Two-dimensional representation of a metal crystal



(a) Random (b) Drift

Fig. 1.2 Electron motion in a metal crystal

Upon application of uniform electric field ε (V/m), electrons while still moving rapidly and randomly, acquire a net (average) component of velocity in opposite direction to that of the field as shown in Fig. 1.2(b). This directed average velocity is known as *drift velocity* u_d and is proportional to ε.

Thus,

$$u_d = \mu(-\varepsilon) \text{ m/s; electrons drift in direction opposite to the electric field} \quad (1.1)$$

where μ = *mobility* (m/s per V/m or m^2/V.s) which is a property of the material and decreases with temperature (with faster vibration of ions in the lattice, chances for collision of electrons with these increase).

The drift of electrons carrying charge $-q$ constitutes electric current. If there are n free electrons/m^3, the current density J(A/m^2) is

$$J = n(-q)u_d = n \times (-q) \times (\mu - \varepsilon) = \sigma\varepsilon \tag{1.2}$$

where $\sigma = nq\mu$ = *conductivity* of the material in siemens/m. The reciprocal of conductivity is called *resistivity* in ohm-m.

It is observed that the conductivity of a material is directly proportional to n, the number of free electrons/unit volume (m^3). For conductors like copper and silver, n is very large ($\approx 10^{28}$) and so these have high conductivity, whereas for insulators like wood, plastic, etc., n is small ($\approx 10^7$) because these materials have a *forbidden energy gap* so wide that few electrons cross it at room temperature.

1.1.2 Semiconductors

Semiconductors differ from conductors and insulators. They have two types of mobile charge carriers. Two of the most important semiconductors are silicon and germanium, which belong to the fourth column of the periodic table. It means that these have four valence electrons in the outer orbit (tetravalent). The crystalline structure of both silicon and germanium is the tetrahedral pattern as shown in Fig. 1.3(a), wherein each atom shares four of its valence electrons with each of the four neighbouring atoms. Such bonding is known as *covalent bonding* and can be represented schematically in two dimensions as shown in Fig. 1.3(b). At zero temperature (K), all the valence electrons are tightly bound to the nuclei and there are no free charge carriers, i.e. the material behaves as an insulator.

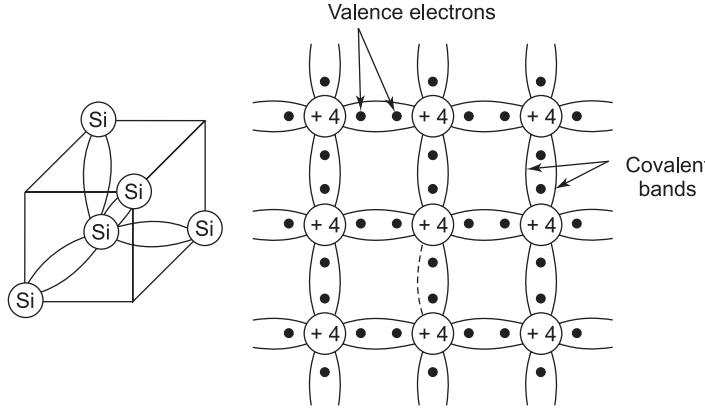


Fig. 1.3 Arrangement of atoms in a silicon crystal

The energy required to break a covalent bond is about 1.1 eV for silicon and about 0.7 eV for germanium at room temperature (about 300 K). Figure 1.4 shows the energy band diagram of silicon from which it is seen that a valence electron must acquire quantum energy jump of well above 1.1 eV to cross over to the conduction band.

*eV (electron-volt) is unit of energy when an electron moves through 1 Volt.

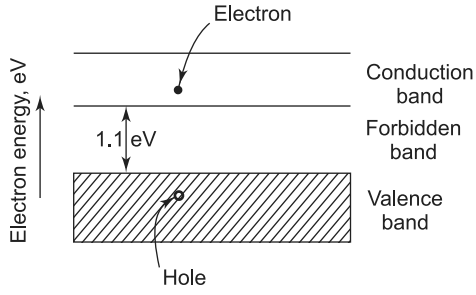


Fig. 1.4 Energy-band structure of silicon

At room temperature a few electrons have sufficient thermal energy to break the covalent bonds becoming free, leaving behind a vacancy with a net positive charge in the region, while the region surrounding the free electron has a negative charge. This is pictorially represented in Fig. 1.5, which shows one free electron, one missing covalent electron with effective positive charge called, the *hole*. Figure 1.5 also shows an electron, which has moved over to conduction band leaving behind a hole. The holes have larger effective mass than electrons.

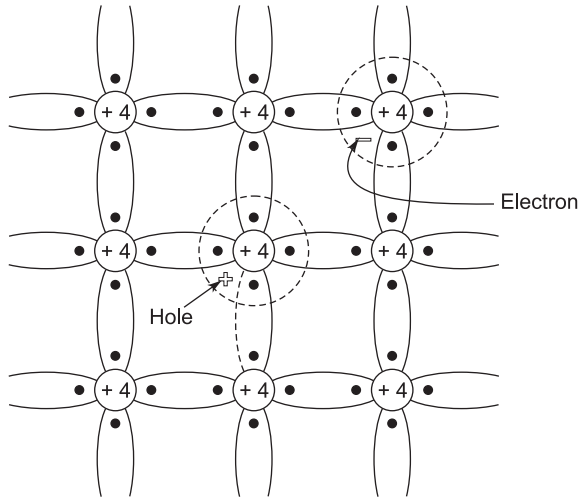


Fig. 1.5 Silicon crystal with one covalent bond broken

An electron released from a covalent bond leaves behind a hole and it may fill the hole vacancy in a neighbouring bond which is equivalent to the movement of the hole from one position to another. In other words, the hole has moved to the position vacated by the electron from that of the earlier vacant bond, which is now occupied by the electron. Thus, both electrons and holes act as charge carriers in semiconductors as different from the conduction in metals (only free electrons are the charge carriers). It must be observed here that where holes are involved, current flow always occurs because of movement of electrons but the effect is that of a hole moving in the opposite direction.

When electric field ε is applied to the semiconductor material, electrons drift in opposite direction to ε while holes drift in the direction of ε both adding to a current in the positive direction. Equation (1.2) for current density now generalises for a semiconductor as

$$J = (n\mu_n + p\mu_p)q\varepsilon \text{ A/m}^2 \tag{1.3a}$$

and conductivity

$$\sigma = (n\mu_n + p\mu_p)q \text{ S/m} \tag{1.3b}$$

where μ_n = electron mobility,
μ_p = hole mobility; the hole mobility being much lower than the electron mobility, and
n, p = number of electrons or holes/unit vol (m^3).

In a pure sample of a semiconductor material, electron-hole pairs are simultaneously generated so that

$$n = p = n_i = \text{number of electrons or holes/m}^3$$

Such a semiconductor is called *intrinsic semiconductor*. Constantly new electron-hole pairs are being generated and older ones are recombining (electron is captured by a hole and the pair vanishes). At room temperature the density of free electron-hole pair (n_i) is not very high, i.e. only a small fraction of valence electrons are in free state (about 1 in every 10^{12} atoms).

Some of the properties of Si and Ge are given in Table 1.1.

Table 1.1 *Properties of silicon and germanium at room temperature (300 K)*

Properties	Silicon	Germanium
Atomic number	14	32
Density (g/m^3)	2.33×10^6	5.32×10^6
Energy Gap (eV)	1.1	0.67
Intrinsic carrier density, n_i ($/m^3$)	14.5×10^{16}	2.4×10^{19}
Intrinsic resistivity (Ω-m)	2300	0.46
Electron mobility (m^2/V.s)	0.135	0.39
Hole mobility (m^2/V.s)	0.048	0.19

* Ref: "Properties of Silicon and Germanium II" Proc. IRE, June 1958, p. 1281.

EXAMPLE 1.1 Estimate the relative concentration of germanium atoms and electron-hole pairs at 300 K (room temperature). Also, predict the intrinsic resistivity. Given atomic weight of germanium, 72.60 g/g-atom; $q = 14.6 \times 10^{-19}$ coulombs.

Solution Using figures for germanium from Table 1.1,

Density = 5.32×10^6 g/m^3

$$n_A = \frac{6.022 \times 10^{23} \text{ atoms/g-atom} \times 5.32 \times 10^6 \text{ g/m}^3}{72.60 \text{ g/g-atom}}$$

$$= 4.41 \times 10^{28} \text{ atom/m}^3$$

Intrinsic concentration n_i (at 300 K) = 2.4×10^{19} electron-hole pairs/m^3 (see Table 1.1). $n_A/n_i = 4.41 \times 10^{28}/2.4 \times 10^{19} = 14.84 \times 10^9$ germanium atoms/electron-hole pair

$$\sigma_i = (n\mu_n + p\mu_p)q = (\mu_n + \mu_p)n_i q$$

$$= (0.39 + 0.19) \times 2.4 \times 10^{19} \times 14.6 \times 10^{-19}$$

$$= 2.227 \text{ S/m}$$

Intrinsic resistivity, $\rho_i = 1/\sigma_i = 1/2.227 = 0.449$ Ω-m

1.2 | DOPED SEMICONDUCTORS

When pentavalent/trivalent impurity atoms are introduced into the crystalline structure of a semiconductor (Si or Ge), there is tremendous increase in the concentration of free electrons/holes with a corresponding enhancement in conductivity of the material which is then known as *doped* or extrinsic semiconductor. *Pentavalent* elements commonly used are nitrogen, phosphorus, arsenic and antimony, while the *trivalent* elements are boron, aluminium, gallium and indium.

Consider the introduction of a small amount of pentavalent (impurity atoms) in the crystal structure of silicon as shown in Fig. 1.6(a). The pentavalent atom has five valence electrons, only four of which participate in the covalent bonding with neighbouring silicon atoms. The fifth electron becomes a free electron leaving behind a positively charged but immobile ion, which cannot take part in the conduction process. This type of impurity is called *donor* or *N-type impurity* as it contributes free electrons from the donor atoms. Of course, the material as a whole is neutral.

The effect of donor impurity on energy-band structure of semiconductors is illustrated in Fig. 1.6(b). It is seen that the energy gap of donor atoms is significantly smaller than that of the intrinsic material. As a consequence, the fifth valence electrons from almost all the donor atoms become free above 300 K. At room temperature in intrinsic silicon there is about one free electron for every 10^{12} atoms (1 for 10^9 Ge atoms). If doping level is 1 donor atom in 10 million (10^7) silicon atoms, the free electrons in doped material rise by ($10^{12}/10^7 = 10^5$) times compared to the intrinsic material. Thus, there is great preponderance of electrons over holes in the extrinsic material with donor impurity atoms; the material being called *n*-type wherein electrons are the *majority charge carriers* and holes are the *minority charge carriers*.

Consider now the introduction of trivalent (impurity) atom into the silicon crystal as illustrated in Fig. 1.7(a). The three valence electrons of the trivalent atom form covalent bonds with the neighbouring silicon atoms while the fourth unfilled bond acts as a hole as it can accept an electron. Such impurity is called an *acceptor* or *p-type* impurity as it leads to generation of holes in the extrinsic material. As one such hole accepts an electron, the impurity atom behaves as a negatively charged immobile ion.

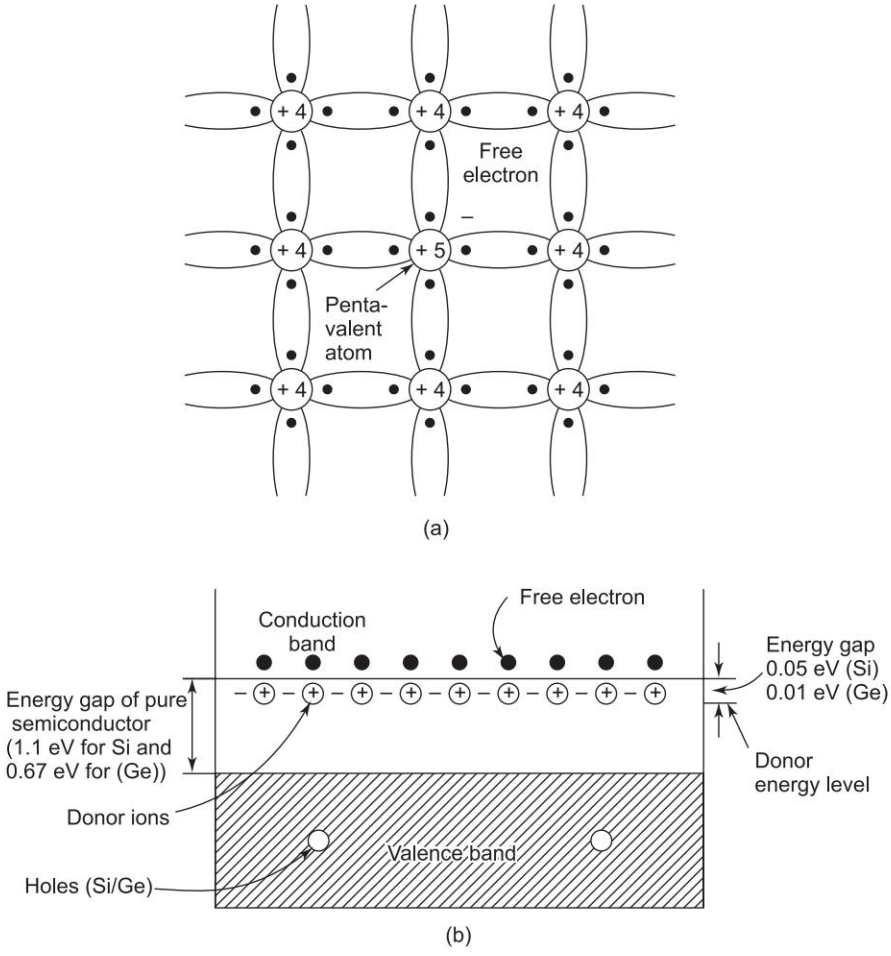


Fig. 1.6 (a) Effect of n-type doping (b) Effect of donor impurity on energy-band structure of semiconductor

Figure 1.7(b) shows the band structure of acceptor type (*P*-type) extrinsic material. In this also we find nearly as many extra holes as the number of acceptor atoms appear in the *P*-type material. Thus, in a *P*-type material, holes are the *majority carriers* and electrons are the *minority carriers*.

As an electron meets a hole, it is captured and the pair vanishes; the process is called *recombination*. The rate of recombination, R (electron-hole pairs/s. m^3) is proportional to all possible meetings of electron holes, i.e. np. Thus,

$$R = rnp;\ r = \text{constant} \tag{1.3c}$$

Under thermal equilibrium conditions, the rate of generation equals the rate of recombination, i.e. $R = g$.

For intrinsic material,

$$g_i = R_i = rn_i p_i = rn_i^2 \ (\text{as } p_i = n_i) \tag{1.3d}$$

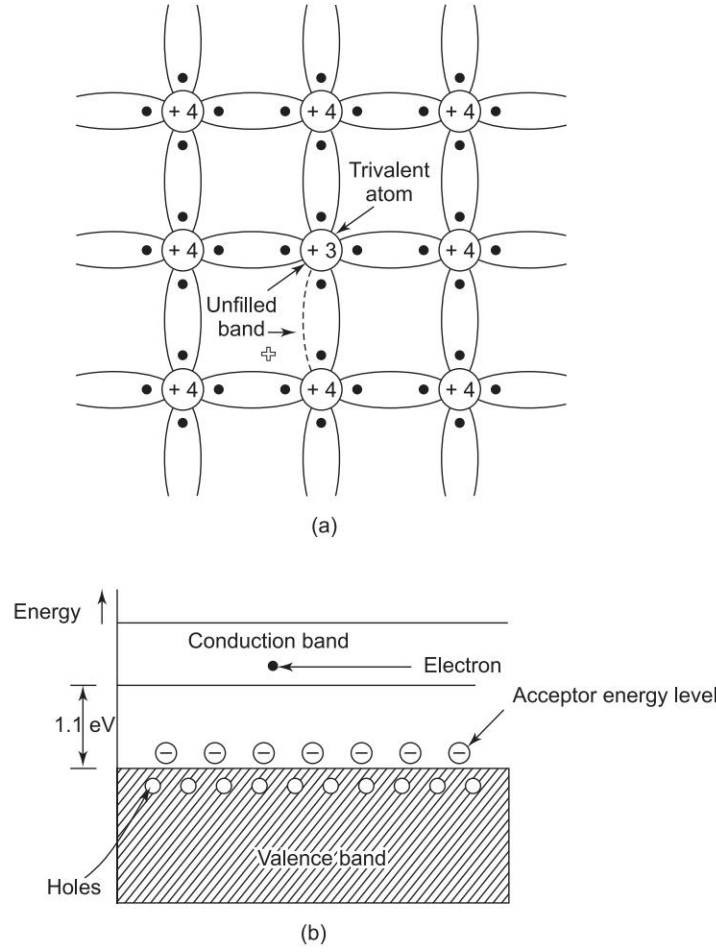


Fig. 1.7 (a) Effect of p-type doping (b) Effect of acceptor impurity on energy-band structure of semiconductor

So long as the number of impurities atoms (donor/acceptor) in doping a semiconductor are small (1 in 10 million is a typical figure), thermal generation in extrinsic semiconductor is the same as intrinsic semiconductor or

$$g = g_i = rn_i^2 \tag{1.3e}$$

In equilibrium of extrinsic semiconductor,

$$g = R$$

or $$rn_i^2 = rnp$$

or $$np = n_i^2 \tag{1.4}$$

where n, p = carrier concentrations in extrinsic semiconductor.

This relationship is known as *mass-action law*.

According to Eq. (1.4), as the concentration of one type of carrier is increased by doping, the concentration of the other type of carrier goes down below the intrinsic value.

In normal doping, the concentration of donor/acceptor atoms is such that

$$N_d >> n_i \text{ or } N_a >> n_i$$

where N_d and N_a are concentrations of donor and acceptor atoms respectively.

$$n_n \approx N_d$$

It immediately follows that the conductivity of extrinsic material is determined mainly by the doping concentration.

Diffusion

All particles in random motion, because of their thermal energies, move from the region of high concentration to that of low concentration. It is because across any imaginary plane, the random motion of particles causes more particles per unit time to cross from the high to low concentration side compared to those crossing from the low to high side. This phenomenon is known as diffusion and is statistical in nature. It is easily observed in everyday life, for example, in the spread of fragrance upon opening a bottle of perfume. It may be noted that this movement is not on account of any external force acting on particles as in the case of drift or any repulsive forces among them but is purely due to random thermal motion of the particles and is related to the concentration gradient.

The phenomenon of the diffusion of electrons and holes in the semiconductors happens in a similar way and constitutes a flow of current known as the *diffusion current*. This current is proportional to the concentration gradient of electrons/holes and is in the direction opposite to that of the positive gradient q (particles flow down the gradient). For electron diffusion, current density can be expressed as

$$\begin{aligned} J &= (-q)D_n(-dn/dx) \\ &= qD_n dn/dx \end{aligned} \tag{1.5a}$$

where D_n = diffusion constant of electrons (m^2/s) and dn/dx = concentration gradient of electrons with reference to the linear dimension in units of electrons/m^3/m.

It is assumed here that electrons are constrained by the geometrical configuration of the semiconductor so as to move in a linear dimension. Similarly, for hole diffusion current density

$$\begin{aligned} J &= qD_p(-dp/dx) \\ &= -qD_p(dp/dx) \end{aligned} \tag{1.5b}$$

where D_p = diffusion constant of holes (m^2/s), and dp/dx = concentration gradient of holes with reference to the linear dimension x in units of electrons/m^3/m.

As mobility and diffusion are both statistical phenomena, it is not surprising that they obey the *Einstein equation*

$$\mu_n/D_n = \mu_p/D_p = q/kT \tag{1.6}$$

where k = Boltzmann constant = 1.38×10^{-23} $m^2 kg\, s^{-2} k^{-1}$.

1.3 HALL EFFECT

The Hall effect is due to the nature of the current in a conductor. Current consists of the movement of many small charge carriers, typically electrons, holes, ions or all three. When a magnetic field is present, which is not parallel to the direction of motion of moving charges, these charges experience a force, called the **Lorentz force**. When such a magnetic field is absent, the charges follow approximately straight, 'line of sight' paths between collisions with impurities, phonons, etc. However, when a magnetic field with a perpendicular component is applied, their paths between collisions are curved so that moving charges accumulate on one face of the material. This leaves equal and opposite charges exposed on the other face, where there is a scarcity of mobile charges. The result is an asymmetric distribution of charge density across the Hall element, perpendicular to both the 'line of sight' path and the applied magnetic field. The separation of charge establishes an electric field that opposes the migration of further charge. Hence, a steady electrical potential is established for as long as the charge is flowing.

In the classical view, there are only electrons moving in the same average direction both in the case of electron or hole conductivity. This cannot explain the opposite sign of the Hall effect observed. The difference is that electrons in the upper bound of the valence band have opposite group velocity and wave vector direction when moving, which can be effectively treated as if positively charged particles (holes) moved in the opposite direction to that of the electrons.

Refer Fig. 1.8. Initially, the electrons follow the curved arrow, due to the magnetic force. After a short time, electrons pile up on the left side and deplete from the right side, which creates an electric field ξ_y. In steady state, ξ_y will be strong enough to exactly cancel out the magnetic force so that the electrons follow the straight arrow (dashed).

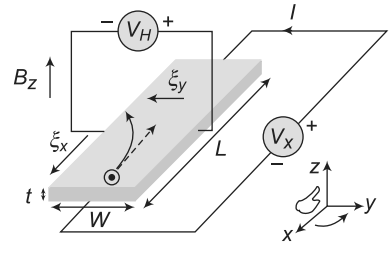


Fig. 1.8 Hall-effect measurement set-up for electrons

For a simple metal, where there is only one type of charge carrier (electrons), the Hall voltage V_H is given by

$$V_H = -\frac{IB}{net} \tag{1.7}$$

where I is the current across the plate length, B is the magnetic field, t is the thickness of the plate, e is the elementary charge, and n is the charge carrier density of the carrier electrons.

The Hall coefficient is defined as

$$R_H = \frac{E_y}{J_x B} \tag{1.8}$$

where J is the current density of the carrier electrons, and E_y is the induced electric field. In SI units, this becomes

$$R_H = \frac{E_y}{J_x B} = \frac{V_H t}{IB} = -\frac{1}{ne} \tag{1.9}$$

(The units of R_H are usually expressed as m^3/C, or Ω·cm/G, or other variants.) As a result, the Hall effect is very useful as a means to measure either the carrier density or the magnetic field.

One very important feature of the Hall effect is that it differentiates between positive charges moving in one direction and negative charges moving in the opposite. The Hall effect offered the first real proof that electric currents in metals are carried by moving electrons, not by protons. The Hall effect also showed that in some substances (especially *p*-type semiconductors), it is more appropriate to think of the current as positive "holes" moving rather than negative electrons. A common source of confusion with the Hall effect is that holes moving to the left are really electrons moving to the right; so one expects the same sign of the Hall coefficient for both electrons and holes. This confusion, however, can only be resolved by modern quantum mechanical theory of transport in solids.

The sample in homogeneity might result in spurious sign of the Hall effect, even in ideal van der Pauw configuration of electrodes. For example, positive Hall effect was observed in evidently *n*-type semiconductors.

Hall Effect in Semiconductors

When a current-carrying semiconductor is kept in a magnetic field, the charge carriers of the semiconductor experience a force in a direction perpendicular to both the magnetic field and the current. At equilibrium, a voltage appears at the semiconductor edges.

The simple formula for the Hall coefficient given above becomes more complex in semiconductors where the carriers are generally both electrons and holes which may be present in different concentrations and have different mobilities. For moderate magnetic fields, the Hall coefficient is

$$R_H = \frac{p\mu_h^2 - n\mu_e^2}{e(p\mu_h + n\mu_e)^2} \tag{1.10a}$$

where n is the electron concentration, p is the hole concentration, μ_e is the electron mobility, μ_h is the hole mobility and e is the absolute value of the electronic charge.

For large applied fields, the simpler expression analogous to that for a single carrier type holds.

$$R_H = \frac{(p - nb^2)}{e(p + nb)^2} \tag{1.10b}$$

with $$b = \frac{\mu_e}{\mu_h}$$

✦ Quantum Hall Effect

For a two-dimensional electron system which can be produced in a MOSFET, in the presence of large magnetic field strength and low temperature, one can observe the quantum Hall effect, which is the quantisation of the Hall voltage.

✦ Spin Hall Effect

The spin Hall effect consists in the spin accumulation on the lateral boundaries of a current-carrying sample. No magnetic field is needed. It was predicted by M I Dyakonov and V I Perel in 1971 and observed experimentally more than 30 years later, both in semiconductors and in metals, at cryogenic as well as at room temperatures.

✦ Quantum Spin Hall Effect

The quantum spin Hall effect has been recently observed for mercury telluride two-dimensional quantum wells with strong spin-orbit coupling, in zero magnetic field, at low temperature.

✦ Anomalous Hall Effect

In ferromagnetic materials (and paramagnetic materials in a magnetic field), the Hall resistivity includes an additional contribution, known as the **anomalous Hall effect** (or the **extraordinary Hall effect**), which depends directly on the magnetisation of the material, and is often much larger than the ordinary Hall effect. (Note that this effect is *not* due to the contribution of the magnetisation to the total magnetic field.) For example, in nickel, the anomalous Hall coefficient is about 100 times larger than the ordinary Hall coefficient near the Curie temperature, but the two are similar at very low temperatures. Although a well-recognised phenomenon, there is still debate about its origins in the various materials. The anomalous Hall effect can be either an *extrinsic* (disorder-related) effect due to spin-dependent scattering of the charge carriers, or an *intrinsic* effect which can be described in terms of the **Berry phase effect** in the crystal momentum space (*k*-space).

✦ Hall Effect in Ionised Gases

The Hall effect in an ionised gas (plasma) is significantly different from the Hall effect in solids (where the **Hall parameter** is always very inferior to unity). In a plasma, the Hall parameter can take any value. The Hall parameter, β, in a plasma is the ratio between the electron gyrofrequency, Ω_e, and the electron-heavy particle collision frequency, ν:

$$\beta = \frac{\Omega_e}{\nu} = \frac{eB}{m_e \nu} \tag{1.11}$$

where

- e is the elementary charge (approx. 1.6×10^{-19} C)
- B is the magnetic field (in Tesla)
- m_e is the electron mass (approx. 9.1×10^{-31} kg).

The Hall parameter value increases with the magnetic field strength.

Physically, the trajectories of electrons are curved by the Lorentz force. Nevertheless, when the Hall parameter is low, their motion between two encounters with heavy particles (neutral or ion) is almost linear. But if the Hall parameter is high, the electron movements are highly curved. The current density vector, J, is no longer collinear with the electric field vector, E. The two vectors J and E make the **Hall angle**, θ, which also gives the Hall parameter:

$$\beta = \tan(\theta) \tag{1.12}$$

✦ Applications

Hall probes are often used as magnetometers, i.e. to measure magnetic fields, or inspect materials (such as tubing or pipelines), using the principles of magnetic flux leakage.

Hall-effect devices produce a very low signal level and thus require amplification. While suitable for laboratory instruments, the vacuum-tube amplifiers available in the first half of the 20^{th} century were too expensive, power consuming, and unreliable for everyday applications. It was only with the development of the low-cost integrated circuit that the Hall-effect sensor became suitable for mass application. Many devices now sold as Hall-effect sensors in fact contain both the sensor as described above plus a high-gain Integrated Circuit (IC) amplifier in a single package. Recent advances have further added into one package an analog-to-digital converter and I²C (Inter-integrated circuit communication protocol) IC for direct connection to a microcontroller's I/O port.

SUMMARY

➢ Semiconductors are discussed in this chapter. They are useful in understanding electronic devices and integrated circuits.

EXERCISES

➥ Review Questions

1. Define electric field intensity, potential energy, electron-volt, mobility and conductivity.
2. Distinguish between intrinsic and extrinsic semiconductors.
3. What is a hole and how does it contribute to conduction?
4. Show the two-dimensional picture of a silicon crystal containing: (a) donor impurity atoms, and (b) acceptor impurity atoms.
5. What type of semiconductor results when doped with (a) donor and (b) acceptor impurities?
6. State the law of mass action.
7. A semiconductor has donor and acceptor concentrations of N_D and N_A respectively. How will you determine the concentration of electrons n and holes p in each type of extrinsic semiconductor?
8. What is meant by recombination?
9. What is meant by diffusion of charge carriers? How is it different from drift? Are the two related? If so, how?
10. Does the resistivity of extrinsic semiconductor increase or decrease with temperature?

➥ Problems

1. A Si bar is doped with 10^{17} boron atoms/cm^3. What is the electron concentration at 300 K? What is the resistivity?
   Given: $n_i = 14.5 \times 10^{10}$/cm^3, $\mu_n = 1350$ cm^2/V.s, $\mu_p = 480$ cm^2/V.s
2. A silicon bar, 0.1 cm long and 100 μm^2 in cross-sectional area, is doped with 10^{17}/cm^3 antimony (*V*th gp.). Find the current at 300 K with 10 V applied. Repeat for 1 in long Si bar.
3. Phosphorous donor atoms with a concentration of 10^{16}/cm^3 are added to a pure sample of silicon. Assume that the phosphorus atoms are distributed homogeneously throughout the silicon sample. What is the sample resistivity at 300 K?
4. An *n*-type sample of silicon has a uniform density of $N_d = 10^{16}$ atoms/cm^3 of arsenic and a *P*-type silicon sample has $N_a = 10^{15}$ atoms/cm^3 of boron. For each of the semiconductor materials, determine the equilibrium minority carrier concentration at 300 K.
5. The electron concentration in a piece of uniformly and lightly doped *n*-type silicon at room temperature varies linearly from 10^{17}/cm^3 at $x = 0$ to 6×10^{16}/cm^3 at $x = 2$ cm. Electrons are supplied to keep this concentration constant with time. Calculate the current density in the silicon, if no electric field is present. Given: Diffusion constant = 35 cm^2/s (*Hint*: Use electron diffusion current density relation).

➥ Multiple-Choice Questions

1. For silicon, the energy gap at 300 K is
   (a) 1.1 J (b) 1.1 eV (c) 1.1 F (d) all of these
2. *N*-Type semiconductors contains
   (a) trivalent impurity (b) tetravalent impurity
   (c) divalent impurity (d) pentevalent impurity
3. Energy gap for germanium at room temperature:
   (a) 0.67 eV (b) 0.12 eV (c) 1.1 eV (d) none of these
4. The electric charge on a hole is
   (a) 1.6×10^{8}C (b) 1.6×10^{-19}C
   (c) 1.6×10^{-11} (d) 1.6×10^{-12}
5. Energy gap for GaAs at room temperature is
   (a) 0.67 eV (b) 14.1 eV (c) 1.43 eV (d) 0.12 eV
6. The thermal voltage equation in diode is
   (a) $V_T = \frac{kT}{q}$ (b) $V_T = qkT$ (c) $V_T = \frac{q}{kT}$ (d) $V = \frac{k}{qT}$
7. The process of adding impurities to a pure semiconductor is called
   (a) doping (b) mixing (c) diffusing (d) filtering
8. The value of Boltzmann's constant *k* is
   (a) 1.38×10^{-24} J/K (b) 1.39×10^{-23} J/K
   (c) 1.38×10^{-23} J/K (d) 1.39×10^{23} J/K
9. The mobility of electrons in a material is expressed in units of
   (a) V/s^2 (b) m^2/s (c) J/K (d) m^2/V-s
10. Hall-effects occurs only in
    (a) *P*-type semiconductor (b) *N*-type semiconductor
    (c) Metal (d) all of these

ANSWERS

✦ Problems

1. $2250/cm^3$, 0.13 cm
2. 2.16 mA, 2.16 A
3. 0.463 cm
4. $22500/cm^3$, $225000/cm^3$
5. $1120\ A/cm^2$, in negative x-direction

✦ Multiple-Choice Questions

1. (b) **2.** (d) **3.** (a) **4.** (b) **5.** (c) **6.** (a) **7.** (a) **8.** (c) **9.** (d) **10.** (b)

CHAPTER 2

Diodes and Applications

GOALS AND OBJECTIVES

- ❏ Discussion of various types of diodes
- ❏ Description of *PN*-junction diode
- ❏ Zener diode and its applications as voltage regulator
- ❏ Types of rectifiers—half wave, full wave and bridge rectifier
- ❏ Rectifier with centre tapped (CT) transformer
- ❏ Different types of wave shaping
- ❏ Explanation of special purpose diodes—Schottky, barrier and photodiode

2.1 INTRODUCTION

A diode is a two-layer (*PN*-junction) device which facilitates conduction in one direction and stops conduction in the other direction. It has a wide range of applications like rectification (converting ac to dc), voltage regulation, protection against high voltage and wave shaping. Furthermore, there are special purpose diodes, e.g. Zener diode, light-emitting diode and several others.

2.2 | *PN*-JUNCTION DIODE

As shown in Fig. 2.1, the junction is formed when thin layers of *P*- and *N*-type semiconductors are joined together that results in following phenomenon immediately.

- The majority holes from *P*-side diffuse into *N*-side and vice versa.
- Recombination of electrons and holes in a narrow region on both sides of the junction results in uncovered fixed positive ions on *N*-side and fixed negative ions on *P*-side.
- This is the *depletion region* where no free electrons and holes are present.
- The electric field set up by the positive and negative ions prevents further flow of electrons and holes.
- The electric field causes the movement of minority carriers in opposite direction that provides a minority carrier *drift* current.
- In steady-state, there is no net current flow across the junction.

The simplified diagram of an open-circuit *PN*-junction diode is drawn in Fig. 2.2 where V_0 is a constant potential. The *P*-side terminal is called the *anode* and the *N*-side terminal is the *cathode*. The symbol of diode is shown in Fig. 2.3.

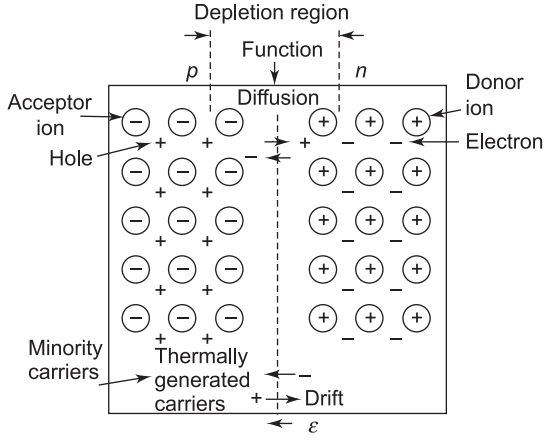


Fig. 2.1 Phenomenon at PN-junction

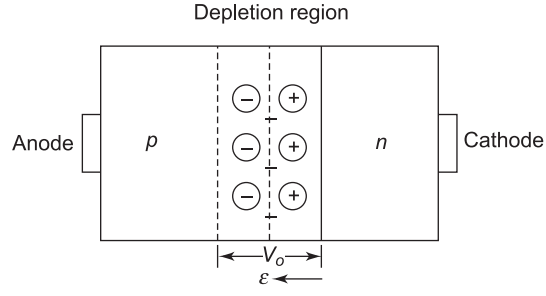


Fig. 2.2 An open-circuited PN-junction diode

2.2.1 Reverse Bias

In reverse bias, the positive terminal of the battery is connected to *N*-side (cathode) and negative terminal of the battery is connected to *P*-side (anode) as shown in Fig. 2.3.

As a result of reverse biasing, the majority of holes and electrons are pulled away from the junction. This causes the width of the depletion region to increase. Therefore, the majority carrier current cannot flow. However, the minority carrier drift current flows but stays at the saturation level I_s as the minority carrier concentrations are very low. I_s is known as the *reverse saturation current*, which is almost of negligible order (nA for Si and μA for Ge).

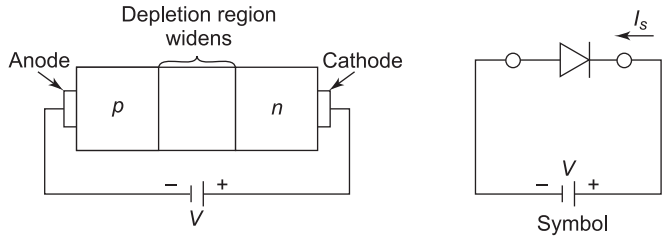


Fig. 2.3 Reverse-biased diode

2.2.2 Forward Bias

The positive terminal of the battery is connected to the anode (*P*-side) and negative terminal, to cathode (*N*-side). The holes from *P*-side and electrons from *N*-side get pushed towards the junction, thereby narrowing the depletion region. As a result, holes easily cross to *N*-side and electrons to *P*-side constituting the injunction current (I_i). The reverse saturation current I_s flows in the opposite direction. The net forward current $I_f = I_i - I_s$ increases sharply and is limited to a value determined by an external series resistance (load).

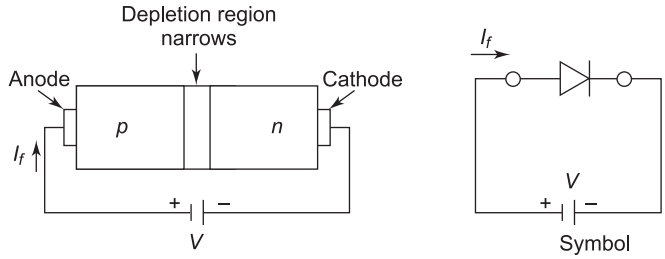


Fig. 2.4 Forward-biased diode

✦ Diode Relationship

$$I_D = I_s(e^{kV_D/T_k} - 1) \tag{2.1}$$

where I_s = reverse saturation current
k = 11,600/η; η = 1 for Ge and η = 2 for Si for low current, below the knee of the curve and
η = 1 for both Ge and Si for higher level of current beyond the knee (see Fig. 2.5).
$T_k = T_c + 273°$, and
T_c = operating temperature (25°C)

The plots of Eq. (2.1) for Ge and Si diodes are drawn to scale in Fig. 2.5. The sharply rising part of the curve extended downward meets the V_D axis, which is indicated as

V_T = offset, threshold or firing potential.

It is quite accurate to assume that $I_D = 0$ up to V_T and then increases almost linearly at a sharp slope. The values of V_T are

V_T = 0.7 V for Si diode

V_T = 0.3 V for Ge diode

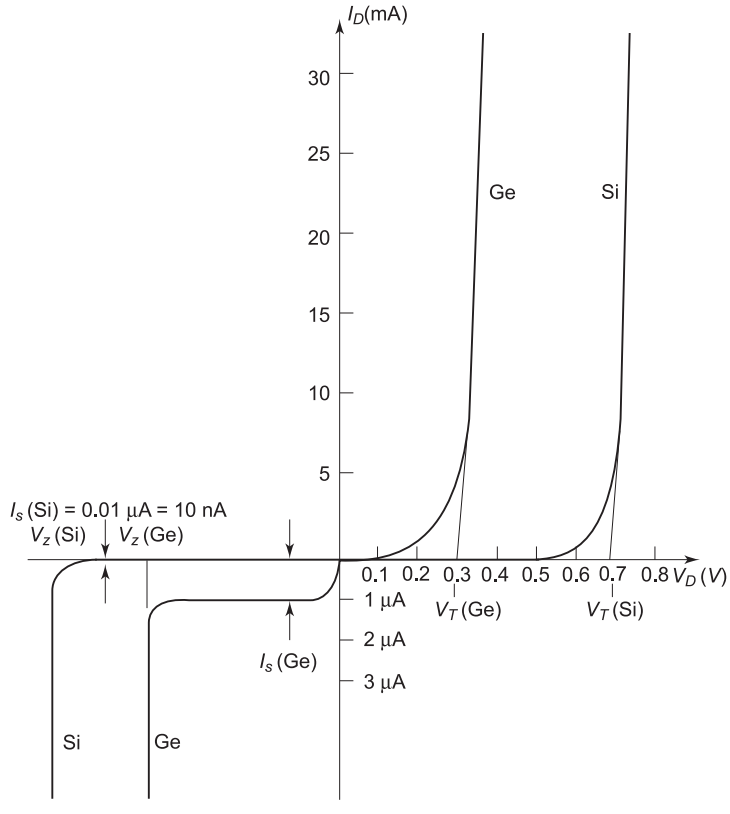


Fig. 2.5 Diode characteristics

EXAMPLE 2.1 An Si diode has I_s = 10 nA operating at 25°C. Calculate I_D for a forward bias of 0.6 V.

Solution We take $\eta = 2$

$$T_k = 25° + 273° = 298°$$

$$k = \frac{11,600}{2} = 5,800$$

$$kV_D/T_k = \frac{5800 \times 0.6}{298} = 11.68$$

$$e^{11.68} = 117930$$

Then $I_D = 10\,(117930 - 1) = 10 \times 0.117929 \times 10^6$ nA

$= 1.18$ mA, negligible.

This justifies the choice of $\eta = 2$

Note: The diode is to conduct current much larger than this value.
Therefore, I_D = 1.18 mA may be approximated as zero.

❑ ***Zener Region*** As the reverse-bias voltage is raised, the diode breaks down at voltage V_z, by *avalanche phenomenon*. The maximum negative voltage that a diode can withstand is at Peak Inverse Voltage (PIV rating).

❑ ***Zener Breakdown*** By heavily doping the *N*- and *P*-regions, the breakdown voltage V_z can be brought as low as –10 V, –5 V. This mechanism of breakdown is different from avalanche. This type of diode is called *zener diode*. When connected at a point in an electronic circuit, it does not allow the potential there to exceed the diode rated voltage.

2.2.3 Equivalent Circuit of Diode

✦ Ideal Diode

It conducts when $V_D > 0$ as shown in Fig. 2.6(a).

✦ Piecewise Linear Model

From the diode characteristic of Fig. 2.5, the piecewise linear characteristic follows and is drawn in Figs. 2.6(a), (b) and (c) along with its circuit model.

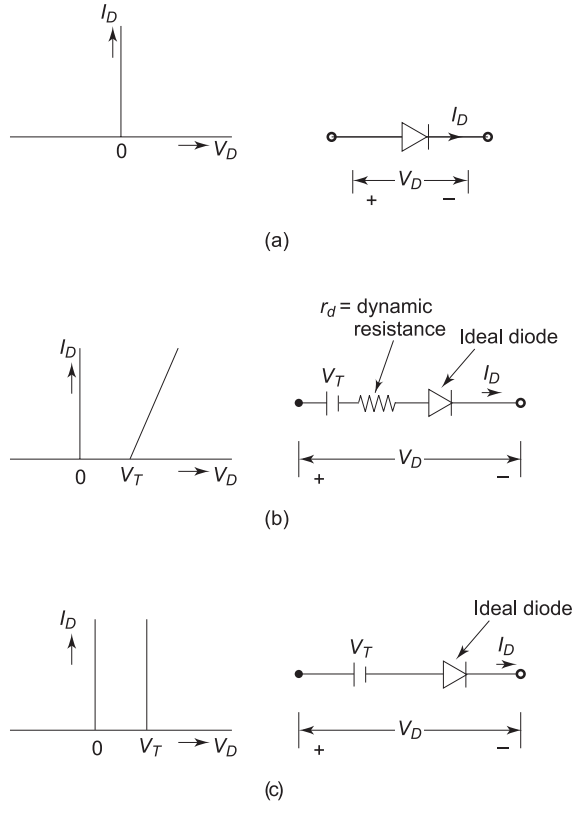


Fig. 2.6 (a) Ideal diode (b) Piecewise linear model (c) Approximate model

✦ Dynamic Resistance

$$r_d = \frac{dV_D}{dI_D} \text{ (average)}$$

It can be proved that dynamic resistance on any point of the actual IV characteristic of a diode is given by

$$r_d = \frac{26\text{ mV}}{I_D\,(\text{mA})} \tag{2.2}$$

The dynamic resistance of r_d is quite small and order of few ohms.

❑ ***Approximate Model*** Assuming $r_d = 0$, the model characteristic and circuit are drawn in Fig. 2.6(c). This equivalent circuit of diode is used most often.

EXAMPLE 2.2 For the diode circuits of Fig. 2.7, find the value of *I*. Use approximate model of the diode.

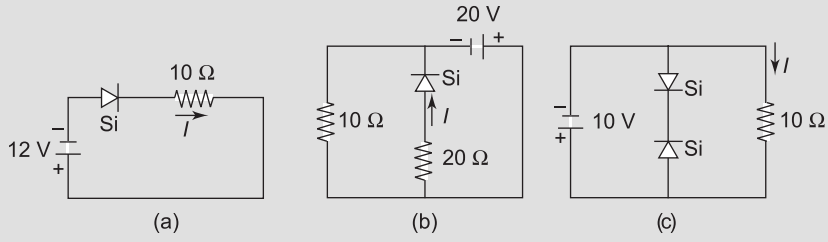


Fig. 2.7

Solution

(a) The Si diode is reverse biased by 12 V. So it does not conduct.

$$I = 0$$

(b) The voltage across diode branch is 20 V independent of 10 Ω resistance. Therefore, the diode conducts. As per equivalent circuit,

$$I = \frac{20 - 0.7}{20} = \frac{19.3}{20} = 0.965\text{ A}$$

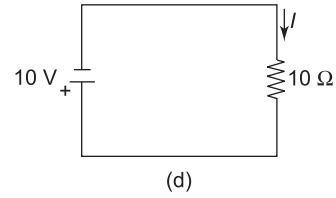


Fig. 2.7(d)

(c) The two diodes are in opposition and cannot conduct (open circuit).
Thus,

$$I = \frac{-10}{10} = -1\text{A}$$

EXAMPLE 2.3 For the diode circuits of Fig. 2.8, determine I_D and V_o using approximate model of the diode.

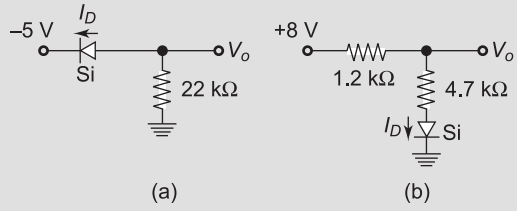


Fig. 2.8

Solution (a) The equivalent circuit is drawn in adjoining Fig. 2.9(a).

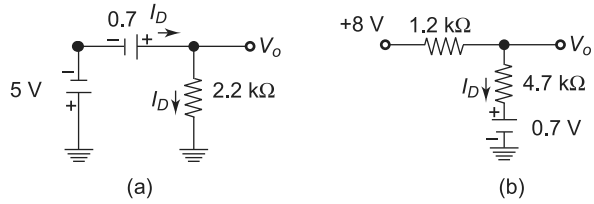


Fig. 2.9

$$I_D = \frac{5 - 0.7}{2.2} = \frac{4.3}{2.2} = 1.95 \text{ mA}$$

$$V_o = 2.2\, I_D = 2.2 \times \frac{4.3}{2.2} = 4.3 \text{ V}$$

or directly,

$$V_o = 5 - 0.7 = 4.3 \text{ V}$$

(b) The equivalent circuit is drawn in Fig. 2.9(b).

$$I_D = \frac{8 - 0.7}{1.2 + 4.7} = \frac{7.3}{5.9} = 1.237 \text{ mA}$$

$$V_o = 4.7 \times 1.237 + 0.7 = 6.51 \text{ V}$$

EXAMPLE 2.4 For the diode circuits of Fig. 2.10(a), determine V_o and I_D.

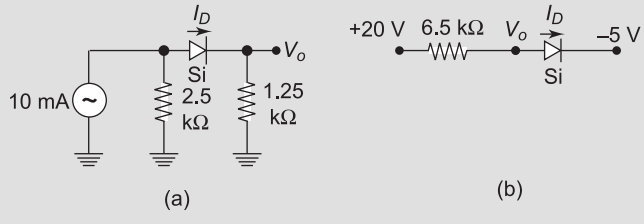


Fig. 2.10

(a) Converting current source to voltage source and diode by its circuit model, we get the circuit of the adjoining figure (Fig. 2.11).

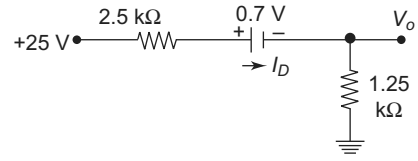


Fig. 2.11

$$I_D = \frac{25 - 0.7}{2.5 + 1.25} = 6.48 \text{ mA}$$

$$V_o = 1.25 \times 6.48 = 8.1 \text{ V}$$

(b) We can proceed directly.

$$I_D = \frac{20 - 0.7 + 5}{6.5} = 3.738 \text{ mA}$$

$$V_o = 20 - 3.738 \times 6.5 = -4.3 \text{ V}$$

or $V_o = -5 + 0.7 = -4.3$ V

EXAMPLE 2.5 For the network of Fig. 2.12, determine V_{o1} and V_{o2}

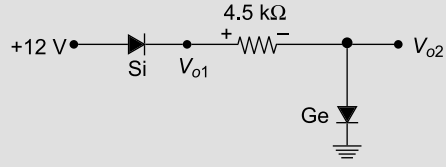


Fig. 2.12

Solution

V_{o2} = 0.3 V or V_T(Ge) = 0.3 V when conducting

V_{o1} = 12 – 0.7 = 11.3 V

Note the result does not depend on 4.5 kΩ.

EXAMPLE 2.6 For the diode network of Fig. 2.13, determine V_o.

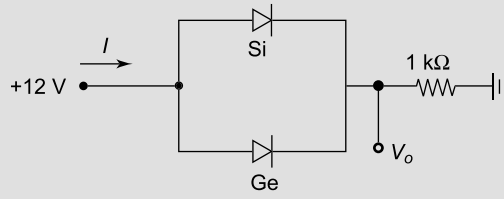


Fig. 2.13

Solution Diode Ge conducts, holding voltage at V_T = 0.3 V. Therefore, diode Si does not conduct as its V_T = 0.7 V

$$I = \frac{12 - 0.3}{1} = 11.7\text{mA}$$

$$V_o = 1 \times 11.7 = 11.7 \text{ V}$$

EXAMPLE 2.7 Determine V_o for the negative logic OR gate.

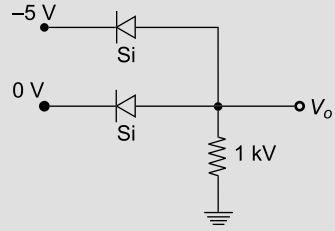


Fig. 2.14(a)

Solution Top diode conducts

$$V_o = -5 + 0.7 = -4.3 \text{ V}$$

Lower diode is negatively biased, so does not conduct.

The output will be $V_o = -4.3$ V (high of negative logic) for input –5 V at any one or both terminals. The output will be zero (low) if both inputs are zero. This is presented in tabular form in Fig. 2.14(b).

In		Out
0	0	0
0	1	1
1	0	1
1	1	1

0 ~ Low (0 V)

1 ~ High (–4.3 V)

Fig. 2.14(b)

EXAMPLE 2.8 Determine V_o for negative logic AND gate of Fig. 2.15.

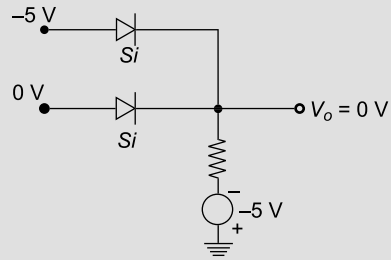


Fig. 2.15

Solution The lower diode will conduct.

$$V_o = 0 \text{ V (low, 0)}$$

If both inputs are 0 V, both diodes conduct. $V_o = 0$ V (low, 0). If both inputs are –5 V, both diodes do not conduct, $V_o = -5$ V (high, 1).

–5 V = high, 0 V = low, 0

The result is presented in the table below.

In		Out
0	0	0
1	0	0
0	1	0
1	1	1

This is negative, AND.

2.3 ZENER DIODE

A zener diode has zener breakdown in reverse bias as shown in *IV* characteristic of Fig. 2.16(a). The symbol of the zener diode is drawn in Fig. 2.16(b). Its equivalent circuit is drawn in Fig. 2.16(c). It is connected in a circuit such that it is reverse biased. It conducts only if reverse bias exceeds V_z. For positive bias, it acts as short circuit.

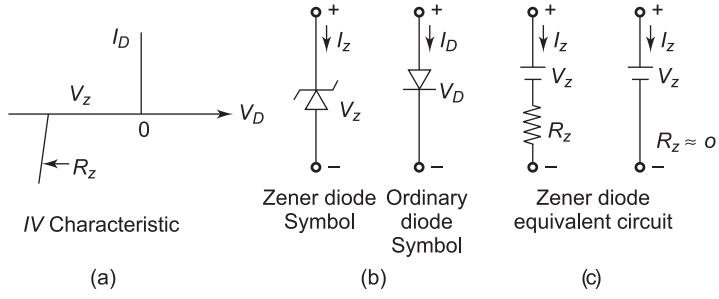


Fig. 2.16 Zener diode

2.3.1 Zener Diode as a Voltage Regulator

Voltage regulators are the devices used to maintain constant voltage across a load despite of fluctuations in the input voltage and load currents. The Zener diode in its reverse bias region is widely used as a voltage regulator as it continues to operate till the magnitude of current becomes less than $I_{Z(min)}$. The typical Zener voltage regulator is shown in Fig. 2.17. The Zener diode of breakdown voltage V_Z is connected to the input supply in reverse direction. For all the values of current within the breakdown region, the voltage across the diode will remain fixed at V_Z, giving a constant supply across its load. The resistance R_S controls the current flowing in the circuit.

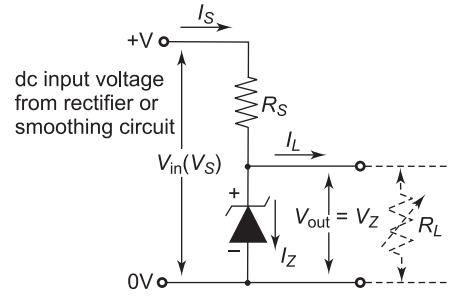


Fig. 2.17 Zener diode as a voltage regulator

❑ ***Case I: When no load is connected ($I_L = 0$)***

The current flowing in the circuit entirely passes through Zener diode. The diode dissipates maximum power. Thus, utmost care must be taken while selecting the series resistor so as to maintain the power dissipation within the range of maximum power dissipating capability of the diode.

❑ ***Case II: When load resistance R_L is connected across the diode***

Here, since the load is parallel to Zener diode, the output voltage will be equal to V_Z. The Zener current must always be above $I_{z(\min)}$ (current for which the stabilisation of voltage is effective). The higher limit of current allowed to flow in the circuit depends upon the power dissipating capability of the components used.

The voltage regulation can be done through two techniques:

1. ***Line Regulation*** In this case, series resistance and load resistance are kept constant and it is assumed that all the variations in voltage arise due to fluctuations in input power supply. The regulated output voltage is achieved for input voltage above certain minimum level. The percentage of regulation is given by

$$\frac{\Delta V_0}{\Delta V_{IN}} \times 100$$

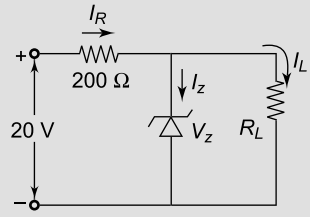
where V_0 is the output voltage, V_{IN} is the input voltage, and ΔV_0 is the change in output voltage for a particular change in input voltage ΔV_{IN}.

2. ***Load Regulation*** In this, the input voltage is fixed while the load resistance is varied. The constant output voltage is obtained as long as the load resistance is maintained above a minimum value. The percentage of regulation is given by

$$\left(\frac{V_{NL} - V_{FL}}{V_{NL}}\right) \times 100$$

where V_{NL} is the voltage across the Zener Diode when no load is applied and V_{NL} is the full load resistor voltage.

EXAMPLE 2.9 The circuit of Fig. 2.18 has a zener diode connected across the load.



V_z = 10 V, R_z = 0
$P_{z(\max)}$ = 350 mW

Fig. 2.18

(a) For R_L = 180 Ω, determine all currents and voltages.
(b) Repeat part (a) for R_L = 450 Ω.
(c) Find the value of R_L for the zener to draw maximum power.
(d) Find the minimum value of R_L for the zener to be just in on-state.

Solution

(a) As R_L is small, assume that the zener does not conduct,
i.e. $I_z = 0.$

Then, $$I_R = I_L = \frac{20}{200 + 180} = 52.6 \text{ mA}$$

$$V_z = V_L = 20 - 200 \times 52.6 \times 10^{-3} = 9.48 < 10 \text{ V}$$

So our assumption is correct.

(b) $$R_L = 450 \ \Omega$$

Assume that the zener conducts.

$$V_L = V_z = 10 \text{ V}$$

$$I_L = \frac{10}{450} \times 10^3 = 22.2 \text{ mA}$$

$$I_R = \frac{20 - 10}{200} \times 10^3 = 50 \text{ mA}$$

$$I_z = 50 - 22.2 = 27.8 \text{ mA}$$

$$P_z = 27.8 \times 10 = 278 \text{ mW} < 350 \text{ mW (rating)}$$

(c) When the zener draws maximum power,

$$I_z = \frac{350}{10} = 35 \text{ mA}$$

Then $$I_R = \frac{20 - 10}{200} \times 10^3 = 50 \text{ mA}$$

$$I_L = I_R - I_z = 50 - 3.5 = 15 \text{ mA}$$

$$R_L = \frac{10}{15} \times 10^{-3} = 667 \ \Omega$$

(d) $$I_z = 0 \text{ (just on state)}$$

$$I_R = 50 \text{ mA} = I_L$$

$$V_L = V_z = 10 \text{ V}$$

$$R_L \text{ (min)} = \frac{10}{50} \times 10^{-3} = 200 \ \Omega$$

EXAMPLE 2.10 Determine the range of V_i in which the zener diode of Fig. 2.19 conducts.

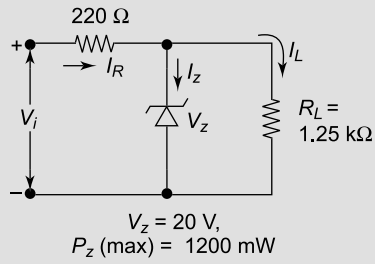


Fig. 2.19

Solution

(a) V_z just in conducting state

$$V_z = 20 \text{ V}, I_z = 0$$

$$I_R = I_L = \frac{20}{1.25} = 16 \text{ mA}$$

$$V_i = 20 + 220 \times 16 \times 10^{-3} = 23.52 \text{ V}$$

(b) $$I_z = I_z(\text{max}) = \frac{1200}{20} = 60 \text{ mA}$$

$$I_L = 16 \text{ mA}$$

$$I_R = 60 + 16 = 76 \text{ mA}$$

$$V_i = 20 + 220 \times 76 \times 10^{-3} = 36.72 \text{ V}$$

For input voltage from 23.52 V to 36.72 V, V_L will remain constant at 20 V.

EXAMPLE 2.11 For network of Fig. 2.20, determine the range of $R_L < I_L$ that will result in V_{R_L} being mentained at 10 V. Also determine wattage rating of diode.

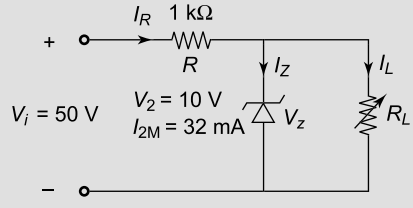


Fig. 2.20

Solution Value of R_L that will turn Zener diode on

$$R_{L_{\text{min}}} = \frac{RV_2}{V_i - V_2} = \frac{1000 \times 10}{50 - 10} = 250 \ \Omega$$

Voltage across R, i.e., $V_R = V_i - V_2 = 50 - 10 = 40$ V

$$I_R = \frac{V_R}{R} = \frac{40}{1000} = 40 \text{ mA} \Rightarrow I_{L_{\min}} = I_R - I_{2M} = 8 \text{ mA}$$

So,
$$R_{L_{\max}} = \frac{V_2}{I_{2_{\min}}} = \frac{10}{8 \text{ mA}} = 1.2 \text{ k}\Omega$$

$$P_{\max} = V_i I_{2M} = 320 \text{ mW}$$

2.4 RECTIFICATION

The diode is an ideal and simple device to convert ac to dc. The process is called rectification. We shall focus our attention on some performance measures of a rectifier: dc voltage, ripple factor, power conversion efficiency, PIV, and voltage regulation.

2.4.1 Half-wave Rectification (Sinusoidal Input)

The half-wave rectification is carried out by the simple circuit of Fig. 2.21 with a single diode. The diode conducts during positive half-cycles of input voltage and cuts off during negative half-cycle. The input and output waveforms are shown in the Fig. 2.21.

The *dc output voltage* can be expressed as

$$V_{\text{dc}} = \frac{1}{2\pi}\left[V_m \int_0^{\pi} \sin \omega t \, d\omega + 0\right] = \frac{V_m}{\pi} = 0.318 \, V_m \tag{2.3}$$

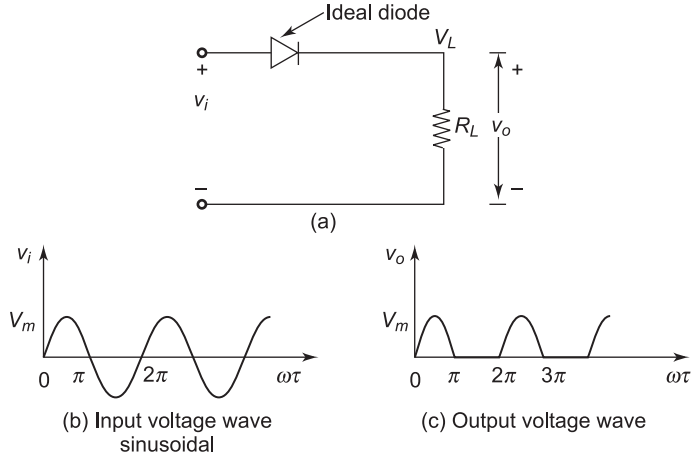


Fig. 2.21 Input/output waveforms of half-wave rectification

Then
$$I_{\text{dc}} = \frac{V_m}{\pi} \cdot \frac{1}{R_L} = \frac{I_m}{\pi} = 0.318 \, I_m \tag{2.4}$$

PIV—During negative half, the input voltage reverse biases the diode. So PIV = V_m

✦ Ripple Factor

Ripple is the variation of output voltage about dc, which is quite large in a half-rectified wave. It has a two times frequency to the frequency of the input voltage.

Ripple factor is defined as

$$\gamma = \frac{\text{rms value of the ac component of load (output) voltage}}{\text{dc component of load voltage}} \tag{2.5}$$

We can write the equality

$$V_{L\,\text{rms}} = \left[V^2_{L\,\text{ac rms}} + V^2_{L\text{dc}}\right]^{1/2}$$

or

$$\gamma = \left[\frac{V^2_{L\text{rms}}}{V^2_{L\text{dc}}} - 1\right]^{1/2} \tag{2.6}$$

$V_{L\,\text{rms}}$ = rms of half sine wave $(0 - \pi)$ over $(0 - 2\pi)$

$$V_{L\,\text{rms}} = \sqrt{\left(\frac{V_m}{\sqrt{2}}\right)^2 \times \frac{1}{2}} = \frac{V_m}{2}$$

Then

$$\gamma = \left[\frac{(V_m/2)^2}{(V_m/\pi)^2} - 1\right]^{1/2} = \left[\left(\frac{\pi}{2}\right)^2 - 1\right]^{1/2} = 1.21 \tag{2.7}$$

We find that ripple factor of a half-wave rectifier is quite high, which is unacceptable.

The value of V_{dc} can be adjusted by providing V_i from a transformer of appropriate turn ratio.

For small values of V_m, we need to replace V_m by $(V_m - V_T)$ in all the above relationships.

✦ Power Conversion Efficiency

It is defined as

$$\eta = \frac{\text{dc power output}}{\text{ac power input}}$$

Assuming the diode to be ideal in a half-wave rectifier,

$$\text{dc output} = I^2_{\text{dc}}\, R_L$$

$$\text{ac input} = I^2_{\text{rms}}\, R_L$$

$$I_{\text{rms}} = \sqrt{\left(\frac{I_m}{\sqrt{2}}\right)^2 \div 2} = \frac{I_m}{2}$$

$$\eta = \left(\frac{I_{\text{dc}}}{I_{\text{rms}}}\right)^2 = \left(\frac{I_m/\pi}{I_m/2}\right)^2 = \left(\frac{2}{\pi}\right)^2 = 0.405 \text{ or } 40.5\% \text{ (ideal)} \tag{2.8}$$

2.4.2 Full-wave Rectification

In order to reduce the ripple factor and raise the dc voltage level, we switch to full-wave rectification in which the phase of the second half of the wave is reversed. The full-wave rectified waveform is drawn in Fig. 2.22.

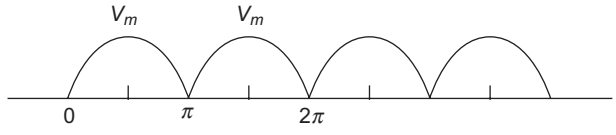


Fig. 2.22 Full-wave rectified waveform

Obviously,

$$V_{dc} = \frac{2V_m}{\pi} = 0.637\, V_m \text{ (double of half-wave)} \tag{2.9}$$

$$I_{dc} = \frac{2I_m}{\pi};\ I_m = \frac{V_m}{R_L}$$

✦ **Ripple Factor**

$$V_{i\,\text{rms}} = \frac{V_m}{\sqrt{2}}$$

Substituting in Eq. (2.7),

$$\gamma = \left[\left(\frac{V_m/\sqrt{2}}{2V_m/\pi}\right)^2 - 1\right] = \left[\left(\frac{\pi^2}{8}\right) - 1\right]^{1/2} = 0.482 \tag{2.10}$$

The ripple factor is reduced from 1.21 to 0.482.

✦ **Power Conversion Efficiency**

$$\text{dc output} = I_{dc}^2\, R_L$$

$$I_{rms} = \frac{I_m}{\sqrt{2}}$$

$$\text{ac input} = I^2_{rms}\, R_L$$

$$\eta = \left(\frac{I_{dc}}{I_{rms}}\right)^2 = \left(\frac{2I_m/\pi}{I_m/\sqrt{2}}\right)^2 = \left(\frac{2\sqrt{2}}{\pi}\right)^2 = 0.81 \text{ or } 81\% \text{ (ideal)} \tag{2.11}$$

We shall now take up two full-wave rectification circuits.

2.4.3 Bridge Rectifier

It is a bridge of four diodes as shown in Fig. 2.23. In positive half-cycle of input v_i, diodes D_2 and D_3 conduct through the load R_L. The conduction path is shown by a solid line. During

negative half-cycle, the polarity of v_i reverses; diodes D_4 and D_1 conduct through R_L in the same direction. The conduction path is shown by a dotted line. Thus, in both half-cycles, the current flows in the same direction through R_L. The voltage across R_L is, therefore, rectified voltage; see the output waveforms shown in Fig. 2.22.

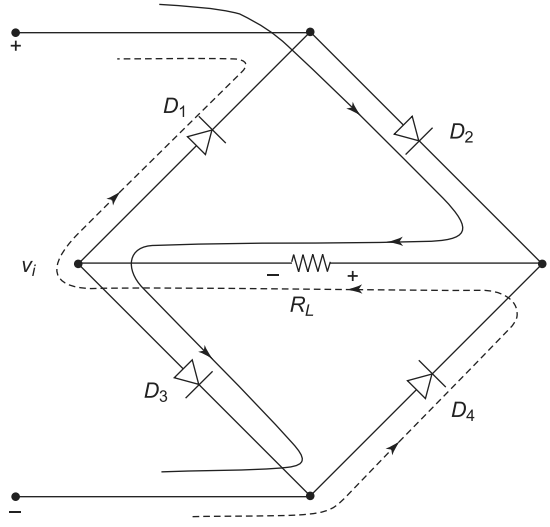


Fig. 2.23 Bridge rectifier

✦ PIV

When D_2, D_3 are conducting reverse voltage across D_1, which in off position, is equal to the voltage across R_L, its peak value is V_m. The same applies when D_1, D_4 conduct. Therefore,

$$\text{PIV} > V_m$$

A bridge rectifier can provide high dc voltages. It is commonly used in electronic circuits.

2.4.4 Rectifier with Centre-Tapped (CT) Transformer

The circuit diagram of a CT transformer with diodes connected at each end and load connected to the central tap is drawn in Fig. 2.24. During positive half-cycle of $v_i(u)$, the diode D_1 conducts feeding the load with polarities shown in the figure. During negative half-cycle, the polarities of $v_i(u)$ and $v_i(l)$ reverse. D_2 conducts feeding the load as shown by dotted lines. The polarity of the load remains the same (rectification) while v_i is sinusoidal (ac), and v_o is rectified sine wave (see Fig. 2.22). All the relationships apply with V_m = peak value of half-secondary.

✦ PIV

When D_1 is conducting, D_2 is 'off', and the reverse voltage across D_2 is $v_i(t) + v_o$. At peak value, $v_d(t) + v_o = 2V_m$. Hence,

$$\text{PIV} > 2V_m$$

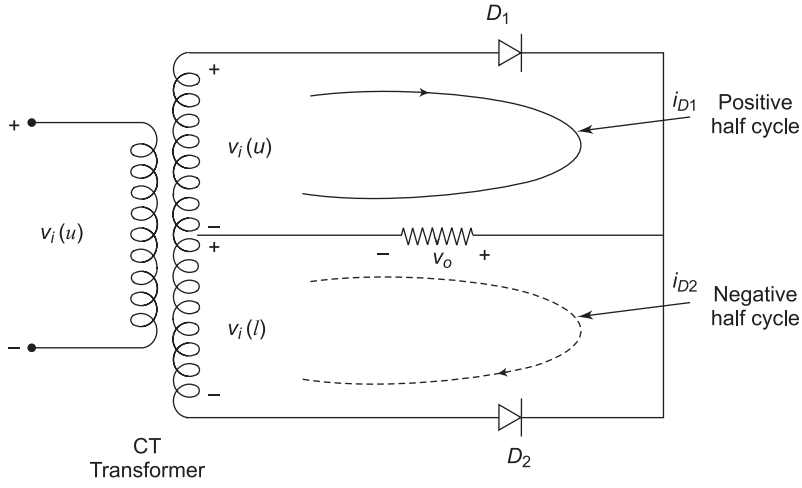


Fig. 2.24 Rectifier with CT transformer

✦ Filtering

Even in full-wave rectification the ripple factor is quite high (0.482). Therefore, the rectified output is filtered to reduce its ac content. This is achieved by connecting a capacitor (large size) across the load as shown in Fig. 2.25. Under steady conditions when v_i is less than v_o, the capacitor feeds the load (at time constant $1/R_LC$). When v_i increases above v_o, the diode conducts till $v_o = v_c$ during which period the source charges the capacitor. The output waveform is shown in Fig. 2.26, which has very much reduced ripple voltage (more smooth dc).

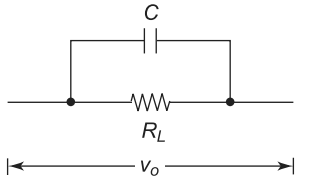
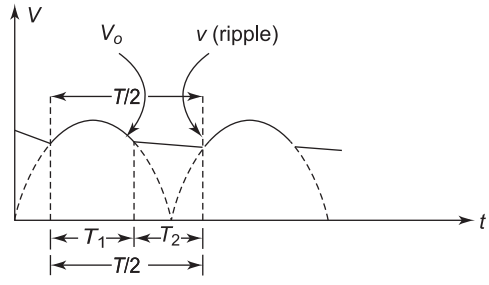


Fig. 2.25 Capacitor across load



T_1—diode conducts
T_2—capacitor conducts
v (ripple) = peak-to-peak;
ripple voltage considerably reduced

Fig. 2.26 Filtered waveform

2.4.5 Choke-capacitor Filter

The choke-capacitor filter of Fig. 2.27 is much more effective and is used for large dc power. The ripple factor can be made almost negligible because of the nonlinearity of choke not being used in low power electronic circuits.

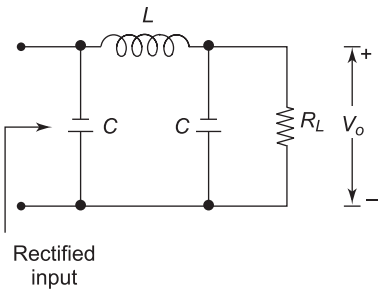


Fig. 2.27 Choke-capacitor filter

2.4.6 Peak Rectifiers

Although, dc output can be attained in a rectifier but the performance of device is limited due to pulsating output. In order to improve the output and reduce ripples, a capacitor filter is often used. In order to see the impact, let us consider a half-wave rectifier as shown in Fig. 2.28 given below. The capacitor *C* is added in parallel to load resistance *R*. The rectifier having the filter capacitor in parallel to diode is known as peak rectifier.

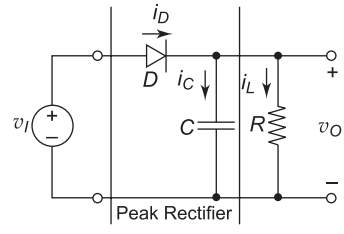


Fig. 2.28 Half-wave rectifier with filtering capacitor

In this case, capacitor charges during positive cycles of input voltage and discharges during the negative cycles. This impact is used to control the ripples across the output by controlling the discharging rate of capacitor; for this reason high values of *C* are used. The improvement in output is shown by blue line in Fig. 2.29.

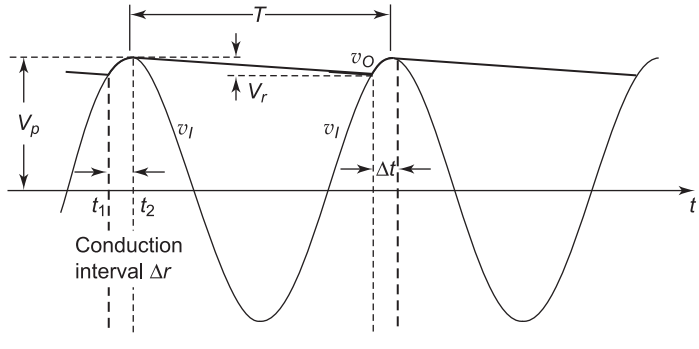


Fig. 2.29 Output waveform of peak rectifier

EXAMPLE 2.12 For the diode network of Fig. 2.30, sketch v_i, v_d and i_d if the input is sinusoidal of 50 Hz.

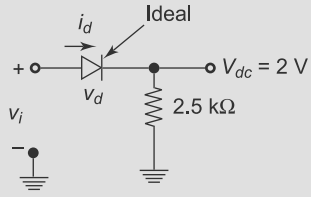


Fig. 2.30

Solution

$$V_{\text{dc}} = \frac{V_m}{\pi} = 2\,\text{V}$$

$$v_i \text{ peak} = V_m = 2\pi\,\text{V}$$

$$i_d \text{ peak} = \frac{V_m}{\pi} = \frac{2\pi}{2.5} = 2.51\,\text{mA}$$

$$v_d = 0, \text{ diode is ideal}$$

The waveform sketches are drawn in Fig. 2.31.

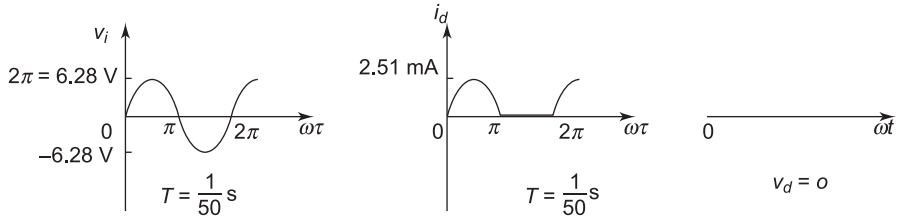


Fig. 2.31

EXAMPLE 2.13 Repeat Example 2.12 if the diode has V_T = 0.7 V.

Solution

$v_d = \dfrac{V_m - V_T}{\pi} = \dfrac{2\pi - 0.7}{\pi} = 1.78$ V, it is an approximation as output voltage is not exactly sinusoidal.

v_i peak $= V_m = 2\pi = 6.28$ V

For $v_i < V_T = 0.7$ V, $i_d = 0$ at time t_1,

$6.28 \sin 2\pi f t_1 = 0.7$ V,

$t_1 = 0.02$ s

$$i_d \text{ peak, } I_m = \frac{V_m - V_T}{2.5} = \frac{2\pi}{2.5} = 2.51 \text{ mA}$$

v_i in sinusoidal with peak 6.98 V, $f = 50$ Hz

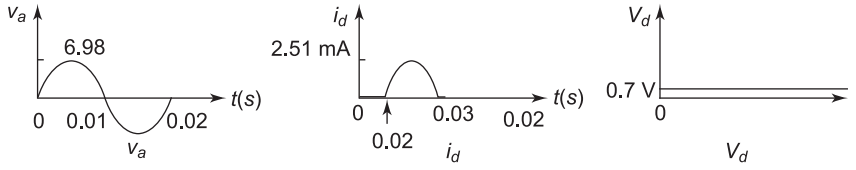


Fig. 2.32

EXAMPLE 2.14 For the network of Fig. 2.33, sketch v_o and determine V_{dc}.

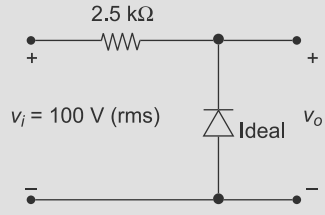


Fig. 2.33

Solution

$$\text{Input } V_m = 100\sqrt{2} = 141.4 \text{ V}$$

$$V_{dc} = \frac{V_m}{\pi} = \frac{141.4}{\pi} = 45 \text{ V}$$

During positive half-cycle of v_i, the diode does not conduct, $v_o = v_i$. During negative half-cycle of v_i, the diode conducts, causing short circuit at output terminals; $v_o = 0$. See Fig. 2.34.

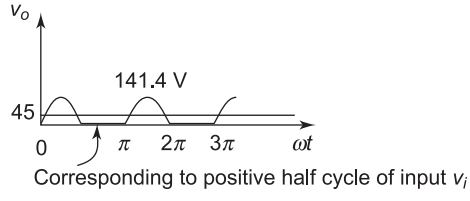


Fig. 2.34

EXAMPLE 2.15 For a sinusoidal input of 10 V peak, sketch i_R and v_o in the network of Fig. 2.35.

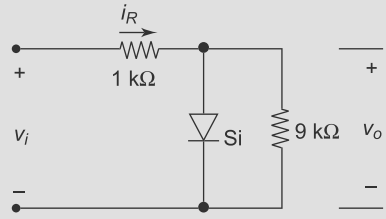


Fig. 2.35

Solution

For Si, $V_T = 0.7$ V

During positive half-cycle, the diode does not conduct up to 0.7 V.

$$v_i \Rightarrow 0 - 0.7 \text{ V}$$

$$v_o = \left(\frac{9}{10}\right) v_i$$

$$= 0.9\, v_i \text{ up to } 0.7 \text{ V}$$

$$i_R = 0.1\, v_i \text{ up to } 0.07 \text{ mA}$$

For $v_i \geq 0.7$ V, diode voltage remains constant.

$$v_o = 0.7 \text{ V},$$

$$i_R = \frac{v_i - 0.7}{1} = (v_i - 0.7) \text{ mA}$$

$$i_R \text{ (peak)} = (10 - 0.7) = 9.3 \text{ mA}$$

During negative half-cycle, the diode open circuits

$$i_R \text{ (peak)} = \frac{10}{1} = 10 \text{ mA}$$

$$v_o = 0.9\, v_i \text{ peak} = 9 \text{ V}$$

v_o and i_R are sketched in Fig. 2.36.

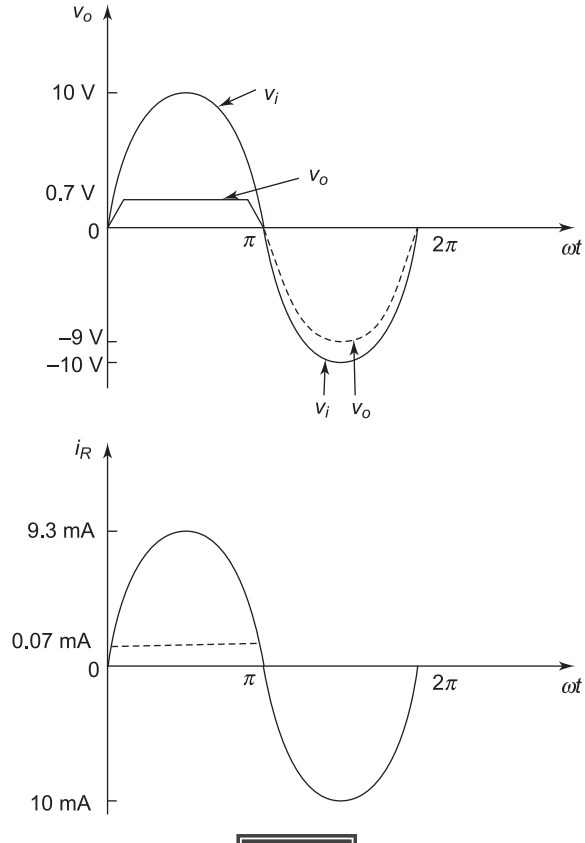


Fig. 2.36

EXAMPLE 2.16 A full-wave bridge rectifier with an input of 100 V (rms) feeds a load of 1 kΩ. $V_T = 0.7$ V

(a) If the diodes employed are of silicon, what is the dc voltage across the load?
(b) Determine the maximum current that each diode conducts and the diode power rating.
(c) Determine the PIV rating of each diode.

Solution

(a) Peak values of input, $V_m = 100\sqrt{2} = 141.4$ V

$$V_{dc} = \frac{2(V_m - 2V_T)}{\pi}, \text{ conduction path is through two diodes.}$$

or $$V_{\text{dc}} = \frac{2(141.4 - 2 \times 0.7)}{\pi} = 44.56 \text{ V}$$

(b) $$I_m = \frac{V_m - 2V_T}{1} = 140 \text{ mA}$$

$$I(\text{rms}) \text{ (each diode)} = \frac{140}{\sqrt{2}} \text{ mA}$$

$$\text{Diode power rating } P_0 = \frac{140}{\sqrt{2}} \times 0.7 = 69 \text{ mW}$$

(c) PIV > V_m = 141.4 V

2.5 | POWER SUPPLY FILTERS

As we know, almost all electronic devices require dc current/voltage source for their operation which is usually obtained using components like rectifiers. The major limitation of the dc power sources is the presence of fluctuations in the output supply. The diversion of voltage and current waveforms from pure dc is often termed as ripples. In order to enhance the dc supply, the frequency and amplitude of ripples is considered by power supply filters for quality enhancement. These filters usually employ capacitors and choke through either pi-, RC, or L-section, combinations with the load resistance.

❑ ***Capacitive filter*** A capacitive power supply filter is shown in Fig. 2.37. Here, a capacitor is used in parallel to load resistor. The fluctuations in voltage are grounded through capacitor. During the alternating cycles of power input, the capacitor charges and discharges itself, thereby maintaining the output current in same direction and hence the output voltage at a little higher level than its unfiltered value. Also, the reactance of capacitor is small as compared with load resistance, thereby providing a path to fluctuating current and pure dc supply appears across the load.

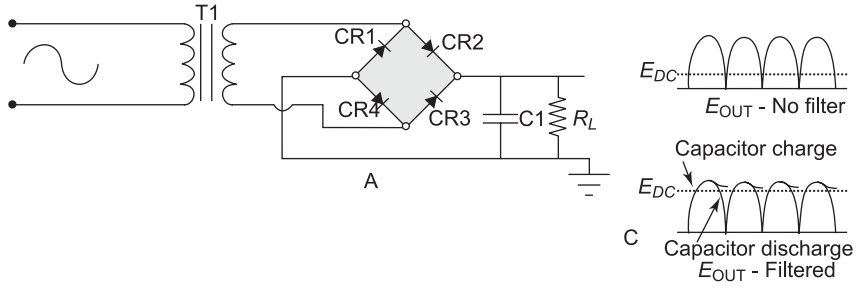


Fig. 2.37 Capacitive power supply filter

❑ ***Choke input filter*** As shown in Fig. 2.38, the choke filter implements inductor in series with the load resistance. The variation in current is opposed by inductor giving constant voltage at the output terminal. It may be noted that capacitor filter is implemented where higher and steady magnitude of voltages is required while inductive filters are implemented for requirement of steady current under varying load conditions.

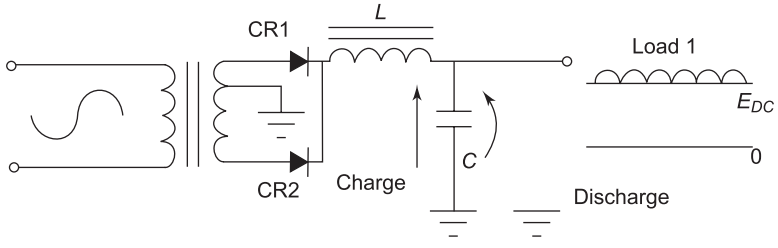


Fig. 2.38 Inductive power supply filter

2.6 | WAVE SHAPING

Diodes find several applications in wave shaping, some of which are discussed here.

2.6.1 Clamping

A clamping circuit fixes the dc level of a signal to a new value. Also, when feeding the signal to an amplifier, a dc component, if present, must be removed. In a TV signal, the dc level need to be clamped to build the peak value of a complex TV signal.

A clamping circuit requires a diode, a capacitor and a resistor. It may also need an independent source. The circuit is so constructed that during charging, there is no series resistance and so the *capacitor charges instantaneously*. While discharging, the time constant $RC >> T = 1/f = 2\pi/\omega$ (signal time period). During this period, the capacitor *practically holds charge*, after which begins the next charging cycle.

Consider the diode circuit of Fig. 2.39(a) in which C and R fulfill the conditions mentioned above. The input voltage wave is drawn in Fig. 2.39(b). During the negative half cycle of input, the diode conducts and the capacitor gets charged to a polarity indicated in Fig. 2.39(b), beyond which the diode open circuits. The output voltage is

$$v_o = A + A \sin \omega t$$

which is sketched in Fig. 2.39(c). The output is clamped at the minimum value of 0. Any small loss of charge is made up when the output touches zero and the diode conducts momentarily.

✦ Observation

In clamping, the peak-to-peak value of the signal at output is same as of input (2 A). The signal has shifted up by +A.

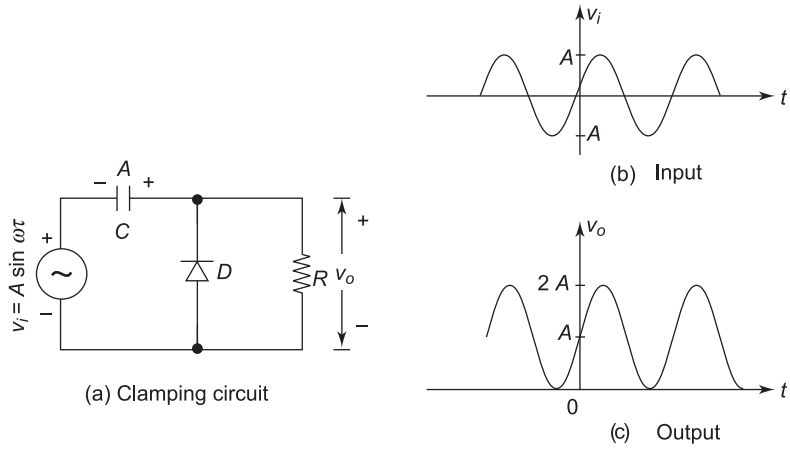


Fig. 2.39 Clamping

2.6.2 Clamping Circuit Diode with Negative Bias

During negative half cycle, the diode conducts and the capacitor charges to $(V_m - 10)$, polarity indicated in the Fig. 2.40. The output:

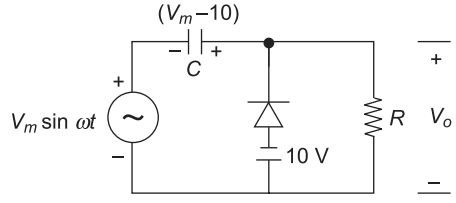


Fig. 2.40 Clamper diode with negative bias

Positive peak

$$V_m + (V_m - 10) = 2\,V_m - 10$$

Negative peak

$$-V_m + (V_m - 10) = 10 \text{ V}$$

The output waveform in sketched in Fig. 2.41. The minimum value is clamped at –10 V.

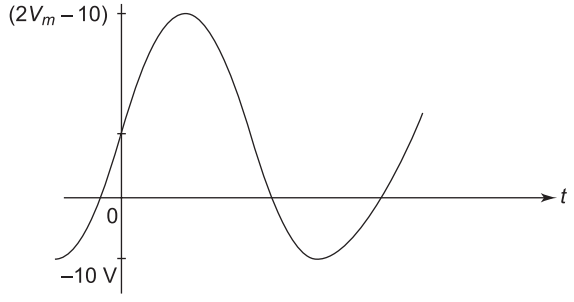


Fig. 2.41 Output waveform

2.6.3 Clamping Circuit with Pulse Input

For the clamping circuit of Fig. 2.42, determine the output for a rectangular pulse train of Fig. 2.43.

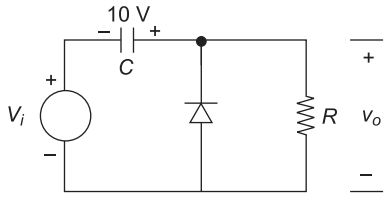


Fig. 2.42

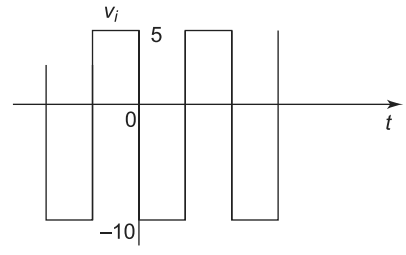


Fig. 2.43

When v_i = –10 V, the diode conducts and the capacitor charges to 10 V, polarity shown in Fig. 2.43. Then

$$v_o = v_i + 10 \text{ V}$$

Thus, the pulse train moves up by 10 V. The pulse varies from (5 + 10) = +15 V to (–10 + 10) = 0 V. The reader may sketch the wave.

2.6.4 Clipping

A part of the input signal is cut off or dipped while the remaining part is intact. A half-wave rectifier is a clipper. For clipping, a diode is used in series or shunt. In a rectifier, the diode is used in series. We will take up the circuit with the diode in shunt.

2.6.5 Shunt Diode

A clipping circuit with shunt diode and voltage source is drawn in Fig. 2.44(a). The diode conducts, when $v_i > V_1$ and the output becomes zero, as a result, this part of the signal is clipped. The part of the signal $v_i < V_1$ appears as such at the output as the diode is open circuited. For a sine wave having $V_m > V_1$, the output wave shape is drawn in Fig. 2.44(b).

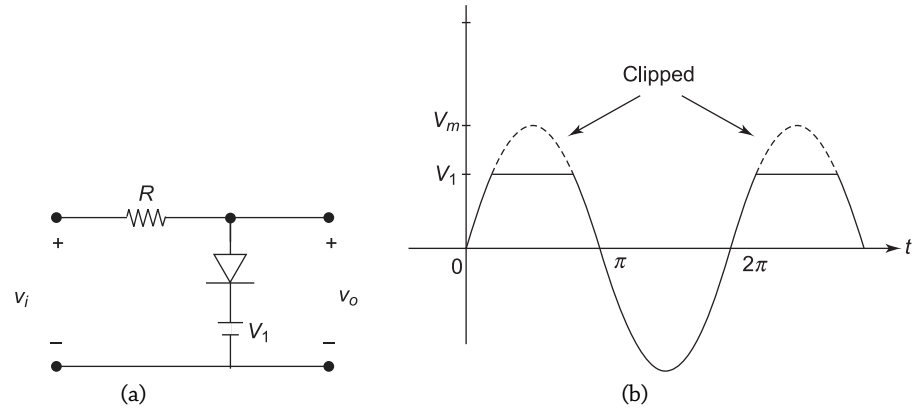


Fig. 2.44 (a) Clipping circuit with shunt diode and voltage source (b) Output wave shape

2.6.6 Voltage Multiplier

Voltage multiplier is a rectifier that provides output many times higher than the input voltage. It steps up the output voltage just as the transformers do in case of ac supplies. The device is widely employed in microwaves and high-voltage test equipment requiring high voltage dc supply. It implements the combination of diode and capacitors to efficiently deliver a dc voltage whose magnitude is integer multiple of the input peak voltage. The different multipliers include half-wave multiplier, full-wave multiplier, etc.

✦ Half-wave

The voltage doubler circuit is drawn in Fig. 2.45.

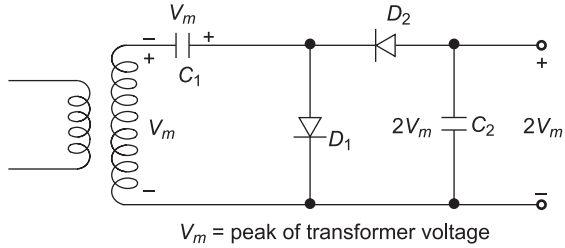


Fig. 2.45 Half-wave voltage doubler circuit

- During the positive half-cycle, D_1 conducts and C_1 gets charged to V_m. D_2 does not conduct and C_2 is uncharged.
- During negative half-cycle, the transformer voltage sign reverses. D_2 conducts through C_2 so that C_2 gets charged to $V_m + V_m = 2\ V_m$, output.

✦ Full-wave

The circuit is drawn in Fig. 2.46

- During positive half-cycle, D_1 conducts and C_1 charges to V_m.
- During negative half-cycle, D_2 conducts and C_2 charges to V_m
- Output $V_m + V_m = 2\ V_m$

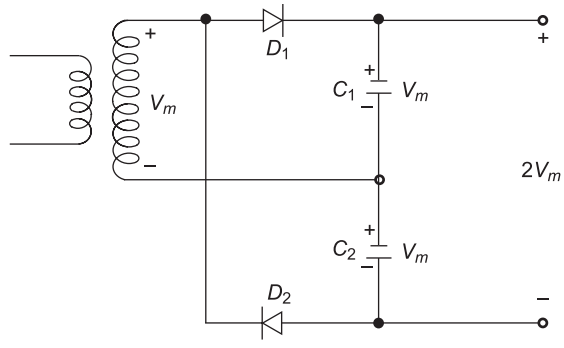


Fig. 2.46 Full-wave voltage doubler

2.6.7 Peak Detector

The circuit of a peak detector is shown in Fig. 2.47. It is indeed a half-wave rectifier with an *RC* filter. Consider the input is sinusoidal as shown in Fig. 2.48. During positive half-cycle, the diode conducts and *C* charges to *A* volts, the diode now stops conducting. If $RC >> T = 2\pi/\omega$, v_o decays exponentially till slightly less than $5\pi/\omega$. The diode again conducts and charges to *A*; the process thereon repeats.

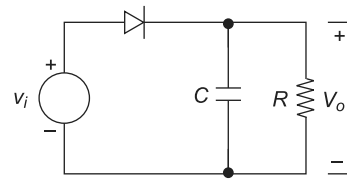
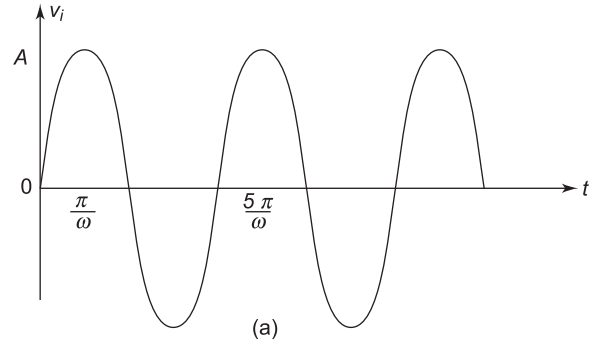


Fig. 2.47 Peak detector

The input has a small ripple voltage and its average value is close to the peak value *A*. This circuit is employed in *envelop detection* of Amplitude Modulated (AM) wave.



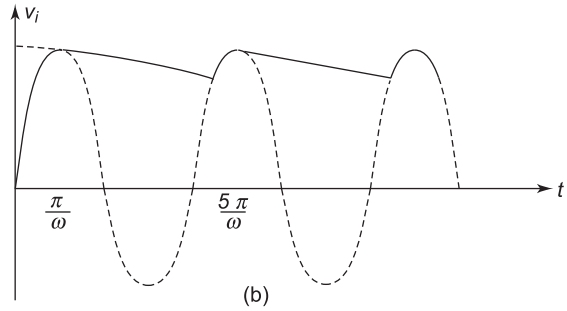


Fig. 2.48 (a) Sinusoidal input (b) Output

2.7 SPECIAL-PURPOSE DIODES

The *IV* characteristics of a *PN*-junction can be considerably modified from that of a normal diode by control of doping level and choice of material. These are other two-terminal devices which are highly sensitive (or produce light). These belong to the general category of photoelectric devices. All these will be briefly introduced here.

2.7.1 Schottky Barrier Diode

It is formed by metal-semiconductor junction, which provides rectifying action. The constructional details are shown in the cross-sectional view of Fig. 2.49(a). The metal (Al) contact with the n^+ semiconductor forms a rectifying junction. The majority electrons from the semiconductor flow in large number into the metal, that has its own pre-electrons by smaller energy. The semiconductor gets somewhat deflected by electrons. The barrier potential so formed prevents any further flow of electrons. The Al metal forms the *anode* and the n^+ region is the *cathode*.

When a forward bias is applied, the conduction is solely by electrons as there are no holes in the metal. The threshold voltage $V_T = 0.3$ V is less than 0.7 V for a silicon ordinary diode. This means less power dissipation in a Schottky diode. Further, the recovery time upon switch-off is very small, about 10–20 ns, as there is no recombination of electron-holes.

The symbol of Schottky diode and equivalent circuit are shown in Fig. 2.49(b) and (c).

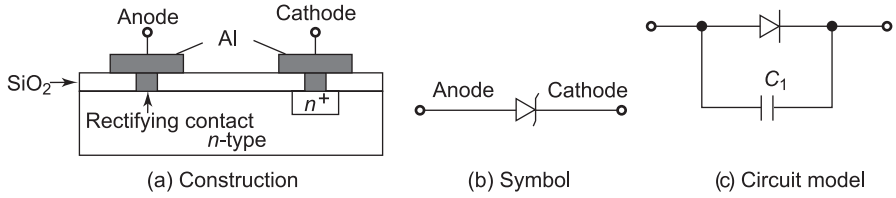


Fig. 2.49 Schottky diode

2.7.2 Tunnel Diode

As the doping levels of *N*- and *P*-sides of a diode are heavily increased, the depletion region becomes very narrow. When the depletion layer width becomes 1 nm or less, low-energy electrons are able to cross the barrier by a process called *tunnelling*. So the conduction begins at much lower values of bias voltage, where there is no injunction current.

The *IV* characteristic of a tunnel diode is sketched in Fig. 2.50.

As the bias voltage increases, the current reaches its maximum value I_p at V_p, beyond which no more electrons are available and holes have decreased. The current now begins to reduce reaching the minimum value of I_v at V_v. In this region, the diode offers negative resistance; beyond V_v the diode behaves like a normal diode.

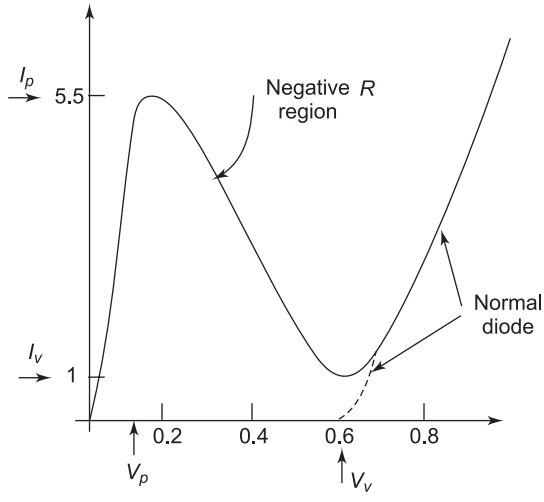


Fig. 2.50 Characteristic of tunnel diode

✦ Application

In a high-frequency *LC* oscillator circuit, tunnel diode is included whose negative portion makes up the loss in *LC* circuit resistance.

The figure of merit of a tunnel diode is the ratio (I_p/I_v). Its value is 10 for germanium and 20 for gallium arsenide.

2.7.3 Varactor (Varicap) Diode

A diode with reverse bias has wide depletion region and positive and negative charge on both side of it. Therefore, it possesses a capacitance.

$$C_T = \in \frac{A}{W_d};$$

A = area of depletion region , W_d = its width, $\in$ = permittivity of semiconductor material C_T is called *transition capacitance*. As the width of the depletion region varies with reverse voltage (V_R), the transition capacitance varies accordingly. The transition capacitance can be approximated as

$$C_T = \frac{K}{(V_T + V_R)^n} \tag{2.12}$$

where V_T = threshold voltage = 0.7 V for Si; 0.3 V for Ge; Barrier potential
V_R = reverse bias (magnitude), can be a maximum of 20 V
K = constant of semiconductor material
n = ½ for alloyed junction and 1/3 for differed junction

A typical plot of C_T vs V_R is shown in Fig. 2.51(a) where $C(0)$ is the capacitance at $V_R = 0$. Equation (2.12) can be expressed as

$$C_T = \frac{C_T(0)}{\left(1 + \frac{V_R}{V_T}\right)^n} \tag{2.13}$$

The equivalent circuit and symbol of a varicap are drawn in Fig. 2.51(b) and (c).

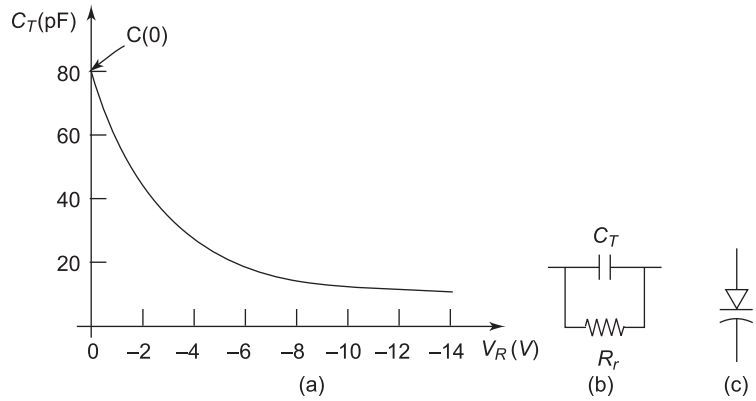


Fig. 2.51 (a) Plot of C_T vs V_R (b) Equivalent circuit (c) Symbol

R_r = reverse resistance of diode, very large (more than 1 MΩ)

Change in temperature affects the thermally generated carriers and therefore, the transition capacitance. The temperature coefficient of $C(0)$ is provided by manufacturers.

Being a reverse biased diode the capacitance is highly temperature dependent. The capacitance temperature coefficient is defined as

$$TC_c = \frac{\Delta C}{C_0(T_1 - T_0)} \times 100\%$$

where C_0 is the capacitance at T_0.

✦ Application

A varactor is employed as variable capacitance of tank-circuit high-frequency oscillators for changing the oscillation frequency.

2.7.4 Photodiode

The field of photoelectrons has quite a variety of applications and has been attracting deep research interest. Here, we will study two kinds of devices—one in which light controls diode current and the other in which diode emits light when carrying current.

✦ Light Definition and Units

As per the quantum theory, light is in the form of photons and each photon delivers an energy packet to the surface on which it falls.

$$W = hf \text{ joules}$$

where h = Planck's constant (6.624×10^{-34} joule seconds)

f = frequency of light waves in Hz.

Note that light also behaves as a travelling wave.

The frequency of light is directly related to its wavelength (distance between successive peaks) as

$$f = v/\lambda$$

where v = velocity of light (3×10^8 m/s)

λ = wavelength in metres

Units of wavelength are angstrom (Å) or micrometer (μm).

1 Å = 10^{-10} m, 1 μm = 10^{-6} m

Intensity of light is measured in units of luminous flux incident on unit area. Units of luminous flux are lumens where

$$1 \text{ lm} = 1.496 \times 10^{-10} \text{ W}$$

Practical unit of intensity of light is

1 lm/ft^2, called foot-candle (fc) = 1.609×10^{-9} W/m^2

A *photodiode* is a *PN*-junction (silicon/germanium) operates in reverse-bias region as shown in Fig. 2.52. The reverse saturation current I_λ (μA) is limited by the availability of thermally generated minority carrier. As light is made to impinge on the junction, the light photons impart energy to the valence electrons causing more electron-hole pairs to be released. As a result, the concentration

of minority carriers increases and so does the current I_λ. The symbol of a photodiode is drawn in Fig. 2.52(b).

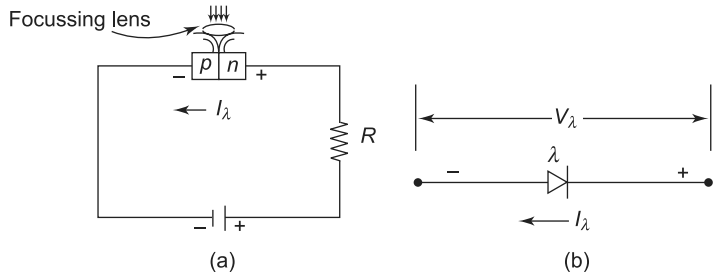


Fig. 2.52 (a) Photodiode in reverse bias (b) Symbol

The *IV* characteristics for various values of light intensity (*fc*) are drawn in Fig. 2.53. The dark current characteristic corresponding to no-light impingement (($I_\lambda = I_s$). By examining the characteristics it is found that at a certain V_λ (say 20 V), I_λ increases almost linearly with f_c.

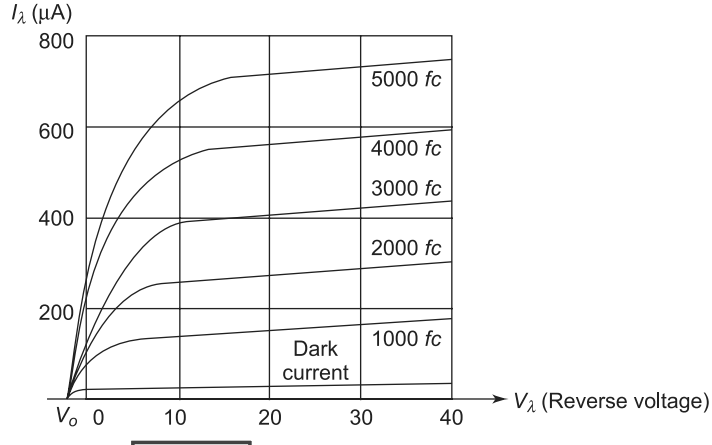


Fig. 2.53 IV characteristics of photodiode

It has been found that Ge photodiode has more overlaps compared to Si, which is in the range of light frequencies to which the human eye is sensitive. Ge is, therefore, more suitable for infra-red (IR) light sources like laser.

2.7.5 Light-Emitting Diode (LED)

In a forward biased *PN*-junction, diode recombination of electrons and holes takes place at the junction and within the body of the crystal, particularly at the location of a crystal defect. Upon capture of a free electron by a hole, the electron goes into a new state and its kinetic energy is given off as heat and as light photons. In a silicon diode, most of this energy is given off as heat but in other materials such as gallium arsenide (GaAs) or gallium phosphide (GaP), sufficient number of photons (light) are generated so as to create a visible source. This process of light emission in *PN*-junctions of such materials is illustrated in Fig. 2.54 and is known as

electroluminescence. The metal contact of *P*-material is made much small to permit the emergence of maximum number of photons so that in an LED, the light lumens generated per watt of electric power is quite high. Intensity of light increases almost linearly with forward current, depending on the material used.

The voltage levels of LEDs are 1.7 V to 3.3 V which is compatible with the solid-state circuits. The response time is short (only a few nanoseconds) and light contrast is good.

LEDs emit light red, green, orange or blue.

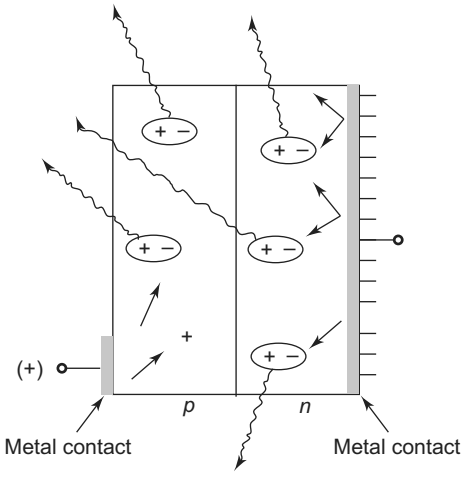


Fig. 2.54 Light emission in PN-junction

✦ Applications

LEDs find several display applications, particularly in 8-segment display of numbers 0 to 9. These are now being used in LED TV's.

2.7.6 Photocoupler

It is a package of an LED and photodiode whereas circuits are electrically isolated as shown in Fig. 2.55. The LED is forward biased and the photodiode is reverse biased. The output is available across R_2.

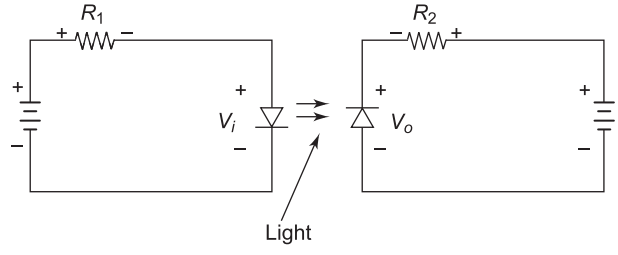


Fig. 2.55 Photocoupler

The key advantage of the photocoupler is the electrical isolation between two circuits. It is employed to couple circuits whose voltage level may differ by several thousand volts.

EXAMPLE 2.17 Determine the capacitance of diffused junction varactor at a reverse potential of 4.5 V, if $C(0) = 85$ pF and $V_T = 0.7$ V. Also determine the values of K in Eq. (2.12).

Solution

$$C_T = \frac{C_T(0)}{\left(1+\frac{V_R}{T}\right)^n} = \frac{85}{\left(1+\frac{4.5}{0.7}\right)^{1/3}} = \frac{85}{1.9511} = 43.565 \text{ pF}$$

$$C_T = \frac{K}{(V_T + V_R)^n}$$

$$43.565 = \frac{K}{(0.7 + 4.5)^{1/3}}, \; K = 75.475$$

EXAMPLE 2.18 For a photodiode, determine I_λ if V_λ = 30 V and intensity of light is 3.22×10^{-6} W/m^2. Read from Fig. 2.52.

Solution Light intensity in *fc*

$$fc = \frac{3.22 \times 10^{-6}}{1.609 \times 10^{-9}} = 2000$$

Reading from Fig. 2.52 we find

$$I_\lambda \approx 300 \; \mu A$$

SUMMARY

➢ Various types of diodes, rectifiers have been described in this chapter along with their characteristics and applications.

EXERCISES

➥ Review Questions

1. Explain what is depletion region in a *PN*-junction diode?
2. What is reverse saturation current in a diode? Does it exist in both reverse-biased and forward-biased diodes.
3. What is threshold voltage of a diode? What is its value for Si and Gi diodes?
4. Write the diode conduction equation. Explain the meaning of each symbol.
5. Draw the circuit equivalent of a forward-biased diode.
6. Explain the operation of a zener diode and draw its circuit equivalent.
7. A zener diode acts as a voltage regulator. Explain the meaning of the statement.
8. Draw the circuit of a bridge rectifier. What is the input and output waveform?
9. Write the expression of dc voltage of half-wave and full-wave rectifiers.
10. What is the ripple factor of a diode rectifier? Derive its expression for a full-wave rectifier. Will the ripple factor be more or less than this value for a half-wave rectifier?
11. What is the conversion efficiency of a rectifier circuit? Derive its expression for a full-wave rectifier. Will the value be more or less than this in a half-wave rectifier?

➥ Problems

1. For the diode circuit of Fig. 2.56, determine I, V_1, V_2, V_0.

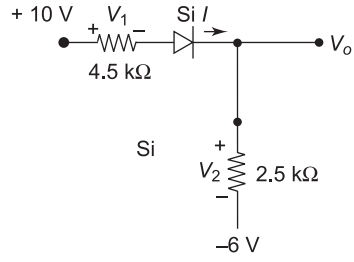


Fig. 2.56

2. For the diode circuit of Fig. 2.57, determine I_1, I_2, I_{D2}.

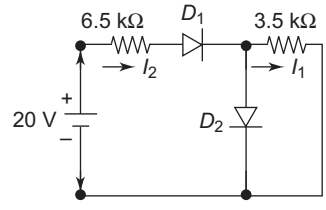


Fig. 2.57

3. For the diode OR gate logic of Fig. 2.58, prepare a table of inputs and corresponding outputs.

4. For the diode AND logic of Fig. 2.59, prepare a table of inputs and corresponding output.

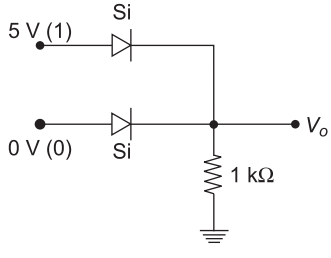


Fig. 2.58 OR gate

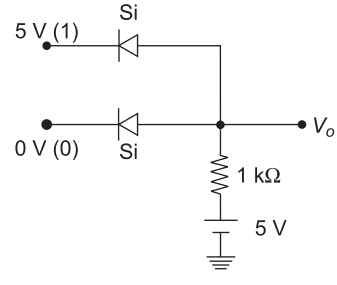


Fig. 2.59 AND gate

5. In the zener diode circuit of Fig. 2.60, V_L is to be maintained constant at 12 V, while I_L varies from 0 to 250 mA. Calculate the values of V_z, R_S and power rating of the zener.

6. For the circuit with zener diode (Fig. 2.61), determine the output for V_i = 50 V and V_i = 5 V.

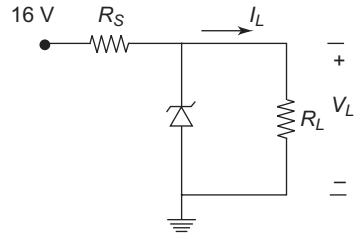


Fig. 2.60

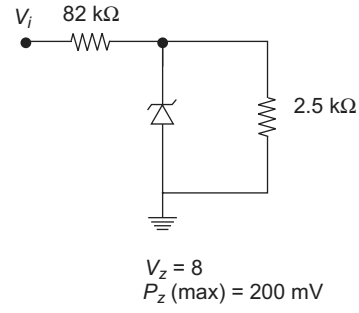


Fig. 2.61 P_z(max) = 200 mW

7. In a Zener diode voltage regulator circuit, the source series resistance R_S = 20 Ω, Zener voltage V_Z = 18 V, and load resistance R_L = 200 Ω. If the source voltage V_S varies from 20 V to 30 V, find the maximum and minimum current in the diode?

8. For the diode bridge of Fig. 2.62, V is sinusoidal with V_m = 100 V. Sketch v_o. Find V_o(dc) and PIV of each diode.

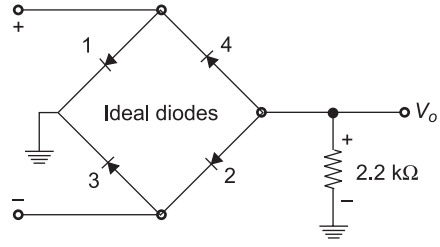


Fig. 2.62

9. For the diode network of Fig. 2.63, which is sinusoidal input with V_m = 170 V, determine V_o(dc).

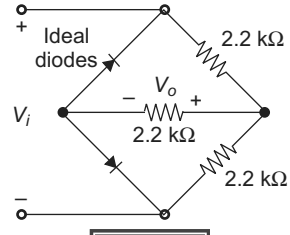


Fig. 2.63

10. A CT transformer full-wave rectifier has ac voltage of each half-secondary of 20 V(rms). The resistance of each half is 1 Ω. Diodes are Si–V_T = 0.7 V, r_d = 0.5 Ω. The load resistance is 15.5 Ω. Determine V_{dc} and I_{dc} of the load.

11. A bridge rectifier has four identical diodes of forward resistance of 5 Ω each. It is supplied from a transformer with output voltage of 20 V(rms) and secondary winding resistance of 10 Ω. Calculate the

(a) dc output voltage at a dc load current of 100 mA

(b) rms value of output voltage at a dc load current of 200 mA

(c) rms value of the ac component of the voltage in part (b).

12. For the circuit of Fig. 2.64 and the given input signal, determine v_i and the output signal.

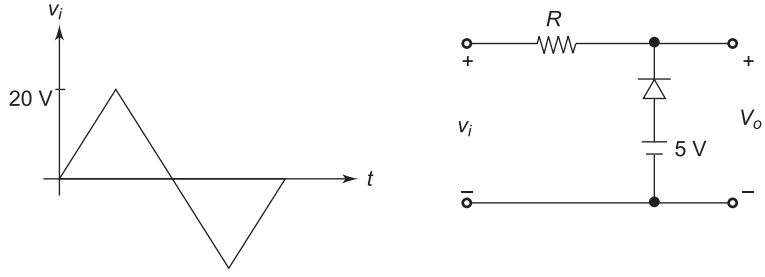


Fig. 2.64

13. Draw the output voltage of Fig. 2.40 when the direction of the diode is reversed.

14. In Fig. 2.41, reverse the direction of diode and bias voltage; sketch v_o.

➥ Multiple-Choice Questions

1. The cut in (knee voltage) of a germanium diode is
(a) 0.2 V (b) 0.3 V (c) 0.7 V (d) 0.8 V

2. Junction breakdown of a *PN*-junction diode occurs
(a) with forward bias (b) with reverse bias
(c) because of improper design (d) All of these

3. A zener diode
(a) is useful as an amplifier
(b) has a negative resistance
(c) has a high forward voltage
(d) has a sharp breakdown at low reverse voltage

4. Which of the following diodes is best suited as a switching diode for very high frequencies?
(a) PIN diode (b) Zener diode
(c) LED (d) Step recovery diode

5. The ripple factor of a half-wave rectifier is
(a) 0.482 (b) 0.812 (c) 1.21 (d) 1.11

6. The ac resistance for a Ge semiconductor diode having a forward bias of 200 mV and reverse situation current of 1 μA at
(a) 10.52 W (b) 11.86 W (c) 12.37 W (d) 15.49 W

7. Which of the following is a type of clipping circuit?
(a) Positive clippers (b) Negative clippers
(c) Biased clippers (d) All of these

8. Level shifter circuits are also known as
(a) clipper circuits (b) clamper circuits (c) diodes (d) transistors

9. In a zener diode, the value of I_{zm}, if power dissipation rating is 500 mW and zener voltage rating is 6.8 V,
(a) 78.6 mA (b) 80.2 mA (c) 82.5 mA (d) 73.5 mA

10. In an LED, light emission occurs because
(a) light gets reflected due to lens action (b) light falling on the diode gets amplified
(c) diode gets heated up (d) recombination of charge carriers takes place

ANSWERS

✦ Problems

1. 2.186 mA, 9.84 V, 5.46 V, –0.535 V

2. 0.2 mA, 2.86 mA, 2.66 mA

3. Input $\begin{cases}\text{High (1)} = 5\text{ V} \\ \text{Low (0)} = 0\text{ V}\end{cases}$; Output $\begin{cases}\text{High (1)} = 4.3\text{ V} \\ \text{Low (0)} = 0\text{ V}\end{cases}$

4. Input $\begin{cases}\text{High (1)} = 5\text{ V} \\ \text{Low (0)} = 0\text{ V}\end{cases}$; Output $\begin{cases}\text{High (1)} = 5\text{ V} \\ \text{Low (0)} = 0.7\text{ V}\end{cases}$

5. $V_z = 12$ V, $R_S = 16\ \Omega$, 3 W

6. 8 V, 0.15 V

9. 27.05 V

10. 16.57 V, 0.663 A

11. (a) 16 V (b) 15.5 V (c) 6.65 V

✦ Multiple-Choice Questions

1. (b) **2.** (b) **3.** (d) **4.** (a) **5.** (a) **6.** (b) **7.** (d) **8.** (b) **9.** (d) **10.** (d)

CHAPTER 3

Bipolar Junction Transistor (BJT) and Other Devices

GOALS AND OBJECTIVES

- Introduction and constructional details of bipolar junction transistor (BJT)
- Operational characteristics of BJT and configurations
- Description of silicon-controlled rectifier (SCR)
- Construction and characteristics of unijunction transistor (UJT)
- Description of phototransistor
- Types of dc biasing—self, voltage divider and voltage feedback bias

3.1 INTRODUCTION

In Chapter 2, we understood how a two-layer, one-junction semiconductor device acts as a diode, which conducts current in a particular direction and whose magnitude is controlled by the external circuit supply. Basically, a transistor is a combination of two back-to-back diodes, provided crystal continuity is maintained. Addition of another layer results in a three-layer two junctions device which has *npn* or *pnp* form and is called a transistor. With a terminal

connected to each layer, it acts as a *two-port device* (input/output ports) wherein one of the terminals is common between the two ports. Such a transistor is known as *Bipolar Junction Transistor* (BJT) which acts as a *current-controlled device* with the output current being controlled by the input current, such that the input-current waveform is replicated at the output. This is the amplifying action of a transistor commonly applied in various types of audio/video *amplifiers*. A BJT can also be made to act as a switch wherein the input current level controls the ON/OFF state of the output current. This mode of operation of a BJT finds wide applications in high-speed digital electronics.

3.2 BJT CONSTRUCTION AND OPERATION

The constructional details and order of dimension for the two types of BJT are shown in Fig. 3.1. In a *pnp* transistor, a thin *N*-type layer is sandwiched between two *P*-type layers, while in an *npn* transistor, a thin layer of *P*-type is sandwiched between two *N*-type layers.

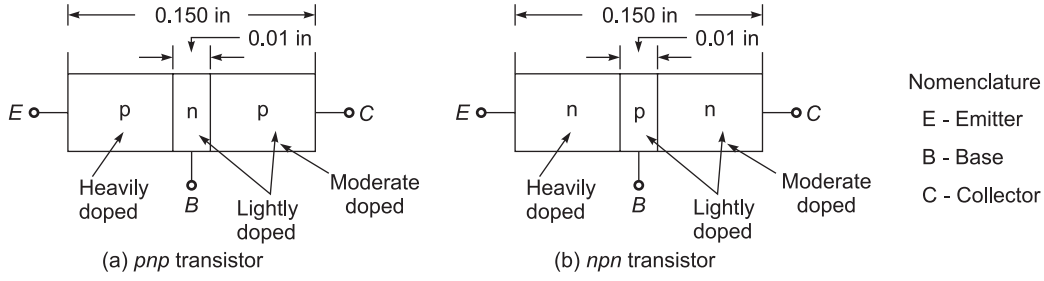


Fig. 3.1 BJT Transistor—construction and order of dimensions

✦ Biasing

There are two junctions: *EB* junction and *CB* junction. A transistor is like two diodes.

EB junction: Forward-biased diode

CB junction: Reverse-biased diode

Voltage biasing of transistors is shown in Fig. 3.2(a).

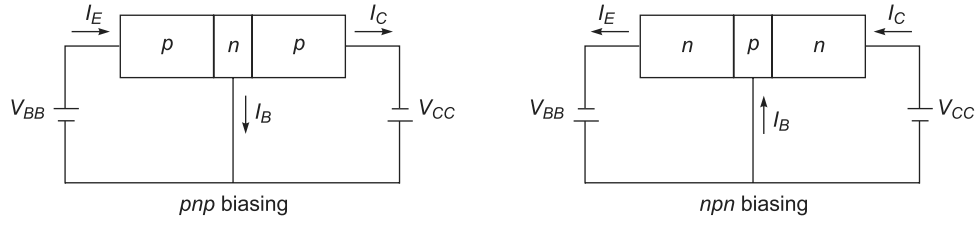


Fig. 3.2(a) Direction of current is conventional (hole current)

✦ Transistor Symbols

Fig. 3.2(b) Transistor symbols

The type of transistor can be recognised from the direction of arrow of *E* (emitter), Fig. 3.2(b).

✦ Operation

An enlarged cross-sectional view of a *pnp* transistor is drawn in Fig. 3.3(a); not to scale.

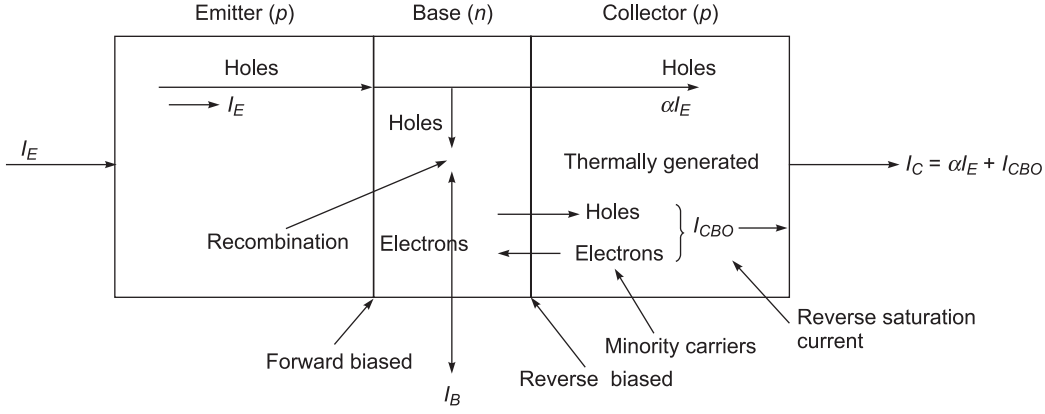


Fig. 3.3(a) Carrier flow in a pnp transistor

- Emitter holes (majority carriers) cross the forward-biased *EB* junction into the base. These constitute the emitter current I_E.
- The major portion of these holes cross over the reverse-biased CB junction constituting current αI_E; $\alpha = 0.98$ to 0.99. This is why the base width is kept very small.
- The reverse saturation current I_{CBO} flows across the *CB* junction. The controller current $I_C = \alpha I_E + I_{CBO}$ of order nA for Si.
- A small number of holes coming from the emitter recombine with electrons in the base. Recombination rate is small as electron concentration is very light in the base. To replenish the recombining electrons, a small electron current I_B flows out of the base. Note that direction of I_B is that of conventional current.

- Very few electrons from the base cross over to the emitter as the base is lightly doped. This current flow can be ignored [not shown in the figure]. As both electrons and holes participate in conduction, hence the name Bipolar Junction Transistor (BJT).

Carrier flow in an npn transistor is shown in Fig. 3.3(b).

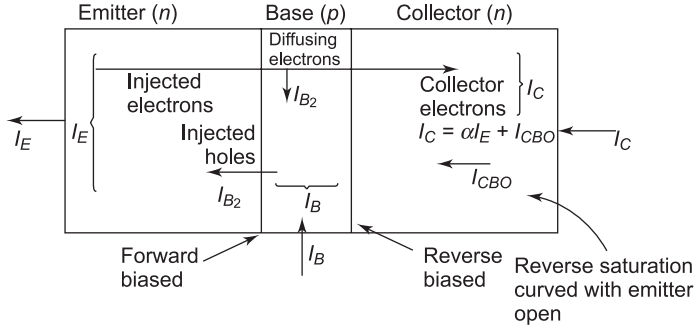


Fig. 3.3(b) Carrier flow in *npn* transistor

- Electrons in emitter region are injected into the base due to forward bias at EB junction.
- Most of the injected electrons reach the edge of CB junction before being recombined if the base is narrow.
- Electrons at the edge of CB junction will be swept into collector due to reverse bias at the junction.
- Emitter injection efficiency (γ) = $I_{En}/(I_{En} + I_{Ep})$.
- Base transport factor (α_T) = I_{Cn}/I_{Ep}.
- Common-base current gain (α) = $I_{Cn}/I_E = \gamma \alpha T < 1$.
- Terminal currents of BJT are as follows:
  - I_E (emitter current) = I_{En} (electron injection from E to B) + I_{Ep} (hole injection from B to E)
  - I_C (collector current) = I_{Cn} (electron drift) + I_{CBO} (CBJ reverse saturation current with emitter open)
  - I_B (base current) = I_{B1} (hole injection from B to E) + I_{B2} (recombination in base region)

The following fundamental dc relationships emerge.

$$I_C = \alpha I_E + I_{CBO} \tag{3.1}$$

As per KCL $I_E = I_C + I_B$ (3.2)

Substituting I_E from Eq. (3.2) in Eq. (3.1),

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

Or
$$I_C = \left(\frac{\alpha}{1-\alpha}\right) I_B + \left(\frac{1}{1-\alpha}\right) I_{CBO} \tag{3.3}$$

We define $\beta = \dfrac{\alpha}{1-\alpha}; \quad \alpha = \dfrac{\beta}{1+\beta}$

Then $I_C = \beta I_B + (1 + \beta) I_{CBO}$ (3.4)

I_{CBO}

If we make $I_E = 0$, open emitter

$I_{CBO} = I_{CO}$, reverse saturation current

as shown in Fig. 3.3(c).

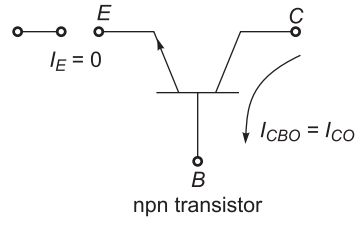


Fig. 3.3(c) Open emitter

Order of β

Let $\alpha = 0.988$

$$\beta = \frac{0.988}{1-0.988} = 82.3$$

Range of $\beta = 50 - 400$

Approximations

As I_{CBO} is the reverse saturation current (nA for Si), it can be neglected in Eq. (3.4) even though it is multiplied by $(1 + \beta)$

Thus, we can write

$I_C \approx \beta I_B$, we shall use this approximation most often (3.5)

Also, from Eq. (3.1),

$$I_C = \alpha I_E$$

or $I_C \approx I_E$ as $\alpha \approx 1$ (3.6)

Thus, we have from Eqs (3.5) and (3.6),

$$\alpha \approx \frac{I_C}{I_E}, \quad \beta \approx \frac{I_C}{I_B} \tag{3.7}$$

3.2.1 Early Effect

Refer Fig. 3.4. At the *CB* junction, there is a depletion region as the *CB* junction is reversed biased, while there is no such region at the *EB* junction as it is forward biased. As V_{CB} increases, the depletion region, and so effective base width, reduces. This base-width modulation is known as *early effect*. Its consequences are the following:

- Reduction of base width results in hole crossover to the collector faster, which mean α increases.
- Concentration of electrons in base increases, which crossover to the emitter thereby increasing I_E.
- At extremely large values of V_{CB}, the base width reduces to zero causing BJT breakdown. This phenomenon is known as *punch through*.

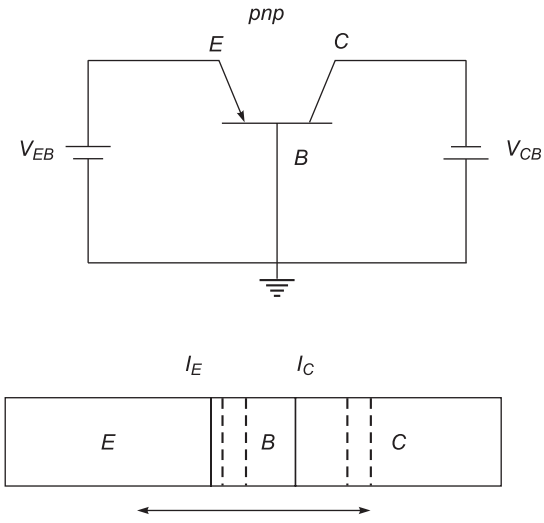


Fig. 3.4 Basewidth modulation due to V_{CB}

3.2.2 Ebers—Moll Model (Two-Diode Model)

It has been seen above that *BJT* has two junctions, *EB* and *CB*, which act as diodes. The *EB* diode is forward biased and the *CB* diode is reverse biased. It can, therefore, be modelled as two diodes shown in Fig. 3.5.

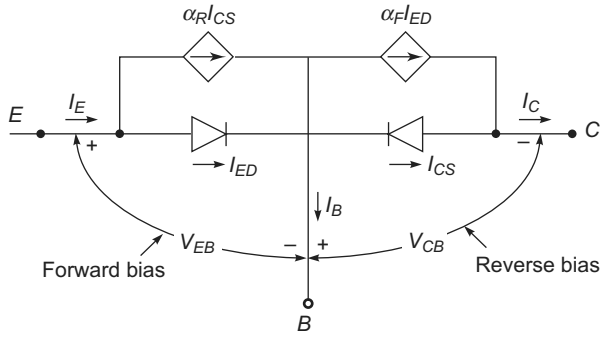


Fig. 3.5 Two-diode model

I_{CS} = reverse saturation current = I_{CBO} (in nA)

Range of α's

$$0.98 \le \alpha_F \le 0.998$$

$$0.4 \le \alpha_R \le 0.8$$

We can, therefore, ignore the dependent current $\alpha_R I_{CS}$. This will result in equations already presented.

3.3 BJT CONFIGURATIONS AND CHARACTERISTICS

As BJT is a three-terminal device, it can be connected in three possible configurations. For circuit operations, we require two terminals for input as well as output. So, one terminal of the transistor is grounded.

1. Common base
2. Common emitter
3. Common collector

To describe the behaviour of any configuration, two characteristics are required.

(i) Driving point or input
(ii) Output

3.3.1 Common-Base (CB) Configuration

The circuit is drawn in Fig. 3.6. Unless otherwise mentioned the transistor is npn.

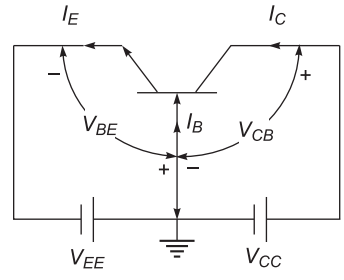
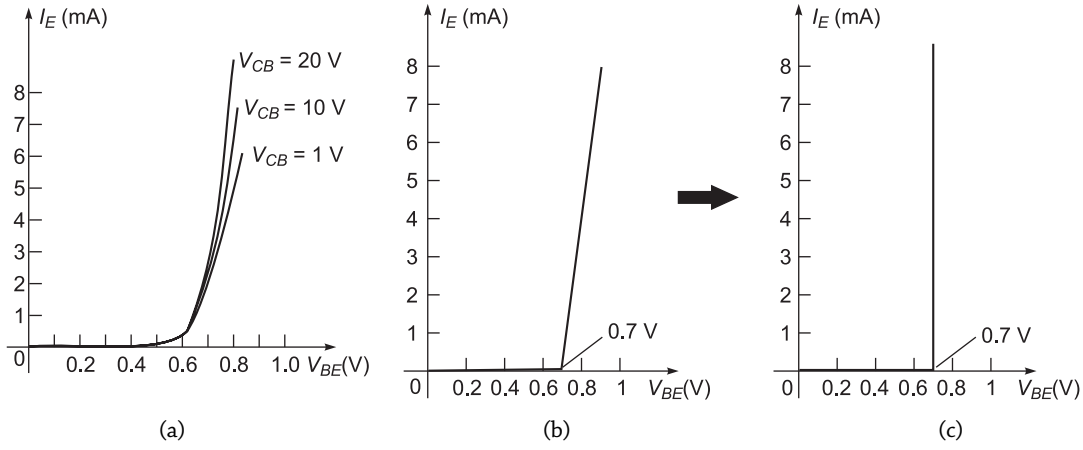


Fig. 3.6 Common base configuration (npn)

✦ Input (Driving-Point) Characteristic (I_E vs V_{BE})

The *EB* junction is a forward-biased diode. A typical characteristic to scale is presented Fig. 3.7. It is found that it is practically independent of V_{CB}. It can be approximated as a diode characteristic. Conduction begins for V_{BE} = 0.7 V as shown in Fig. 3.7(b).



(a) (b) (c)

Fig. 3.7 Driving point characteristic

The current I_E is controlled by adding a resistance in series with V_{EE}, while it will be assumed that

$$V_{BE} = 0.7 \text{ V}$$

At any value of I_E, once the transistor starts conducting, ON state is shown in Fig. 3.7(c).

✦ Output (Collector) Characteristics

These relate output current (I_C) with output voltage (V_{CB}) for varying values of input current (I_E). Typical output or collector characteristics are drawn in Fig. 3.8.

The characteristics can be divided into three regions.

✦ Active Region

The base-emitter junction is forward biased, while the collector-base junction is reverse biased. All the carriers that are injected into the emitter are swept away through the base to the collector. As a result, as already shown that

$$I_C = \alpha I_E \quad \text{or} \quad I_C \approx I_E \text{ as } \alpha \approx 1$$

α is the common-base forward current gain.

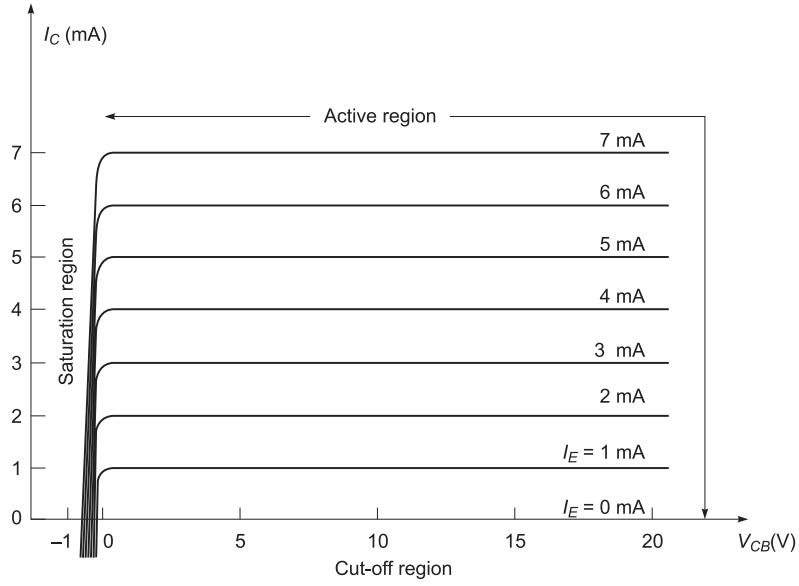


Fig. 3.8 Collector characteristics

This is easily seen from Fig. 3.8. Also, I_E is not affected by reverse bias V_{CB}. This is the linear region in which amplifying action of the circuit takes place.

✦ Cut-off Region

In this region, both base-emitter and collector-base junctions are both reverse biased. As a result, $I_E = 0$ and so $I_C = 0$.

✦ Saturation Region

In this region, both base-emitter and collector-base junctions are forward biased. This region is to the left of $V_{CB} = 0$. In this region, the collector current rises exponentially to the I_E value set by V_{BE} circuit as V_{CB} increases towards reverse bias.

✦ Amplifying Action

Current amplification ≈ 1. The input current is transferred by the transistor to output.

Voltage amplification will result from low driving-point resistance (10 – 100 Ω) [see Fig. 3.7(b)] and high output resistance.

The configuration is not used for amplification but serves certain special purposes.

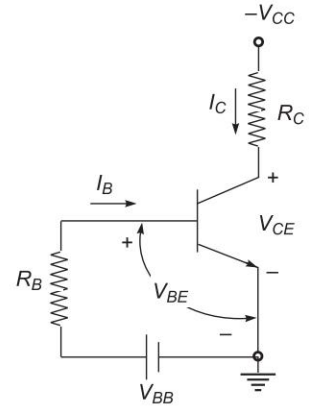


Fig. 3.9 CE configuration

3.3.2 Common-Emitter (CE) Configuration

It is the most frequently used configuration. Its circuit is drawn in Fig. 3.9.

✦ Input (Base) Characteristics

I_B vs V_{BE} for varying V_{CE} is drawn in Fig. 3.10(b).

✦ Output (Collector) Characteristics

I_C vs V_{CE} for various values of I_B is drawn in Fig. 3.10(a).

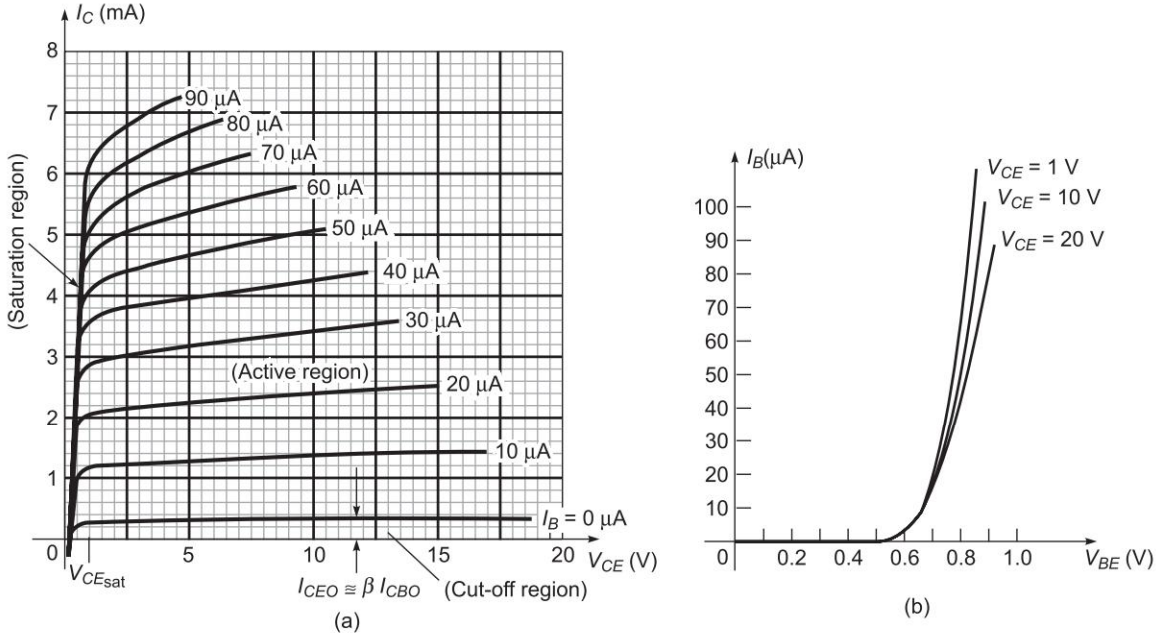


Fig. 3.10 (a) Collector characteristics (b) Base characteristics

It is seen from the base characteristics that I_B is practically independent of V_{BE}. The base-emitter junction goes 'on' at V_{BE} = 0.7 V and then stays there, while I_B is adjusted by the external resistance R_B.

✦ Active Region

Base-emitter is forward biased and collector-base is reverse biased. The collector characteristics are drawn in Fig. 3.10(a). It is seen that I_B is in μA and I_C in mA. These are related by β. The middle of this region is linear w.r.t. I_B and V_{CE}.

✦ Cut-off Region

It is below $I_B = 0$; the *EB* junction becomes reverse biased but the corresponding $I_C \neq 0$. From Eq. (3.4),

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

or $$I_C = (1 + \beta) I_{CBO} \approx \beta I_{CBO}; \text{ for } I_B = 0$$

$$\beta I_{CBO} = I_{CEO} \tag{3.8}$$

If I_{CBO} = 1 μA, resulting from $I_B = 0$, then

$$I_{CEO} = 250 \times 1 \times 10^{-2} = 0.25 \text{ mA at } \beta = 250$$

✦ Saturation Region

It is to the left of $V_{CE(sat)}$ = 0.2 V. In this region, the *EB* junction becomes forward biased and I_B no longer controls I_C.

The β: β_{dc} and β_{ac} lie in the middle region of Fig. 3.10(a), at any point corresponding to a certain V_{CE} say 10 V.

$$\beta = \beta_{dc} = \frac{I_C}{I_B}, \text{ large signal gain} \tag{3.9}$$

For variation about this point along V_{CE} = 10 V,

$$\beta_{ac} = \frac{\Delta I_C}{\Delta I_B} \tag{3.10}$$

Study on typical collector characteristics shows that

$$\beta_{ac} \approx \beta_{dc}$$

We will use the symbol β for β_{dc} or β_{ac} which will be obvious from the study under investigation.

We observe here that *β is the common-emitter forward-current gain*.

3.3.3 Common Collector (CC) Configuration

This connection is similar to common emitter, except that output is taken from the emitter. This causes, the output to be in phase with input (signal). It offers a high input resistance and low output resistance. It is, therefore, employed for impedance matching.

In this configuration, α_R factor will exist which shows the amplification of input at output.

$$\gamma = \frac{\Delta I_E}{\Delta I_B} \tag{3.11}$$

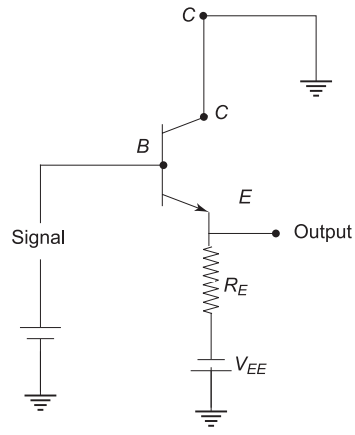


Fig. 3.11 Common collector configuration

✦ Limits of Operation

For each transistor, there are limits of operation which identify the region on its characteristics within which the signal exhibits least distortion. As shown in Fig. 3.12, the region is bounded by cut-off region, saturation region $I_{C(max)}$, maximum power dissipation $P_{D(max)} = V_{CE}\, I_C$, an inverse hyperbola and $V_{CE(max)}$.

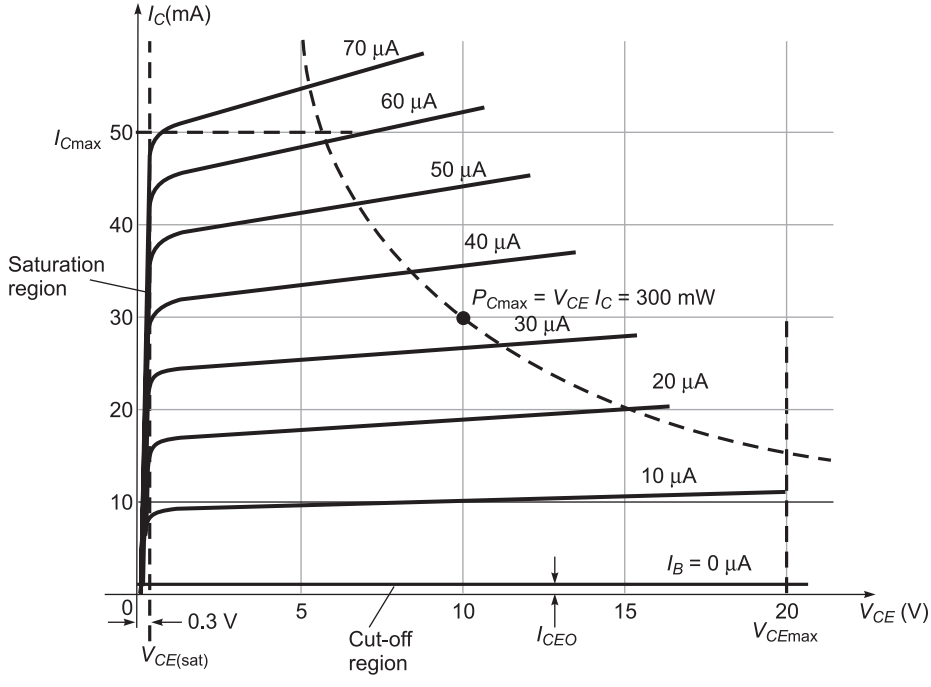


Fig. 3.12 Limits of operation

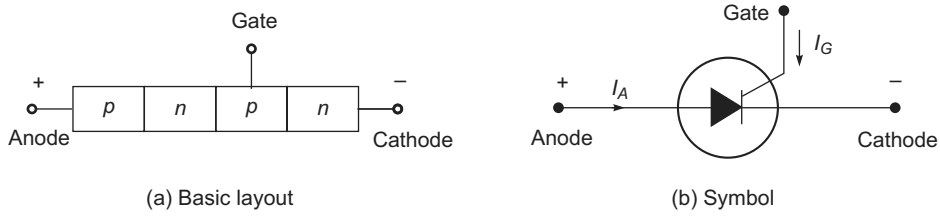
3.4 SILICON-CONTROLLED RECTIFIER (SCR)

It is a four-layer device which along with its associated circuitry has a very wide range of applications—rectifiers, regulated power supplies, dc to ac conversion (inverters), relay control, time-delay circuits and many more.

SCRs are now available to control power as high as 10 MW with individual rating of 2 kA and 1.8 kV. The frequency range has now been extended to 50 kHz, which are employed in high-frequency applications like induction heating and ultrasonic cleaning.

3.4.1 Basic Operation and Symbols

The material used for SCR is silicon because of high-temperature requirement of handling large current and power. Its four layers are arranged as *pnpn* shown in Fig. 3.13. The outer layers are connected to terminals to form **anode** (positive terminal) and **cathode** (negative terminal). The *P*-layer closer to the cathode is connected to the **gate** terminal. The SCR symbol is drawn in Fig. 3.13(b). It is similar to that of a diode, the difference being the indication of the gate terminal.



(a) Basic layout (b) Symbol

Fig. 3.13 Silicon Controlled Rectifier (SCR)

As a forward voltage is applied across the anode (+) and cathode (–), no conduction takes place as the middle *np*-junction is reverse biased. If a positive pulse is applied at the gate, such that a current of magnitude equal to more than I_G (turn-on) flows into the gate, the processes in the device cause it to go into conduction. The forward current (anode to cathode) is offered a resistance as low as 0.01 to 0.1 Ω. However, because of regenerative action, removing the gate current does not cause the device to turn off.

The dynamic reverse resistance of an SCR is as high as 100 kΩ or more.

3.4.2 Two-transistor Model

The cross-sectional view of an SCR with its four layers is drawn in Fig. 3.14(a). The middle *n* and *p* layers can be imagined to be subdivided into two halves, as shown by the dotted line. It is now immediately recognised that the device comprises one *PNP* and one *NPN* transistor. As there is an electrical continuity between the two halves of each of these layers, the base of

PNP is connected to the collector of *NPN*; and the collector of *PNP* is connected to the base of *NPN*, while the gate is connected to the base of *NPN*.

The corresponding two-transistor equivalent circuit is drawn in Fig. 3.14(b); observe that the connections are consistent with Fig. 3.14(a). This circuit will now be used to explain the action of the gate pulse I_G.

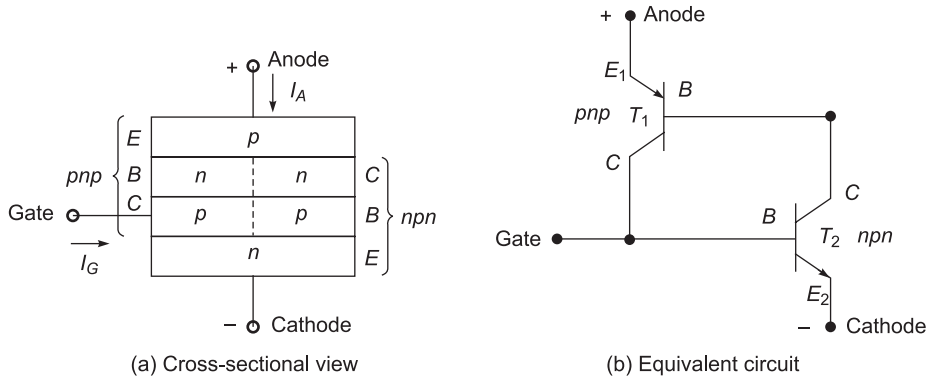


Fig. 3.14 Two-transistor model of SCR

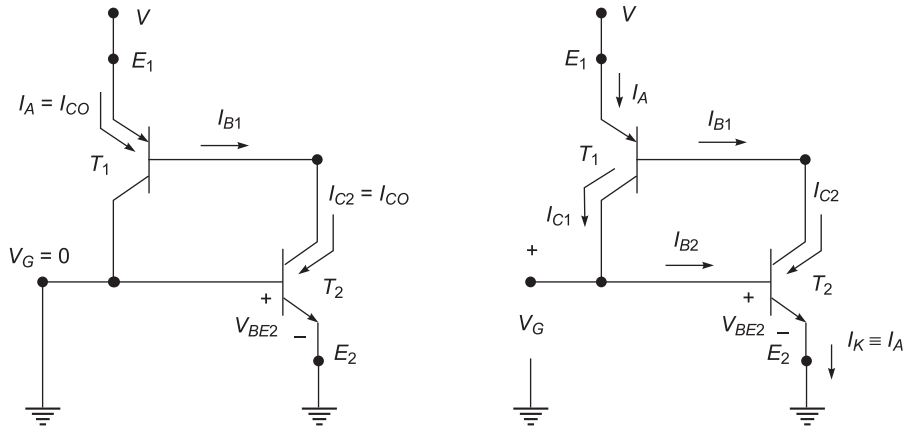
3.4.3 Switching Action

Let a positive voltage V be applied to the anode (E_1), and the cathode (E_2) and gate (G) be both grounded as shown in Fig. 3.15(a). As $V_G = V_{BE2} = 0$, the transistor T_2 is in 'off' state. It means that *CB*-junction of T_2, through *EB*-junction of T_1, is reverse biased. Therefore, $I_{B1} = I_{CO}$ (minority carrier current) is too small to 'turn-on' T_1. Thus both T_1 and T_2 are 'off' and so anode current

$$I_A = I_{B1} = I_{CO}$$

is of negligible order. It means that SCR is in 'turn-off' state, that is the switch between anode (E_1) and cathode (E_2) is open.

Now, let a voltage $+V_G$ be applied at the gate as shown in Fig. 3.15(b). As $V_{BE2} = V_G$, on making V_G sufficiently large, I_{B2} will cause T_2 to turn on and the collector current I_{C2} becomes large. As $I_{B1} = I_{C2}$, T_2 turns on causing a large collector current I_{C1} ($I_A = I_{C1}$) to flow. This in turn, increases I_{B2} causing a regenerative action to set in (this is indeed a positive internal feedback). The result is that the SCR is turned on, that is, the switch between the anode (E_1) and cathode (E_2) is closed (*turn-on*). The current I_A must be limited by the external circuit, say a series resistance between the source and E_1.



(a) With $V_G = 0$ (b) With voltage V_G applied

Fig. 3.15 Switching action of a two-transistor SCR

The turn-on time of an SCR is typically 0.1 to 1 μs. However, for high-power devices in the range of 100–400 A, turn-on time may be 10–25 μs.

✦ Turn-off

As has been shown earlier, when the SCR is in conduction mode, the gate is ineffective in turning it off. The turn-off mechanism is called *commutation* and it can be achieved in two ways explained below.

❑ ***Natural Commutation*** When the source that feeds the current to anode of SCR is such that it naturally passes through zero, the SCR turns off at the current zero. This is the case when the SCR is fed from the ac source. In this situation, the commutation is also known as *line commutation*.

❑ ***Forced Commutation*** In this method of commutation, the current through the SCR is forced to become zero by passing a current through it in opposite direction from an independent circuit. A variety of SCR turn-off circuits are available in books and in manufacturer's manuals.

One basic turn-off circuit which illustrates the principle is drawn in Fig. 3.16. A transistor and dc battery source in series are connected to the SCR. When the SCR is in conduction mode (on), $I_B = 0$ and when the transistor is off, it is almost an open circuit. To turn off the SCR, a positive I_B pulse of magnitude large enough to drive the transistor into saturation is applied at the transistor base. The transistor acts almost like a short circuit. This causes flow of very large I_{off} through the SCR in the opposite direction to its conduction current. The total SCR current reduces to zero in a very short time causing it to turn off. The transistor has to withstand a large current but for a very short time.

Turn-off time of an SCR is typically 5–30 μs.

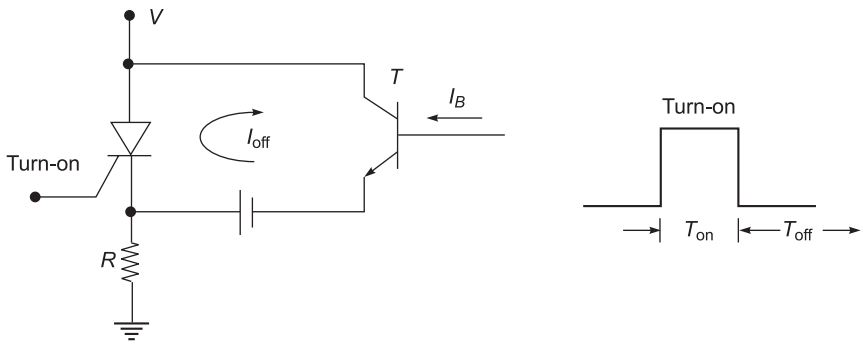


Fig. 3.16 Turn-off circuit using SCR

3.4.4 SCR Characteristics

The symbol and *I-V* characteristics of an SCR are given in Fig. 3.17. Various voltages and currents which provide important information for SCR applications are described below.

1. *Forward Breakover Voltage* V_F (BR) is the voltage at which for a given I_G, the SCR enters into conduction mode. As is seen from Fig. 3.17, that this voltage reduces as I_G increases. V_F(BR) has dependence on the circuit connection between *G* and *K* terminals.
2. *Holding current* I_H is the value of the current below which SCR switches from conduction state to forward blocking regions of specified conditions.
3. Forward and reverse blocking regions are those regions in which the SCR is open circuited and no current flows from anode to cathode.
4. Reverse breakdown voltage corresponds to Zener or avalanche region of a diode.

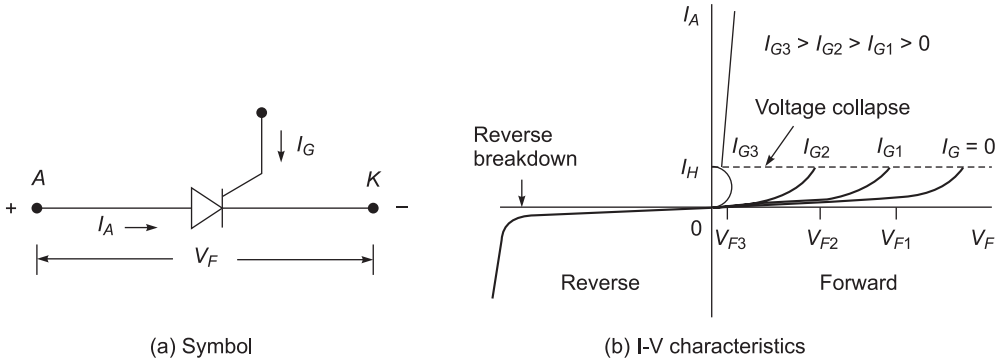


Fig. 3.17 SCR characteristics

Other specifications of interest in designing SCR circuitry and systems are the grid voltage, current and power, their maximum permitted values symbolised as

$$V_{GFM}, I_{GFM}, P_{GFM}$$

The manufacturer's manual, apart from indicating these, identifies the preferred region of operation.

3.4.5 Applications of SCR

The range of applications of SCRs given in the beginning of this section is by no means exhaustive. As for applications in power and drives, these form the subject matter of a separate course for which several excellent books are available. We shall give here a single application for illustrative purpose.

✦ Variable Resistance Phase Control

On triggering, an SCR permits flow of only forward current but blocks the current in reverse direction. This action is the same as that of a diode. On application of alternating voltage, it causes rectified ac to flow but it needs to be triggerd for each positive half cycle of ac. It then produces constant dc (average value) current through load and dc voltage across load. Adjusting the triggering time on positive half cycle of ac voltage would yield variable dc output. This method is known as phase control.

A variable-resistance phase-control circuit is provided in Fig. 3.18(a). The SCR gate current is controlled through R and the variable R_1. Let R_G be adjusted to high value so that even at the peak value v_i (positive), $I_G < I_{G(\text{turn-on})}$ and no conduction takes place. As R_1 is reduced, I_G rises to turn-on value at a particular angle (time) of v_i. The conduction then begins and continues till v_i reaches zero (180°). Varying R_1 allows the adjust of SCR firing angle from 0° to 90° as shown in Fig. 3.17(b).

At R_1, corresponding to the firing angle of 90°, $v_i = v_i$ (max). If R_1 is adjusted for firing at α [see Fig. 3.17(b)], the firing will take place at angle $\alpha < 90°$ but not at angle $\beta = (180° - \alpha) > 90°$ as the angle α is reached earlier in time on the v_i-wave. So the operation of this circuit is known as *half-wave, variable-resistance phase control*.

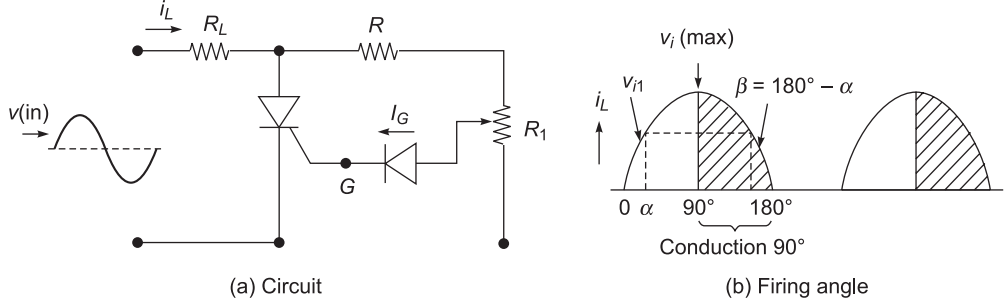


Fig. 3.18 Variable resistance phase control

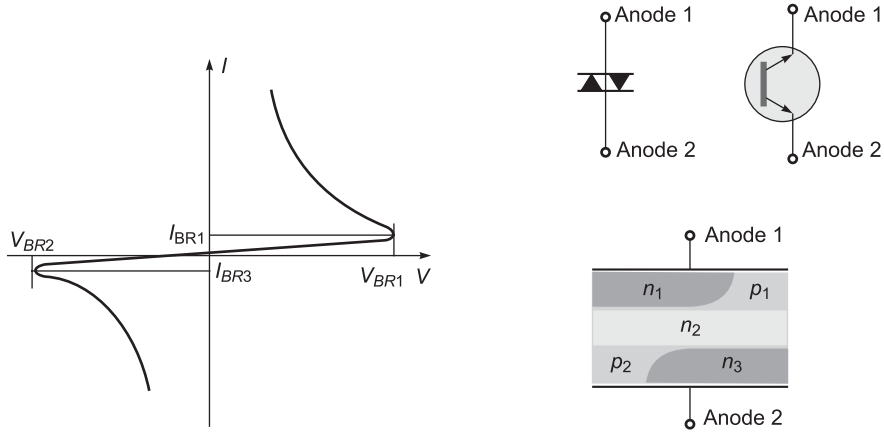
Thus, i_L(dc) can be adjusted to the maximum value at 0° to the minimum value at 90°.

It may be noted that a diode is provided in the firing circuit to prevent the flow of reverse gate current.

3.4.6 DIAC

The DIAC is basically a two-terminal, parallel-inverse combination of semiconductor layers that permits triggering in either direction. The characteristics of the device, presented in Fig. 3.19(a), clearly demonstrate that there is a breakover voltage in either direction. This possibility of an on condition in either direction can be used to its fullest advantage in ac applications.

The basic arrangement of the semiconductor layers of the DIAC is shown in Fig. 3.19(b), along with its graphical symbol. Note that neither terminal is referred to as the cathode. Instead, there is an anode 1 (or electrode 1) and an anode 2 (or electrode 2). When the anode 1 is positive with respect to the anode 2, the semiconductor layers of particular interest are $p_1n_2p_2$ and n_3. For the anode 2 to be positive with respect to the anode 1, the applicable layers are $p_2n_2p_1$ and n_1.



(a) Characteristic (b) Structure and Symbol

Fig. 3.19 DIAC

3.4.7 TRIAC

The TRIAC is fundamentally a DIAC with a gate terminal for controlling the turn-on conditions of the bilateral device in either direction. In other words, for either direction, the gate current can control the action of the device in a manner very similar to that demonstrated for an SCR. The characteristics, however, of the TRIAC in the first and third quadrants are somewhat different from those of the DIAC, as shown in Fig. 3.20. Note the holding current in each direction not present in the characteristics of the DIAC.

The graphical symbol for the device and the distribution of the semiconductor layers are provided in Fig. 3.20 with photographs of the device. For each possible direction of conduction, there is a combination of semiconductor layers whose state will be controlled by the signal applied to the gate terminal.

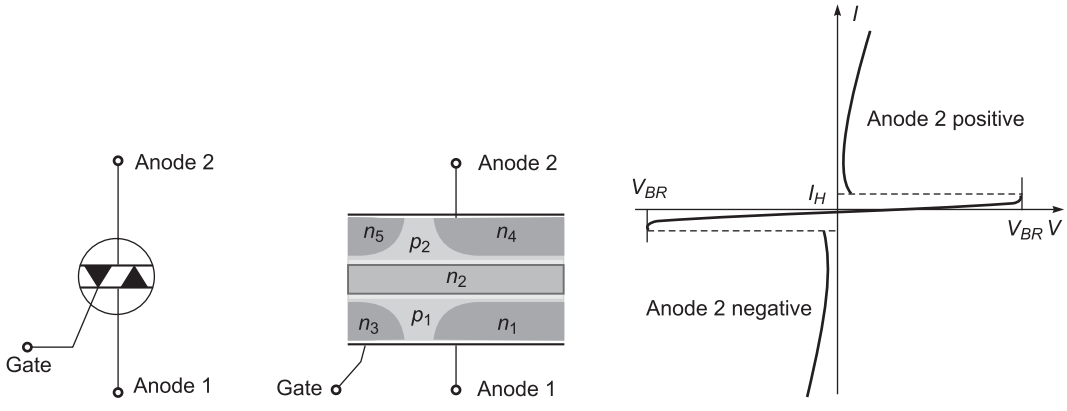


Fig. 3.20 TRIAC

3.5 UNIJUNCTION TRANSISTOR (UJT)

It is a low-cost device, which, because of its excellent characteristic and low power loss (in it), is commonly employed in a wide variety of applications like oscillators, trigger circuits, sawtooth generators, phase control, timing circuits, bistable networks and regulated supplies.

A UJT is a three-terminal device (basic construction shown in Fig. 3.21(a)) formed from a lightly doped slab on *N*-type material having high resistance characteristics. Two base contacts are made at each end of one side of the slab, while an aluminium rod is fused on the other side to form a single *pn*-junction, and hence the name unijunction. The rod is located closer to the base terminal 2 which is made positive with respect to the base terminal 1 by V_{BB}. The symbol and biasing of a UJT is shown in Fig. 3.21(b).

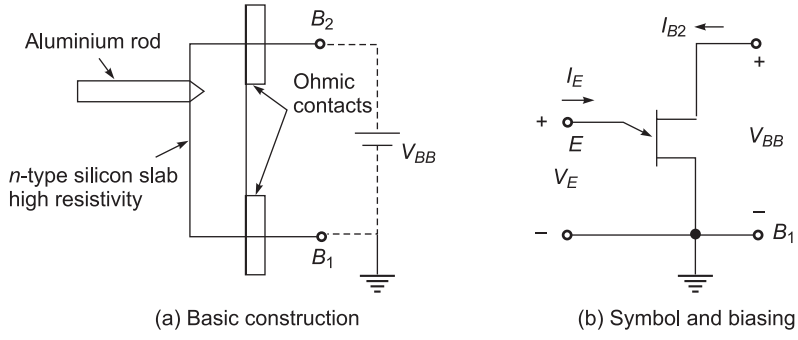


Fig. 3.21 Unijunction transistor (UJT)

The circuit equivalent to the UJT is drawn in Fig. 3.22. Here the input diode represents the *pn*-junction operation; R_{B2} is a fixed resistance and R_{B1} is a variable resistance, which reduces with increase in emitter current I_E. The range of variation of R_{B1} may be 50 kΩ to 50 Ω as I_E increases from 0 to 50 mA. The interbase resistance is defined as

$$R_{BB} = (R_{B1} + R_{B2})\Big|_{I_E = 0} \tag{3.12}$$

The range of this resistance is typically 4–10 kΩ. At $I_E = 0$, the voltage

$$V(R_{B1}) = \frac{R_{B1}V_{BB}}{R_{B1} + R_{B2}}\Bigg|_{I_E = 0} = \eta V_{BB} \tag{3.13}$$

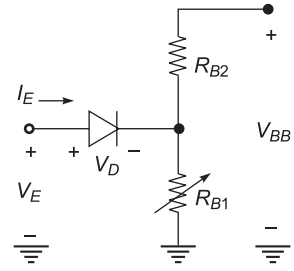


Fig. 3.22 Circuit equivalent of UJT

where η, called *intrinsic stand-off ratio*, is given as

$$\eta = \frac{R_{B1}}{R_{B1} + R_{B2}}\Bigg|_{I_E = 0} \tag{3.14}$$

This ratio (η) is controlled by the location of the aluminium rod (Fig. 3.21). The emitter firing potential is given by

$$V_P = \eta V_{BB} + V_D \tag{3.15}$$

The static characteristic of a typical UJT is shown in Fig. 3.23 for V_{BB} = 10 V, wherein the various regions are indicated. As V_E crosses V_P, the emitter fires and holes are injected into the slab from the *P*-type aluminium rod. This causes increase in the hole content of the *N*-type slab with consequent increase in the number of free electrons in it, and so, increased conductivity. Thus, V_E drops off while I_E increases. This is the negative resistance region of the *UJT*, which passes through the valley point (I_V, V_V), while the devices get saturated.

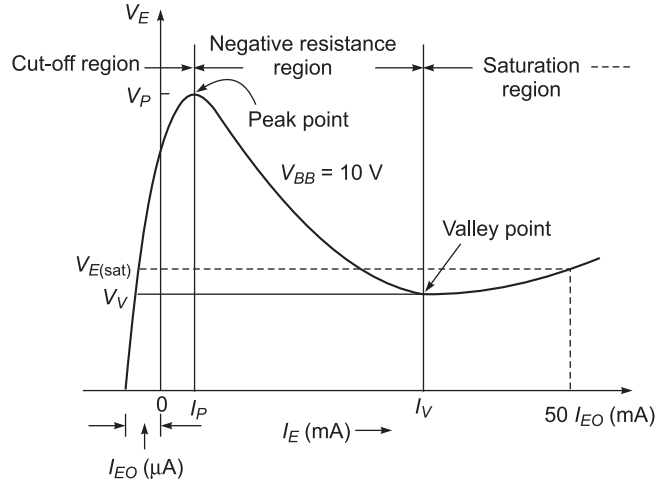


Fig. 3.23 UJT static emitter characteristic

Three important parameters of a UJT are I_P, V_V and I_V. As V_{BB} increases, it causes V_P to increase [as per Eq. 3.15].

3.6 PHOTOTRANSISTOR

Phototransistor is a photosensitive device which converts light energy into electrical energy. The device is designed with large base and collector regions as compared with ordinary BJTs with photosensitive material being used for base. Figure 3.24 given below shows the two types of transistors: (a) homojunction transistor, and (b) heterojunction transistor.

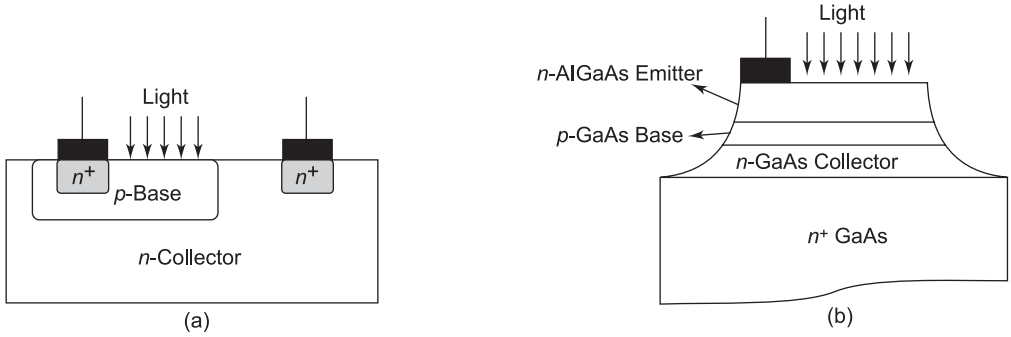


Fig. 3.24 (a) Homojunction phototransistor (b) Heterojunction phototransistor

The symbol of npn phototransistor and physical phototransistor are shown in Fig. 3.25(a) and Fig. 3.25(b), respectively. The lines pointing towards the base represent the light input to base for the required voltage generation. The metal casing is usually used to improve the exposure of base of transistor to light.

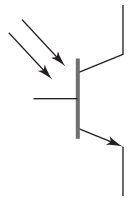


Fig. 3.25(a) Symbol of npn phototransistor

Fig. 3.25(b) Phototransistor

The biasing circuit and output characteristics of phototransistor are shown in Fig. 3.25(c) and Fig. 3.25(d), respectively. At the initial stage, when no exposure of light is there, a minute reverse saturation current flows owing to the presence of minority charge carriers. When exposed to light, transistor starts conducting through reverse biasing. The response of transistor is different for different intensities of light. Here, the incident photons have the energy level $h\nu$, where h is Plank's constant and ν is frequency of travelling wave. Since wavelength and frequency are interrelated terms, the wavelength of light used plays a crucial

role in deciding the material of the device. The number of free electrons generated in each material is proportional to the intensity of incident light. The material used for construction of phototransistor includes silicon, germanium with non-identical material like gallium arsenide on either side of p-n junction. The different applications of phototransistor include optocoupling, switching and controlling, optical isolation, optical sensor, etc.

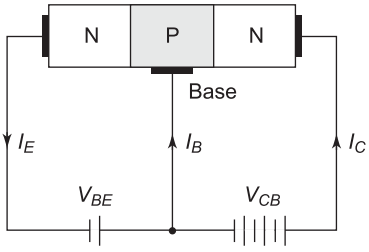


Fig. 3.25(c) Biasing circuit

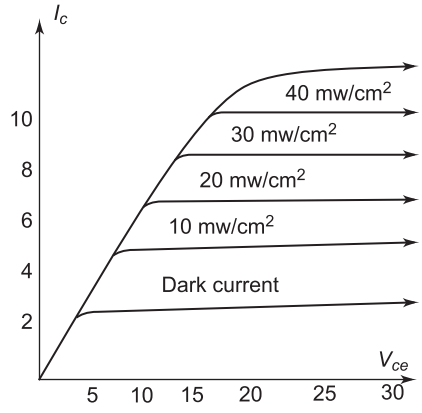


Fig. 3.25(d) Output characteristics

3.7 DC BIASING

It has been seen above that for BJT operation, certain dc levels must be set by dc sources (batteries). This process is called biasing the BJT. The biasing locates an operating point, also called *quiescent point* (*Q*), on the characteristics, about which signal-caused variation takes place [ac input causing ac output suitably modified (amplified)]; dc biasing and ac analysing can be carried out separately (results superimposed if necessary). In order to isolate the ac signals from dc sources, isolating capacitors are used, which act as short circuits for ac signals. From the signal point of view, these are *coupling capacitors*. We shall now proceed to analyse and design the dc biasing which determines the location of the *Q*-point appropriately.

An important requirement of the biasing circuit is its insensitivity to BJT's parameter changes due to temperature, i.e. a stable *Q*-point.

✦ Fixed Bias

The fixed bias circuit is drawn in Fig. 3.26 with a single battery source. The collector characteristics are drawn in Fig. 3.27. The *Q*-point is to be located on the characteristic

$I_B = I_{BQ}$ as shown in Fig. 3.27.

KVL equation for *BE* loop

$$V_{CC} - R_B I_{BQ} - 0.7 \text{ V} = 0;\ V_{BE} = 0.7 \text{ V} \qquad (3.16)$$

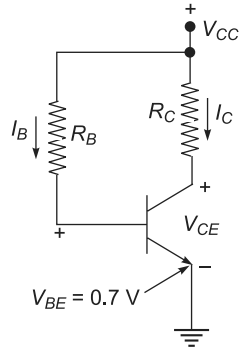


Fig. 3.26 Fixed bias

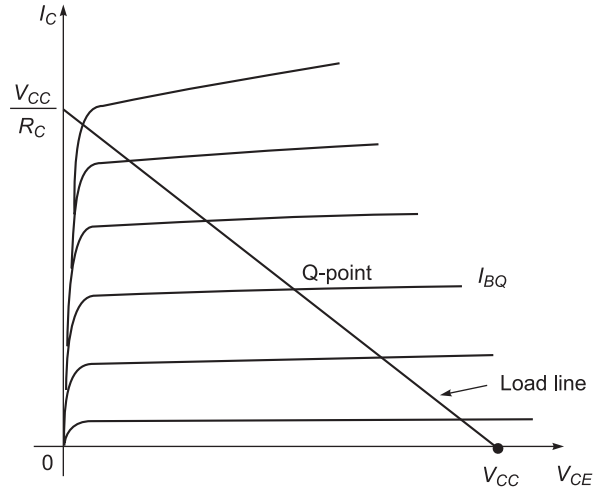


Fig. 3.27 Location of *Q*-point

$$I_{BQ} = \frac{V_{CC} - 0.7}{R_B} \tag{3.17}$$

✦ Load Line

KVL equation for the collector loop is

$$V_{CC} - R_C I_C - V_{CE} = 0 \tag{3.18}$$

or $$I_C = \frac{V_{CC} - V_{CE}}{R_C}$$

It is a straight line in $I_C - V_{CE}$ coordinates.

Along V_{CE} axis. $I_C = 0$; intercept on V_{CE} – axis is

$$V_{CE} = V_{CC}$$

Along I_C axis, $V_{CE} = 0$, intercept on I_C – axis

$$I_C = \frac{V_{CC}}{R_C}$$

The line joining these two intercepts is the *load line* drawn on the collector characteristics. It intersects the characteristics for I_{BQ} at the operating point Q.

By adjusting I_B (by R_B) the Q-point can be located anywhere. If I_B is increased, the Q-point moves up along the load line. If R_C is increased, the load line slope, starting from V_{CC}, decreases and the Q-point moves on I_{BQ} characteristic to the left side.

The fixed bias cannot counter the thermal effect on the BJT characteristics. As β rises with temperature, collector characteristics shift upwards as $I_C \approx \beta I_B$. Correspondingly, the Q-point moves up on the load line.

This deteriorates the BJT's performance.

3.7.1 Bias Stabilisation by Emitter Resistance (Self-Bias)

The fixed-bias circuit of Fig. 3.26 is modified to include a small resistance R_E in the emitter by as shown in Fig. 3.28. The circuit also shows coupling capacitors (C_c) which are open circuit for dc.

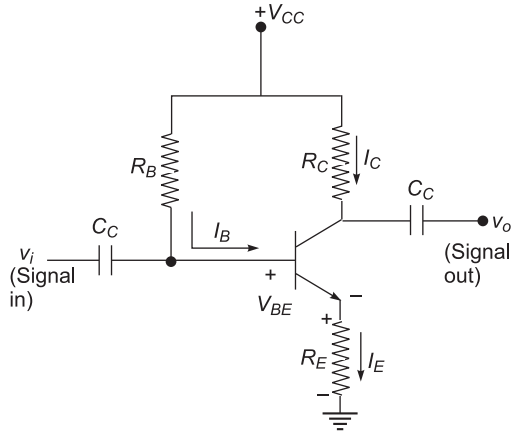


Fig. 3.28 Self-bias circuit

The emitter includes voltage drop $R_E I_E$, which acts as negative feedback to stabilize the bias (Q-point).

$$I_E = I_B + I_C = I_B + \beta I_B = (1 + \beta) I_B$$

Base-emitter loop

$$V_{CC} - R_B I_B - 0.7 - [(1 + \beta)R_E] I_B = 0 \tag{3.19}$$

$$I_B = \frac{V_{CC} - 0.7}{R_B + (1 + \beta) R_E} \tag{3.20}$$

Any increase in β reduces I_B slightly as $R_B >> R_E$.

Corresponding $I_C = \beta I_B$ increases to compensate the shift in Q-point. So there is no significant shift in Q-point, which has been stabilised by inclusion of R_E.

3.7.2 Voltage-Divider Bias

The voltage-divider bias circuit is drawn in Fig. 3.29(a) wherein bias voltage is applied to the base from the voltage-divider resistances R_1, R_2 and self-bias emitter resistance is also included.

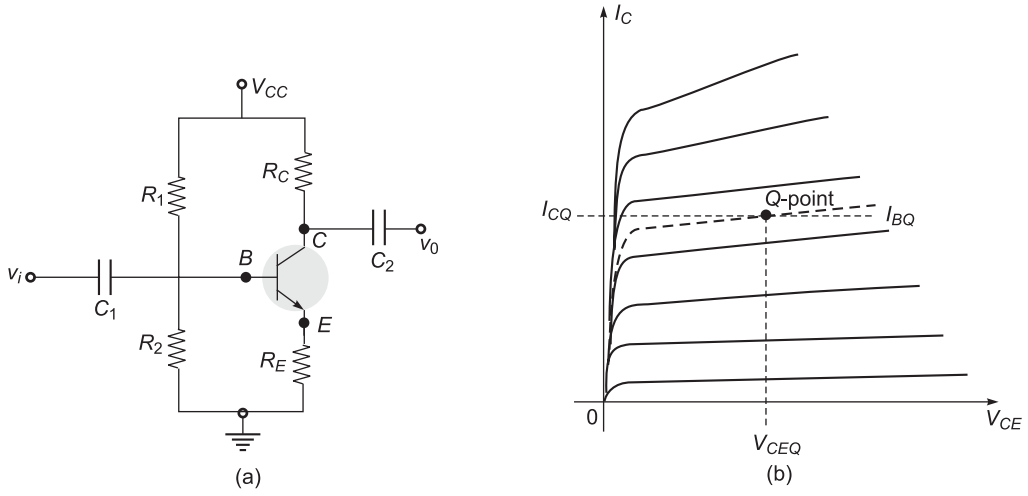


Fig. 3.29 (a) Voltage-divider bias circuit (b) Q-point remains almost unchanged

It can be shown by exact analysis that the sensitivity of the *Q*-point to changes in β is quite small by proper design of circuit parameters. If β changes, the level of I_{BQ} will change because of the negative feedback effect of R_E but the collector characteristics also change accordingly. As a result, the operating point defined by I_{CQ} and V_{CEQ} remains unchanged.

3.7.3 Analysis Equations

The Thevenin equivalent seen at *B*

$$V_{\text{th}} = \left(\frac{R_2}{R_1 + R_2}\right) V_{CC} \tag{3.21}$$

Thevenin resistance

$$R = R_1 \| R_2 \tag{3.22}$$

The biasing circuit is redrawn in Fig. 3.30.

As $$I_E = (1 + \beta)\, I_B \tag{3.23}$$

We find from the *BE* loop

$$I_B = \frac{V_{\text{th}} - 0.7}{R_{\text{th}} + (1 + \beta) R_E} \tag{3.24}$$

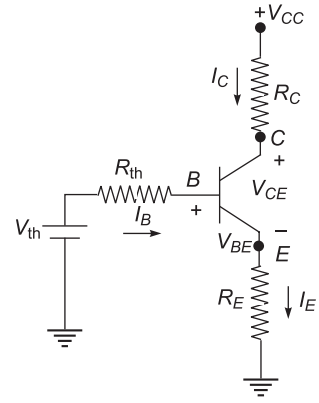


Fig. 3.30 Biasing circuit

From KVL for *CE* loop, we have

$$V_{CE} = V_{CC} - (R_C + R_E)\, I_C;\ I_E \approx I_C \tag{3.25}$$

These equations are sufficient to determine all voltages and currents after the choice of V_{CC} and I_{CQ} and V_{CEQ} from the collector characteristics by locating a *Q*-point.

3.7.4 Voltage Feedback Bias

The emitter self-bias circuit of Fig. 3.28 is modified by voltage feedback to base from V_C the collector voltage instead of V_{CC} as drawn in Fig. 3.30.

Writing the KVL equations for *BE* and *CE* loops and recognising $I_B = I_E - I_C$, we can obtain the following result $(1 + \beta \approx \beta)$

$$I_B = \frac{(V_{CC} - 0.7) = V'}{R_B + \beta(R_C + R_E)} \tag{3.26}$$

It is further found that

$$I_{CQ} = \beta I_B = \frac{\beta V'}{R_B + \beta R'}\,;\; R' = R_C + R_E \tag{3.27}$$

As $\beta R' >> R_B$,

$$I_{CQ} = \frac{V'}{R'}\text{, Independent of } \beta \tag{3.28}$$

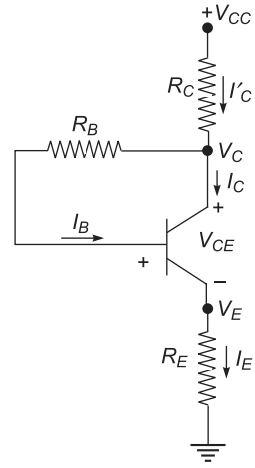


Fig. 3.31 Voltage feedback bias

In fact, Eq. (3.26) is the general form of the I_B equation for the biasing circuits presented. Voltage feedback bias is somewhat better than emitter self-bias because of additional voltage feedback. *But the voltage-divider bias is the best of all the schemes of BJT bias and is universally adopted*.

***Remark*:** In order that emitter resistance R_E does not affect the ac performance, it should be shunted by a capacitor called *bypass capacitor*.

***I_C*(sat):** To gauge how far the *Q*-point is from saturation, we need to find I_C(sat). BJT saturates at $V_{CE} = 0.2$ V, see Fig. 3.10. It can be assumed that in saturation $V_{CE} = 0$, with $R_E = 0$,

$$I_C(\text{sat}) = \frac{V_{CC}}{R_C} \tag{3.29}$$

EXAMPLE 3.1 In a BJT, the emitter current is 8 mA and $I_B = I_C/100$. Determine I_C and I_B.

Solution

$$I_E = I_C + I_B$$

$$I_E = I_C + \frac{1}{100} I_C = \frac{101}{100} I_C$$

$$I_C = 8 \times \frac{1}{101} = 7.92 \text{ mA}$$

$$I_B = \frac{1}{100} I_C = 0.0792 \text{ mA}$$

EXAMPLE 3.2 Given $\alpha_{dc} = 0.997$, (a) determine I_C if $I_E = 5$ mA, (b) determine α_{dc} if $I_E = 2.8$ mA and $I_B = 20$ μs, and (c) find I_E if $I_B = 40$ μA and $\alpha_{dc} = 0.98$.

Solution

(a) $I_C = \alpha_{dc} I_E$; I_{CBO} neglected

$$I_C = 0.997 \times 5 = 4.985 \text{ mA}$$

(b) $$I_E = I_C + I_B$$

$$(1 - \alpha_{dc}) I_E = I_B$$

$$1 - \alpha_{dc} = \frac{20 \times 10^{-3}}{2.8}$$

$$\alpha_{dc} = 0.993$$

(c) $$I_B = (1 - \alpha_{dc}) I_E$$

$$I_E = \frac{40 \times 10^{-3}}{(1 - 0.98)} = 2 \text{ mA}$$

EXAMPLE 3.3 Refer characteristics of Fig. 3.10.

(a) Find I_C corresponding to $V_{BE} = +750$ mV, $V_{CE} = +5$ V.

(b) Find V_{CE} and V_{BE} corresponding to $I_C = 3$ mA and $I_B = 30$ μA.

Solution

From Fig. 3.10(b), at $V_{BE} = 0.75$ V, $I_B = 40$ μA

From Fig. 3.10(a) at $I_B = 40$ μA, $V_{CE} = 5$ V, $I_C = 4$ mA

EXAMPLE 3.4 Using the collector characteristics of Fig. 3.10(a), determine

(a) β_{dc} at $I_B = 80$ μA and $V_{CE} = 5$ V

(b) Repeat part (a) at $I_B = 30$ mA and $V_{CE} = 10$ V

Solution

(a) We find from Fig. 3.10(a) at $I_B = 80$ mA, $V_{CE} = 5$ A

$$I_C = 6.7 \text{ mA}$$

$$\beta_{dc} = \frac{I_C}{I_B} = \frac{6.7}{80 \times 10^{-3}} = 83.75$$

(b) $$I_B = 30 \text{ μA}. \; V_{CE} = 10 \text{ V}$$

$$\beta_{dc} = \frac{3.4}{30 \times 10^{-3}} = 113.3$$

***Remark*:** $\beta_{dc} = 113.3$ is the proper value for use as it lies in the linear part of active region.

EXAMPLE 3.5 On the common-emitter collector characteristics of Fig. 3.10(a), locate the operating point at V_{CE} = 10 V and I_C = 2 mA.

(a) Find the value of α for this operating point.

(b) Find the value of β for this operating point.

(c) At V_{CE} = 10 V read the corresponding V_{CEO}.

(d) Calculate the corresponding value of V_{CBO}.

Solution

(a) The operating point is located at Q in Fig. 3.10(a)

$$I_C = 2 \text{ mA}, I_B = 16 \text{ μA}$$

$$I_E = I_C + I_B$$

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

$$\left(\frac{1}{\alpha} - 1\right) I_C = I_B \text{ or } \left(\frac{1}{\alpha} - 1\right) = \frac{I_B}{I_C} = \frac{16 \times 10^{-3}}{2}$$

$$\alpha = 0.992$$

(b) $$\beta = \frac{I_C}{I_B} = \frac{2}{16 \times 10^{-3}} = 125$$

(c) At $V_{CE} = 10 \text{ V}, I_{CEO} = 0.4 \text{ mA}$

(d) $$I_{CEO} \approx \beta . I_{CBO}$$

$$I_{CBO} = \frac{0.4}{125} \times 10^3 = 3.2 \text{ μA}$$

EXAMPLE 3.6 For the transistor circuit of Fig. 3.32, V_{CC} = 10 V. β = 100.

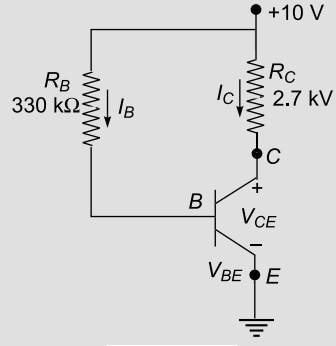


Fig. 3.32

Calculate (a) I_B, (b) I_C, (c) V_{CE}, and (d) V_{CB}.

Solution

Assume active mode.

BE conducts with $V_{BE} = 0.7$ (remains constant)

(a) $I_B = \dfrac{10 - 0.7}{330} = 0.028$ mA

(b) $I_C = \beta I_B = 2.8$ mA

(c) $V_{CE} = 10 - 2.7 \times 2.8 = 2.44$ V > 0.2 V = $V_{CE(sat)}$, active mode

(d) $V_{CB} = V_{CE} - V_{BE} = 2.44 - 0.7 = +1.74$ V, reverse bias; active mode

Remark: $V_{CB} > 0.5$ V (cut-in voltage) for active mode

Correspondingly,

$V_{CE} = V_{CB} + V_{BE} = -0.5 + 0.7 = 0.2$ V = $V_{CE(sat)}$

EXAMPLE 3.7

For the fixed-bias configuration of Fig. 3.33, determine the following at quiscent point: (a) I_B, I_C, V_{CE} (b) V_C, V_B, V_E (c) V_{CB}

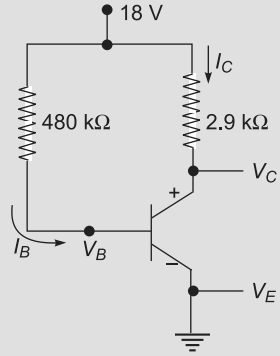


Fig. 3.33 $\beta = 95$

Solution

(a) $I_B = \dfrac{18 - 0.7}{480} = 36$ μA

$I_C = 95 \times 36 = 3.42$ mA

$V_{CE} = 18 - 2.9 \times 3.42 = 8.08$ V

(b) $V_C = 8.08$ V,

$V_B = 0.7$ V,

$V_E = 0$

(c) $V_{CB} = 8.08 - 0.7 = 7.38$ V, reverse biased (OK)

EXAMPLE 3.8 For the emitter-stabilised BJT bias circuit of Fig. 3.34, determine I_{BQ}, I_{CQ}, V_{CEQ}, V_B, V_C, V_E.

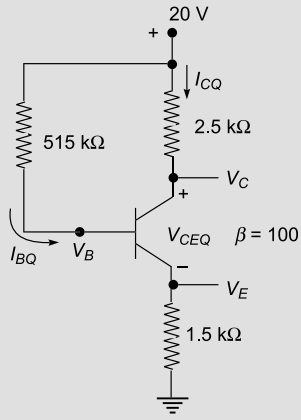


Fig. 3.34

Solution

$$I_{BQ} = \frac{(20 - 0.7) \times 10^3}{515 + (1 + 100) \times 1.5} = 29\ \mu A$$

$$I_{CQ} = 100 \times 29 \times 10^{-3} = 2.9\ mA$$

$$V_{CEQ} = 20 - 2.9 \times 2.5 - 2.9 \times 1.5 \qquad (I_E \approx I_{CQ})$$

$$= 8.4\ V$$

$$V_C = 20 - 2.9 \times 2.5 = 12.75\ V$$

$$V_E = 2.9 \times 1.5 = 4.35\ V$$

EXAMPLE 3.9 For the voltage-divider bias circuit of Fig. 3.33, determine I_{RQ}, I_{CQ}, V_{CEQ}, V_C, V_E, V_B.

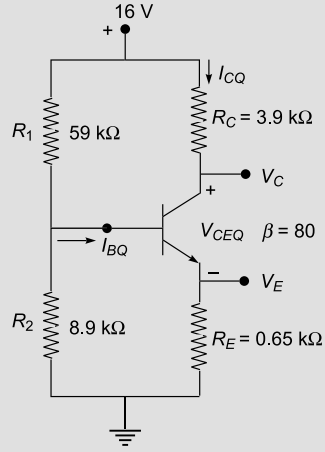


Fig. 3.35 $\beta = 80$

Solution

Thevenin equivalent as seen from the base

$$V_{th} = \left(\frac{8.9}{59 + 8.9}\right) \times 16 = 2.1 \text{ V}$$

$$R_{th} = \frac{59 \times 8.9}{59 + 8.9} = 7.73 \text{ k}\Omega$$

The circuit is now drawn in Fig. 3.36.

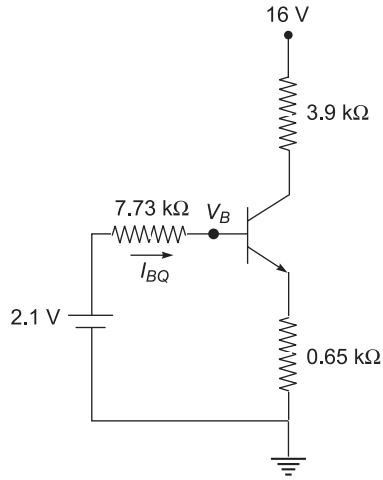


Fig. 3.36

$$I_{BQ} = \frac{(2.1 - 0.7) \times 10^3}{7.73 + (1 + 80) \times 0.65} = 23.2 \,\mu\text{A}$$

$$I_{CQ} = 80 \times 23.3 \times 10^{-3} = 1.856 \text{ mA}$$

CE loop KVL equation

$16 - 3.9\, I_{CQ} - V_{CEQ} - 0.65\, I_{EQ} = 0$

$16 - 3.9 \times 1.805 - V_{CEQ} - 0.65 \times 1.865 = 0$

$V_{CEQ} = 7.51$ V

$V_C = 16 - 3.9 \times 1.856 = 8.76$ V

$V_E = 8.76 - 7.51 = 1.25$ V

$V_B = 1.25 + 0.7 = 1.95$

Check

$V_B = 2.1 - 7.73 \times 23.2 \times 10^{-3} = 1.92$ V (OK)

EXAMPLE 3.10 For the voltage feedback circuit of Fig. 3.37, determine I_B, I_C, V_C.

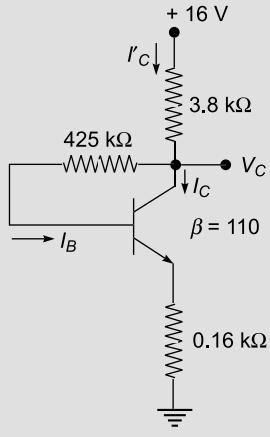


Fig. 3.37

Solution

$$I'_C = \frac{16 - V_C}{3.8} = I_B + I_C$$

or $$16 - V_C = 3.8\, I_B + 3.8\beta\, I_B$$

or $$V_C = 16 - 3.8\,(8 + 110)\, I_B \quad \text{(i)}$$

$$V_C = 16 - 421.8\, I_B \quad \text{(ii)}$$

BE loop KVL

$$V_C = 425\, I_B - 0.7 - 0.61\,(1 + 110)\, I_B = 0$$

$$V_C = (425 + 67.7)\, I_B - 0.7 \quad \text{(i)}$$

$$V_C = 492.7\, I_B - 0.7 \quad \text{(ii)}$$

From Eqs (i) and (ii)

$$16 - 421.8\, I_B = 492.7\, I_B - 0.7$$

$$914.5\, I_B = 16.7$$

$$I_B = 18.26 \text{ mA}$$

$$I_C = \beta I_B = 110 \times 18.26 \times 10^{-3} = 2 \text{ mA}$$

From Eq (ii),

$$V_C = 492.7 \times 18.26 \times 10^{-3} - 0.7 = 8.3 \text{ V}$$

Check

$$I'_C = I_C + I_B = 111\, I_B = 111 \times 18.26 \times 10^{-3} = 2.027 \text{ mA}$$

$$V_C = 16 - 3.8 \times 2.027 = 8.29 \text{ V (OK)}$$

EXAMPLE 3.11 The collector characteristics of a BJT are drawn in Fig. 3.38.

(a) Locate an operating point midway between the active region.

(b) What are the values of I_{CQ} and V_{CEQ}?

(c) Determine the value of R_B to establish this operating point for fixed-bias configuration. Calculate P_{CQ}, given V_{CC} = 20 V.

(d) Calculate the value of β and α at the operating point.

(e) What is the value of the saturation current , $I_{C(sat)}$.

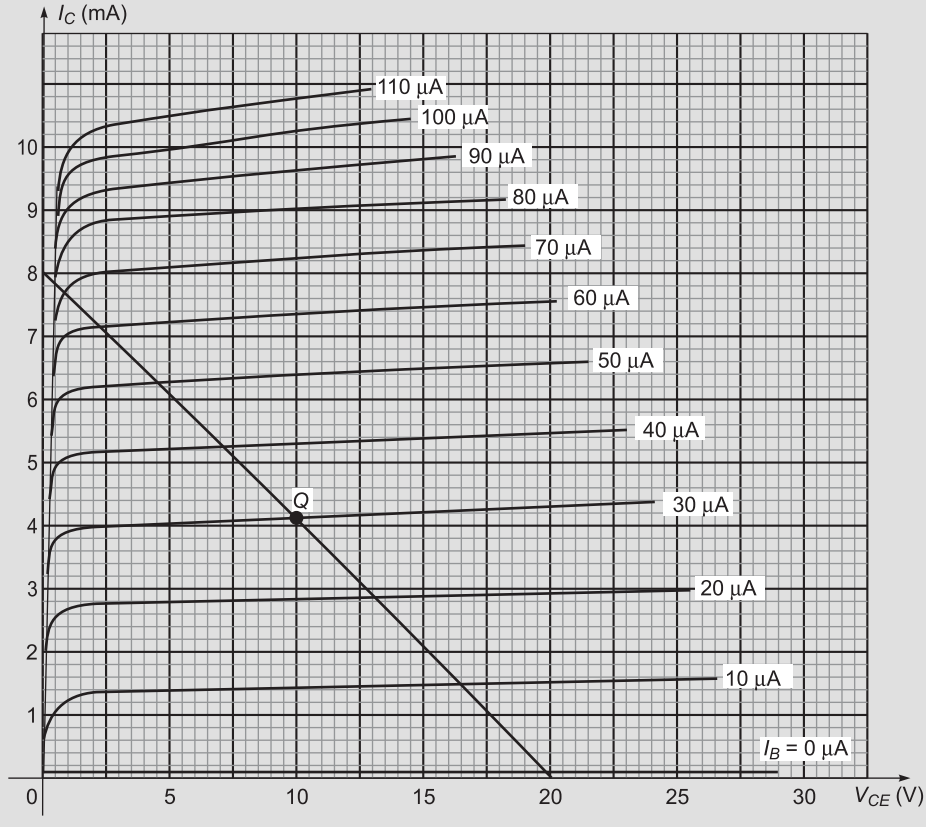


Fig. 3.38

Solution

(a) and (b) Q-point is located on Fig. 3.38 where

$$I_{CQ} = 4.1 \text{ mA}, V_{CEQ} = 10 \text{ V}, I_{BQ} = 30 \text{ μA}$$

(c) From $V_{CE} = V_{CC} = 20$ V, draw a line through Q which intersect

I_C = axis at $I_C = 8$ mA

$$\frac{V_{CC}}{R_C} = 8\text{ mA} \rightarrow R_C = \frac{20}{8} = 2.5\text{ k}\Omega$$

(d) $$R_B = \frac{20 - 0.7}{30\,\mu\text{A}} = 643\text{ k}\Omega$$

$$\beta = \frac{4.1 \times 10^3}{30} = 136.7$$

$$\alpha = \frac{\beta}{\beta + 1} = \frac{136.7}{137.7} = 0.993$$

(e) $I_{C\,(sat)} \approx 10$ mA

SUMMARY

➢ BJT is introduced with constructional details. Its operational characteristics and configurations are explained. SCR and UJT are introduced with their characteristics and biasings.

EXERCISES

➥ Review Questions

1. Comment on the doping levels of the three components of a BJT.
2. Why is the base layer of a BJT made very thin compared to emitter and collector layer?
3. What are the three regions of operation of a BJT. Explain with help of *CE* configuration collector characteristics.
4. In the three regions of operation how are the BJT junctions biased?
5. Define α of a BJT.
6. Define β of a BJT.
7. How are α and β related?
8. Distinguish between α_{dc} and α_{ac}
9. Distinguish between β_{dc} and β_{ac}.
10. What is reverse saturation current in a BJT? How can this be observed independently?
11. Draw the symbols for *pnp* and *npn* BJT. What distinguishes one from the other?
12. How are I_{CBQ} and I_{CEO} related?
13. What is meant by biasing a transistor?
14. What is self-bias? How does it help in stabilising the *Q*-point?
15. Draw the circuit of a voltage-divider bias.
16. What is voltage feedback bias? Draw the circuit. Is this type of bias almost independent of β?
17. What is meant by β insensitive bias?

➥ Problems

1. For common-base configuration for V_{BE} = 0.7 V, I_E is adjusted to 5 mA. What is the value of I_C? Does it depend on V_{CB} in the active region. The collector characteristics are drawn in Fig. 3.8.
2. For the common emitter configuration, I_B = 30 mA, V_{CE} = 7.5 V.
   (a) Calculate β_{dc} and β_{ac}.
   (b) Find the value of I_E.
   (c) Find the change in I_E for ΔI_B = +5 μA.
   The collector characteristics are drawn in Fig. 3.10(a).
3. From the collector characteristics of Fig. 3.10(a), the operating point is at I_B = 20 μA and V_{CE} = 10 V. Calculate α_{dc} and the corresponding I_E.
4. A BJT is connected in *CE* configuration as in Fig. 3.39; *BJT* β = 100.

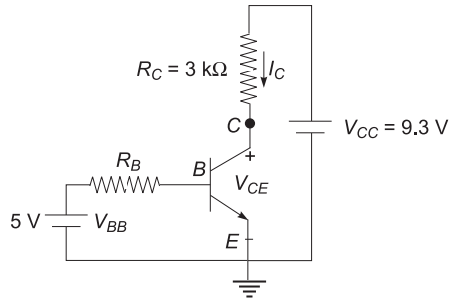


Fig. 3.39

Determine
(a) I_B (b) I_C (c) V_{CE} (d) V_{CB} assuming the transistor to be in active region. Check from the value of V_{CB} if it is so.

Hint: *CB* junction should be reverse biased.

5. For the fixed-bias configuration of Fig. 3.40, determine R_B, R_C, I_C, V_{CE}.
6. For the emitter-stabilised bias circuit of Fig. 3.41, determine R_B, R_C, R_E and V_B.

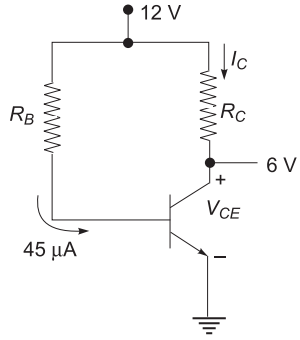


Fig. 3.40 β = 100

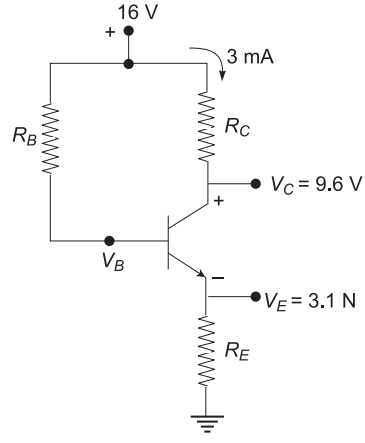


Fig. 3.41 β = 80

7. For the voltage-divider bias circuit of Fig. 3.42, determine I_C, V_E, V_{CC}, V_B, R_1.
8. For the voltage feedback network of Fig. 3.43, determine V_E, I_C, V_C, V_{CE}, I_B, β.

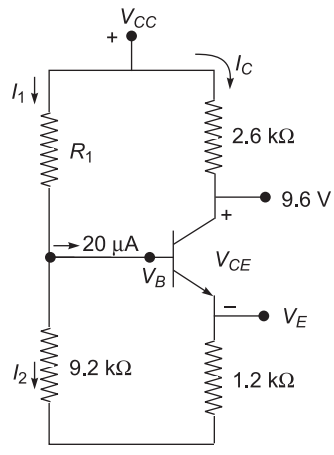


Fig. 3.42 $\beta = 101$

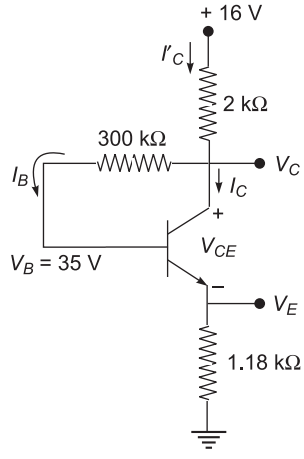


Fig. 3.43

9. For the network of Fig. 3.44, determine
I_E, V_C, V_{CE}

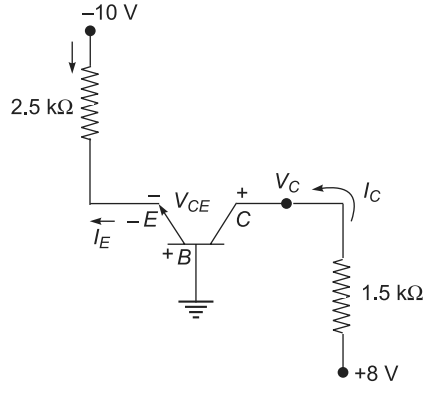


Fig. 3.44

➥ Multiple-Choice Questions

1. Current gain of BJT in common base is
   (a) α (b) β (c) γ (d) none of these
2. An SCR device has
   (a) four layers (b) three layers (c) two layers (d) one layer
3. A UJT device has
   (a) four layers (b) three layers (c) two layers (d) none of these

4. The value of total collector current in a *CB* circuit is
(a) $I_C = \alpha I_E + I_{CO}$ (b) $I_C = \alpha I_E$ (c) $I_C = \beta I_E$ (d) $I_C = \alpha I_{CO} - I_E$

5. When the collector junction in transistors is biased in the reverse direction and the emitter junction in forward directions, the transistor is said to be in the
(a) cut-off region (b) saturation region (c) active region (d) none of these

6. Which of the following acts like a diode and two transistors?
(a) UJT (b) SCR (c) TRIAC (d) DIAC

7. Early effect in BJT refers to
(a) avalanche breakdown (b) zener breakdown
(c) base narrowing (d) none of these

8. In a BJT, largest current flows
(a) in the base (b) in the emitter
(c) in the base and emitter (d) in the collector

9. The "cut-in" voltage of a silicon small-signal transistor is
(a) 0.5 V (b) 0.3 V (c) 1.0 V (d) 0.9 V

10. An SCR consists of
(a) five *PN* junctions (b) two *PN* junctions (c) four *PN* junctions (d) three *PN* junctions

ANSWERS

✦ Problems

1. $I_C \approx 5$ mA, no
2. $\beta_{dc} = 113$ $\beta_{ac} \approx 100$
3. $\alpha_{dc} = 0.992$, $I_E = 2.4$ mA
4. (a) 21.5 µA (b) 2.15 mA (c) +2.85 V
5. $R_B = 251$ kΩ, $I_c = 4.5$ mA, $R_c = 1.33$ kΩ, $V_{CE} = 6$ V
6. $R_C = 2.13$ kΩ, $R_E = 1.03$ kΩ, $R_B = 324.6$ kΩ, $V_B = 3.83$ V
7. $I_C = 2$ mA, $V_{CC} = 14.8$ V, $V_E = 2.4$ V, $V_B = 3.1$ V, $R_1 = 32.8$ kΩ
8. $V_E = 2.74$ V, $I_C = 2.32$ mA, $V_c = 12.3$ V, $V_{CE} = 9.56$ V, $I_B = 4$ µA
9. $I_E = 3.72$ mA, $V_C = 2.42$ V, $V_{CE} = 3.12$ V

✦ Multiple-Choice Questions

1. (a) **2.** (a) **3.** (b) **4.** (a) **5.** (c) **6.** (a) **7.** (c) **8.** (b) **9.** (a) **10.** (d)

CHAPTER 4

Field Effect Transistors (FETs)

GOALS AND OBJECTIVES

- ❏ Introduction of field effect transistors (FETs)
- ❏ Construction and characteristics of JFETs
- ❏ Constructional features and analysis of metal oxide field effect transistors (MOSFETs)
- ❏ Explanation of FET circuit configurations, CMOS circuits and FET biasing
- ❏ Understanding of small signal model analysis
- ❏ Common source and drain amplifier configurations

4.1 | INTRODUCTION

Like BJTs, FETs are three-terminal devices. These differ from BJTs in two respects.

- While BJT is a current-controlled device (input I_B controls output I_C), "FET is voltage-controlled device (input voltage controls output current)."
- BJT is bilateral while FET is unilateral, which means that only one type of carrier participates in conduction.

FETs are of two types: Junction Field Effect Transistor (JFET) and Metal Oxide Semiconductor Field-Effect Transistor (MOSFET).

4.2 | JFETs

4.2.1 Construction and Characteristics

An *N*-type material channel forms the major part of the structure. It is embedded on both sides with *P*-type material as shown in Fig. 4.1 (*N* and *P* could be interchanged). The two *p*-type materials are joined together through ohmic contacts and connected to the terminal gate (*G*).

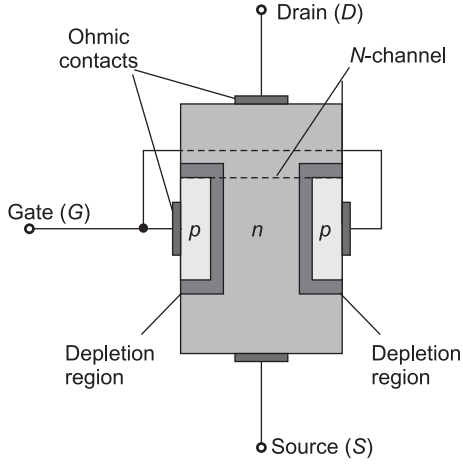


Fig. 4.1 Junction Field Effect Transistor (JFET) N-channel

The *N* and *P* materials form *PN*-junctions on the two sides of the channel. The top and bottom of the *N*-channel make ohmic contact with the Drain (*D*) and source (*S*) terminals, respectively. Around both the *PN*-junctions there is the depletion region (like in a diode).

Operation with $V_{GS} = 0$ and $V_{DS} > 0$ (variable)

✦ Connection Diagram

In Fig. 4.2, voltage $V_{DS} > 0$ is applied across the *D* and *S* terminals, V_{DS} can be varied by the source V_{DD}. The gate terminal *G* is connected to the source terminal *S* so that $V_{GS} = 0$. V_{DS} causes the channel electrons to flow from *S* to *D*, the conventional current I_D flows into *D* and I_S flows out of *S*; obviously, $I_D = I_S$.

4.2.2 Operation

The voltage V_{DS} reverse biases both the *PN*-junctions. The reverse biasing reduces towards the *S* terminal because of voltage drop in the channel from *D* to *S*. As a result, the depletion region widens at the *D*-end of the channel. The width of depletion region reduces along the channel towards the *S*-end. The depletion region is therefore non-uniform as shown in Fig. 4.2. As V_{DS}

is increased from zero, I_D increases. As the channel width is decreasing with increase of V_{DS}, I_D begins to level off as shown in Fig. 4.3. At $V_{DS} = V_P$ (called pinch-off voltage), I_D becomes saturated at I_{DSS}, and becomes independent of V_{DS}. The initial behaviour of the channel is that of voltage-controlled resistor.

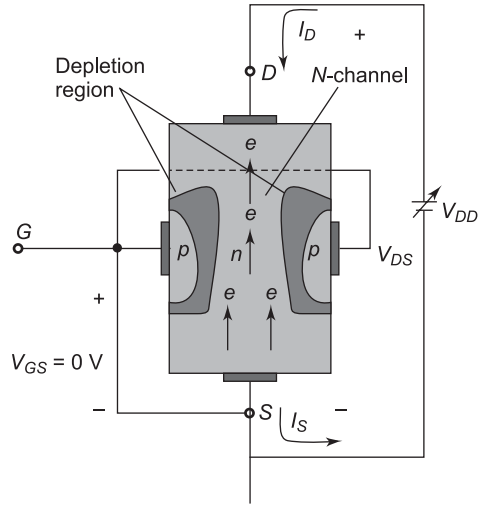


Fig. 4.2 Connection diagram of JFET, $V_{GS} = 0$, $V_{DS} > 0$

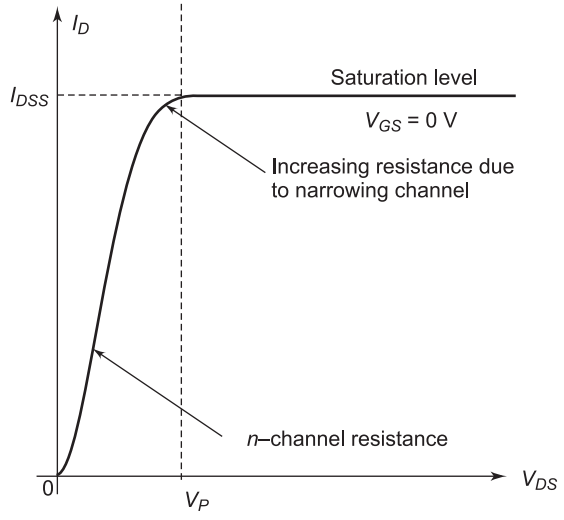


Fig. 4.3 JFET characteristic; $V_{GS} = 0$, V_{DS} increasing

✦ Operation $V_{GS} < 0$, $V_{DS} > 0$, Both Variable

A voltage source is connected between the gate G and source S terminals causing V_{GS} to be negative as shown in Fig. 4.4. This negative V_{GS} reverse biases both the junctions uniformly, reducing the channel width throughout. This is over and above the effect of V_{DS}. The channel

therefore pinches off at value $V_{DS} < V_P$ for $V_{GS} = 0$ as shown in Fig. 4.5. The pinch-off values of V_{DS} continue to decrease as V_{GS} is made more negative. At a value of $V_{GS} = V_P$(–4 V in Fig. 4.5), the channel completely closes and so $I_D = 0$ irrespective of the value of V_{DS}. Then pinch-off is caused by negative V_{GS} (= V_P).

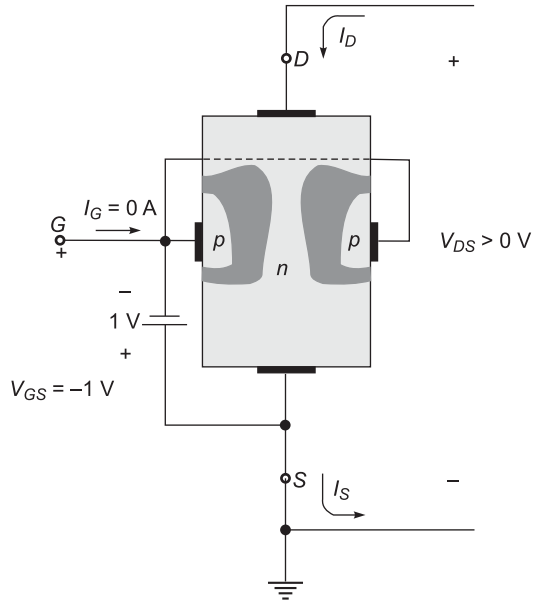


Fig. 4.4 $V_{GS} < 0$, $V_{DS} > 0$, both variable

The complete characteristics (typical) are drawn in Fig. 4.5. It is to be noted that as the two junctions are reverse biased, $I_G \approx 0$.

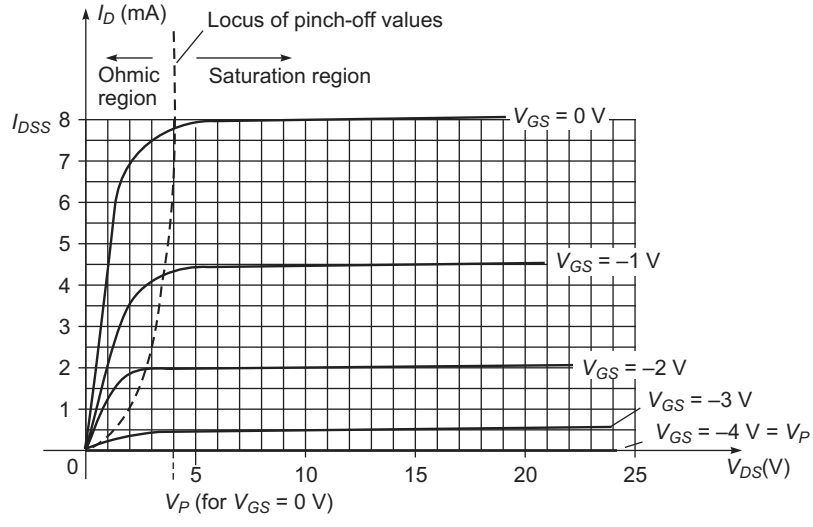


Fig. 4.5 N-Channel JFET characteristics with I_{DSS} = 8 mA and V_P = –4 V

✦ **Symbols**

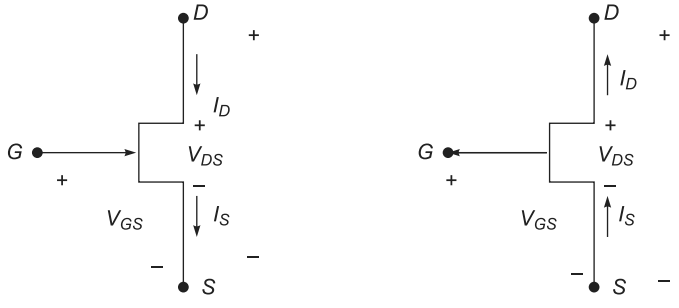


Fig. 4.6 (a) n-channel JFET (b) p-channel JFET

4.2.3 Voltage-controlled Resistor

It is observed from the characteristics of Fig. 4.5 that these are linear (almost) to lie left of the V_P locus. It is further observed that the slope decreases as V_{GS} increases. It means that the JFET acts as voltage-controlled resistor; and resistance increases with V_{GS}. To a good approximation, the resistance offered by JFET in the ohmic region can be expressed as

$$r_d = \frac{r_0}{(1 - V_{GS}/V_P)^2} \tag{4.1}$$

where r_0 = resistance with $V_{GS} = 0$
r_d = drain resistance

4.2.4 Transfer Characteristic

It is observed from the characteristics of Fig. 4.5 that the characteristics to the right of V_P locus, the saturation region (the major part of the characteristics), that I_D is dependent on V_{GS} but is independent of V_{DS}. For any value of V_{DS}, (preferably in the middle), we can read I_D for various values of V_{GS} (from zero to V_P) and plot I_D vs V_{GS} as shown in Fig. 4.7. This plot is known as the transfer characteristic [it transfers V_{GS} (input) to I_D (output)]. That is why JFET is a voltage-controlled device in which input voltage controls the output current.

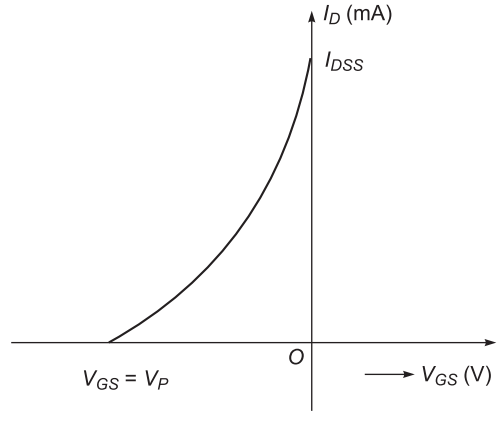


Fig. 4.7 Transfer characteristics

4.2.5 Shockley's Equation

The transfer characteristic can be expressed analytically as

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{4.2}$$

At $V_{GS} = 0$, $I_D = I_{DSS}$ and at $V_{GS} = V_P$. $I_D = 0$

Equation (4.2) can be written in the form to determine V_{GS} for given I_D as

$$V_{GS} = V_P\sqrt{1 - \frac{I_D}{I_{DSS}}} \tag{4.3}$$

4.2.6 *P*-Channel FET

In *p*-channel JFET, holes act as majority charge carriers. The current flow is due to the movement of these holes. Like *n*-channel, *p*-channel JFET has three terminals, namely—gate, source and drain terminal (Fig. 4.8(a)). It carries two *n*-type silicon terminals positioned on both sides and connected to the gate (*G*) terminal. The drain and source leads are connected to both sides of *p*-type channel. When V_{GS} is zero, the current (due to holes as majority carriers) flows freely.

✦ Operation

For positive V_S and zero gate voltage, the current flows through drain source channel (maximum current) and FET is in ON state or in active region. When positive voltage is applied to the gate terminal, the drain source current starts decreasing till it reaches its cutoff and transistor enters into OFF state.

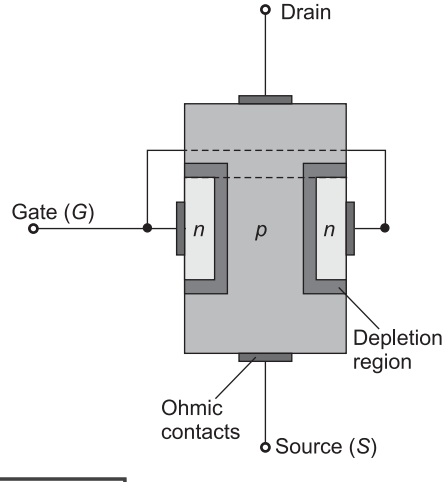


Fig. 4.8(a) *P*-channel construction diagram

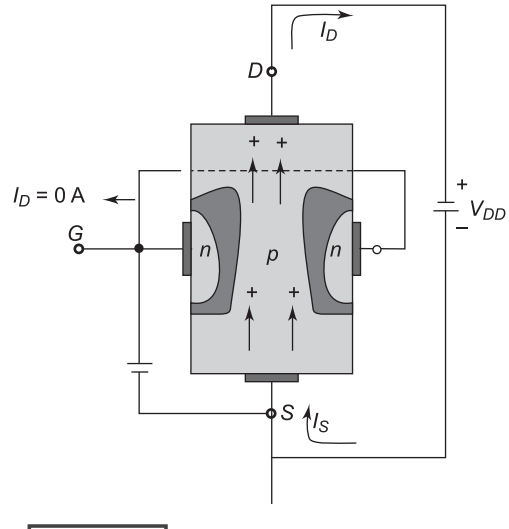


Fig. 4.8(b) *P*-channel connection diagram

4.2.7 Drain Characteristics of *P*-channel FET

From the transistor characteristics, it is shown that gain decreases with increase in positive voltage at gate terminal. For zero voltage at the gate, the drain current attains its maximum value.

P-type JFET turns on with positive voltage at source terminal and ideally no voltage at gate terminal, and turns off when voltage at gate increases typically by +4V (Fig. 4.9).

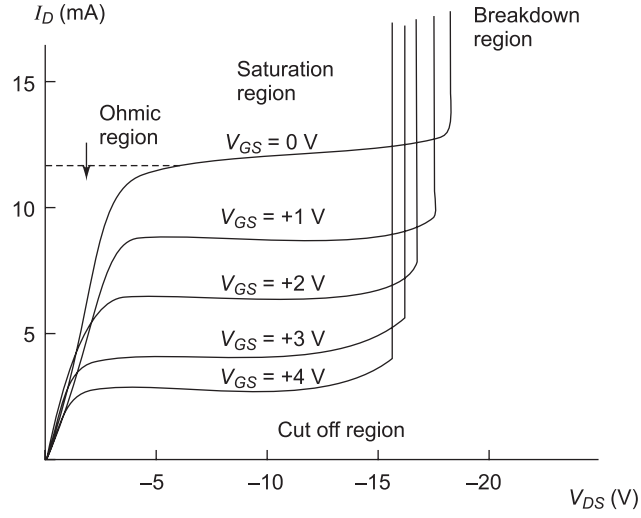


Fig. 4.9 *P*-channel JFET characteristics curve

❑ ***Cutoff Region*** The region when JFET is in OFF state, i.e. no current I_D flows.

❑ ***Ohmic Region*** The region when JFET transistor starts showing resistance to the drain current I_D, i.e. current starts flowing through drain source region.

❑ ***Saturation Region*** This is the fully operational region and maximum current flows during this region. During this region transistor is in ON state.

❑ ***Breakdown Region*** It is the region when the voltage supplied to the source exceeds the required voltage. The transistor loses its ability to resist current and breaks down. During this the current flows from source to drain.

EXAMPLE 4.1 From the drain characteristics of JFET (*N*-channel) of Fig. 4.5, read I_{DSS}, V_P. Using Shockley's equation, check the values of I_D for $V_{GS} = -1$ and $V_{GS} = -2$.

Solution From the figure, $I_{DSS} = 8$ mA, $V_P = -4$ V

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

At $V_{GS} = -1$ V,

$$I_D = 8\left(1 - \frac{-1}{-4}\right)^2 = 8 \times \left(\frac{3}{4}\right)^2 = \frac{9}{2} = 4.5 \text{ mA (checks)}$$

At $V_{GS} = -2$ V,

$$I_D = 8\left(1 - \frac{2}{4}\right)^2 = 8 \times \frac{1}{4} = 2 \text{ mA (checks)}$$

EXAMPLE 4.2 From the drain characteristics of Fig. 4.5, write down the progressive difference in I_D against the difference in V_{GS}. What is the nature of their relationship?

Solution

ΔV_{GS} (V)	ΔI_D(A)
–1	0.5
–1	1.5
–1	2.5
–1	3.5

The relationship in nonlinear.

EXAMPLE 4.3 At a Q-point of JFET, I_{OQ} = 3.5 mA and V_{GS} = –3 V; determine I_{DSS} if V_P = –6 V.

Solution

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

$$3.5 = I_{DSS}\left(1 - \frac{-3}{-6}\right)^2 = \frac{1}{4} I_{DSS}$$

or $$I_{DSS} = 3.5 \times 4 = 14 \text{ mA}$$

EXAMPLE 4.4 For JFET, I_{DSS} = 6 mA, V_P = –4.5 V

Determine (a) I_D at V_{GS} = –2 and –4 V; (b) V_{GS} at I_D = 3 and 5.5 mA

Solution

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

(a) (i) $$I_D = 6\left(1 - \frac{-2}{-4.5}\right)^2 = 1.85 \text{ mA}$$

(ii) $$I_D = 6\left(1 - \frac{-4}{-4.5}\right)^2 = 0.074 \text{ mA}$$

(b) (i) $$3 = 6\left(1 - \frac{V_{GS}}{-4.5}\right)^2$$

$$0.5 = (1 + 0.222\ V_{GS})^2$$

$$1 + 0.222\ V_{GS} = \pm 0.707$$

$$0.222\ V_{GS} = -1.707, -0.293$$

$$V_{GS} = -1.32 \text{ V}, -7.73 > V_P \text{ rejected}$$

(ii) $$5.5 = 6(1 + 0.222\ V_{GS})^2$$

$$0.917 = (1 + 0.222\ V_{GS})^2$$

$$1 + 0.222\ V_{GS} = \pm 0.957$$

$$0.222\ V_{GS} = -0.043, -1.957$$

$$V_{GS} = -0.194 \text{ V}$$

EXAMPLE 4.5 For the JFET drain characteristics of Fig. 4.10, determine the difference in I_D for $V_{GS} = -2$ V and -1 V. Check the result using Shockley's equation.

Solution From the characteristics, ΔI_D for $V_{GS} = -2$ V, -1 V, $\Delta I_D = 7 - 4.5 = 2.5$ mA

Further,

$$I_{DSS} = 10 \text{ mA}, V_P = -6 \text{ V}$$

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

At $V_{GS} = -2$ V,

$$I_D = 10\left(1 - \frac{2}{6}\right)^2 = 4.44 \text{ mA}$$

At $V_{GS} = -1$ V,

$$I_D = 10\left(1 - \frac{1}{6}\right)^2 = 6.94 \text{ mA}$$

$$\Delta I_D = 6.94 - 4.44 = 2.5 \text{ mA}$$

From characteristics, $\Delta I_D = 2.6$ mA

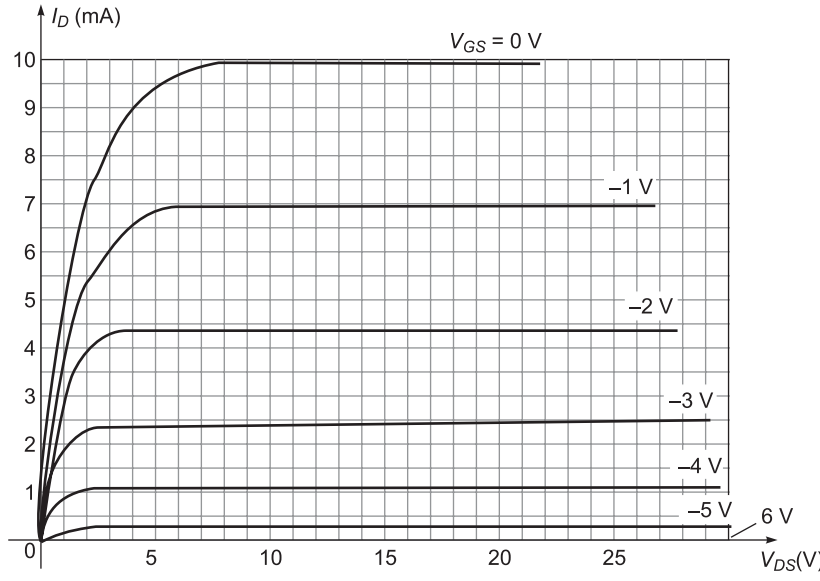


Fig. 4.10 Drain characteristic

4.2.8 Small Signal Model

Small signal modeling is used to estimate the behaviour of electronic circuits and is related to the small changes in FET current and voltage about *Q* point. Small signal modeling analysis is done by replacing all signal sources with their ideal internal resistance, leaving the ac voltages and currents in the circuit. Small signal models are different for low and high frequencies.

✦ Small Signal (Low Frequency) FET Model

From the diagram given below (Fig. 4.11), it is clear that the source gate function is kept as an open circuit, hence no current is drawn by this terminal and input resistance of gate source terminal becomes very high. Though the resistance of input terminal is very high but gate source voltage affects the value of drain current which is represented by voltage controlled current source (g_m , V_{GS}) with value proportional to gate source voltage.

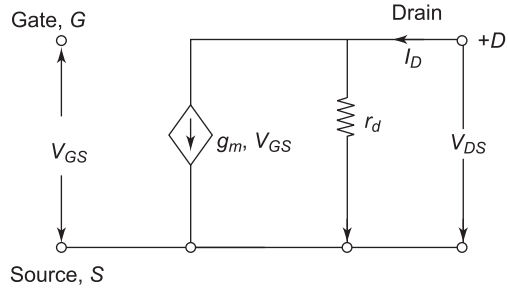


Fig. 4.11 Small signal model analysis

The output resistance or the drain resistance is represented by $r_{d.}$ Drain resistance may have a typical value from 10-50 kΩ.

g_m is the transconductance that can be measured as mA/V or mS or millimho with typical values from 0.5–10 mA/V.

I_D drain current is the function of V_{GS} and V_{DS}.

$$i_D = f(V_{GS} \cdot V_{DS})$$

The change in drain current I_D due to variation in both drain and gate voltage is given below wing Tylor's series.

$$\Delta I_D = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}} \Delta V_{GS} + \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{GS}} \Delta V_{DS}$$

replacing

$$\Delta I_D = i_D$$

$$\Delta V_{GS} = v_{GS}$$

$$\Delta V_{DS} = v_{DS}$$

Hence,

$$i_D = g_m v_{GS} + \Upsilon_{r_d}^{v_{DS}}$$

where

$$g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}} \approx \left.\frac{\Delta I_D}{\Delta V_{GS}}\right|_{V_{DS}} = \left.\frac{i_D}{V_{GS}}\right|_{V_{DS}}$$

Output resistance is defined as:

$$r_d = \left.\frac{\partial V_{DS}}{\partial I_D}\right|_{V_{GS}} \approx \left.\frac{\Delta V_{DS}}{\Delta I_D}\right|_{V_g}$$

✦ Small Signal (High Frequency) FET Model

At low frequency, the interelectrode capacitance is not considered, but in higher frequency devices capacitance between the terminals causes reduction in gain and hence is accounted. From Fig. 4.12, C_{GS} represents barrier capacitance (gate and source), C_{GD} represents barrier capacitance (between gate and drain), and C_{DS} is the capacitance between drain and source. Due to internal capacitances, the feedback exits between the input and output terminals. With increase in frequency, the voltage amplification drops.

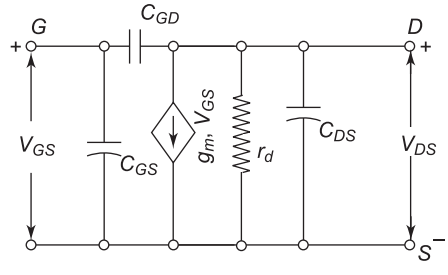


Fig. 4.12 Small signal FET model

EXAMPLE 4.6

For given Fig. 4.13 I_{Dss} = 9mA, Vp = −5V, and Y_{as} (max) = 50 µs. Find mid-band ac gain, output and input impedance.

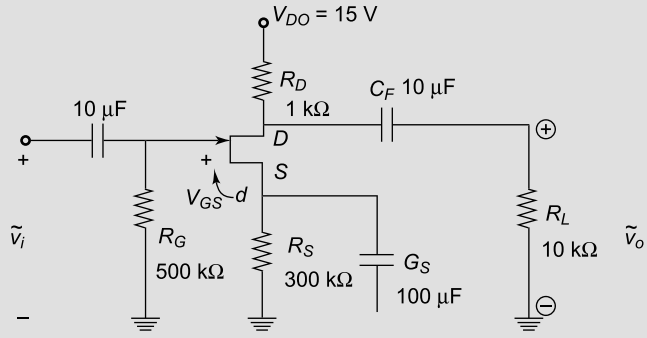


Fig. 4.13

Solution

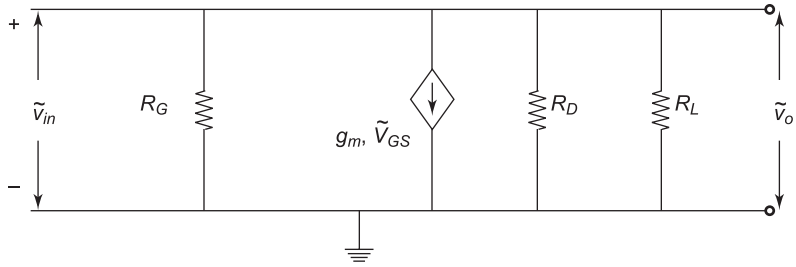


Fig. 4.14

$$A_v = \frac{\tilde{V}_0}{\tilde{V}_1} = \frac{-g_m \tilde{V}_{GS} (R_D \| R_L)}{\tilde{V}_{GS}} = -g_m \frac{R_D R_L}{R_D + R_L}$$

$$\boxed{A_V = -2.3}$$

$$\tilde{Z}_i = R_G = 500\ \text{k}\Omega$$

$$\tilde{Z}_0 = R_D \| R_L = 909\ \Omega$$

4.3 METAL OXIDE SEMICONDUCTOR FIELD EFFECT TRANSISTOR (MOSFET)

The MOSFET transistor has become the most important device for the fabrication of integrated circuits for digital computers. Its thermal stability and other general features makes it suitable for IC design fabrication because of smaller silicon-chip area.

Depletion-Type MOSFET Constructional Features

Refer Fig. 4.15. On a *p*-type substrate, an *n*-channel is formed which is connected to *D* and *S* terminals through heavily doped *n*-regions.

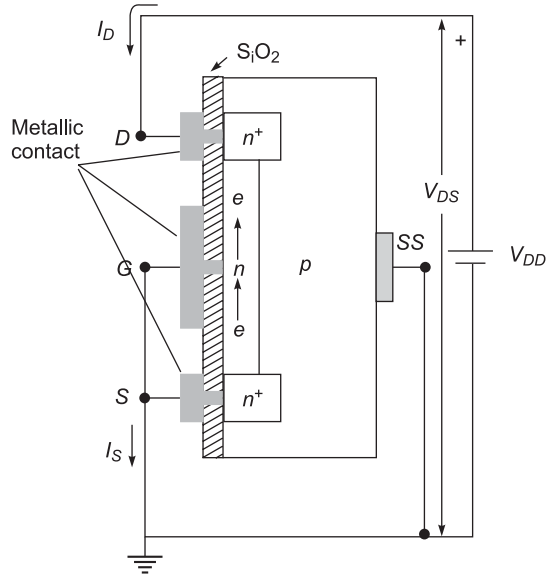


Fig. 4.15 MOSFET—*n*-channel depletion type $V_{GS} = 0$, $V_{DS} > 0$

The *N*-channel is insulated from the gate *G* terminal by an SiO_2 layer, which extends over the complete device. The great quality of the insulated gate is $I_G = 0$ irrespective of any gate voltage which means, extremely high input resistance. This is not so in JFET where I_G is the reverse saturation current though very small (nA).

✦ Operation

When $V_{GS} = 0$ (i.e. positive V_{DS}), saturation current $I_D = I_S = I_{DSS}$ will flow.

When a negative voltage is applied to the gate, say $V_{GS} = -1$, the holes from the substrate are attracted by the gate and so the holes flow into *N*-channel. The holes recombine with electrons being repelled by the gate, thereby reducing the concentration of electrons in the channel as shown in Fig. 4.16.

The result is reduction in saturation current $I_{D(sat)}$. So to make V_{GS} more negative I_D keeps reducing till at $V_{GS} = V_P$, the channel pinches off. The drain characteristics and transfer characteristics are similar to those of JFET, as shown in Fig. 4.17.

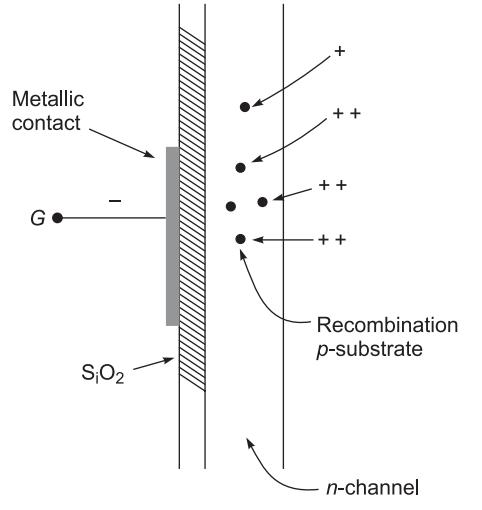


Fig. 4.16 Channel depletion process

However, unlike JFET, if V_{GS} is made positive, the minority electrons get attracted into the *N*-channel and so I_D begins to increase sharply. Even at small values of positive V_{GS}, I_D may exceed the prescribed limit.

Depletion mode and enhancement mode are both shown in Fig. 4.17.

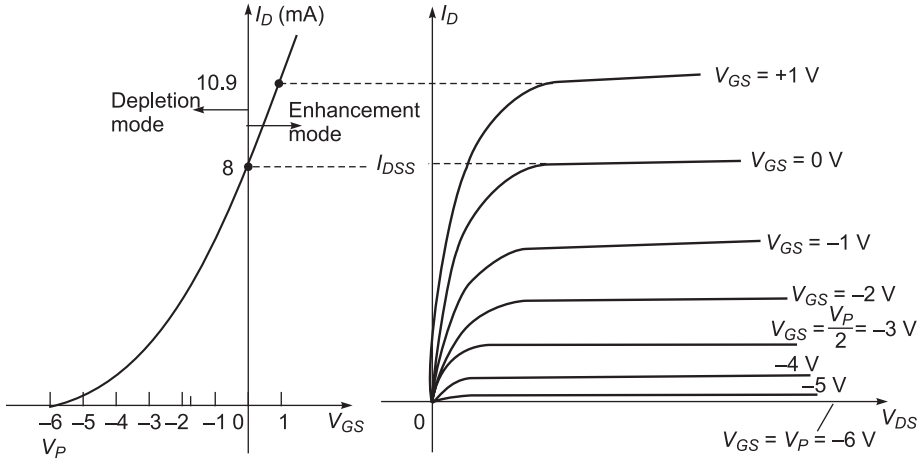


Fig. 4.17 MOSFET *N*-channel depletion type

4.3.1 *P*-channel MOSFET

P-type MOSFET is composed of holes as majority carriers. In its ON state, the current carrier holes move through the channel. *P*-type MOSFET is further divided as enhancement mode and depletion mode FET.

✦ *P*-channel Depletion Type

Compared to *N*-channel type, V_{GS} is positive for depletion; V_{DS} reverses, –ve at *D* and +ve at *S*. The pinch-off voltage V_P will be positive; see Fig. 4.18.

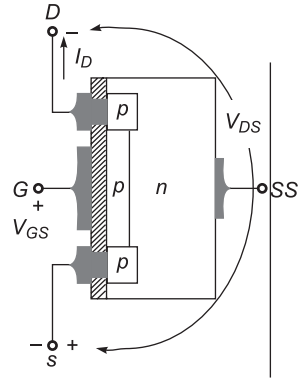


Fig. 4.18 *P*-channel depletion type

When the voltage between the gate and source terminal is zero then depletion type MOSFET (p-channel) is in ON state and current flows form source to drain. When voltage applied to gate terminal is increased, the drain to source path becomes more resistive. If the voltage at gate terminal is made higher than that required, the transistor completely shuts off.

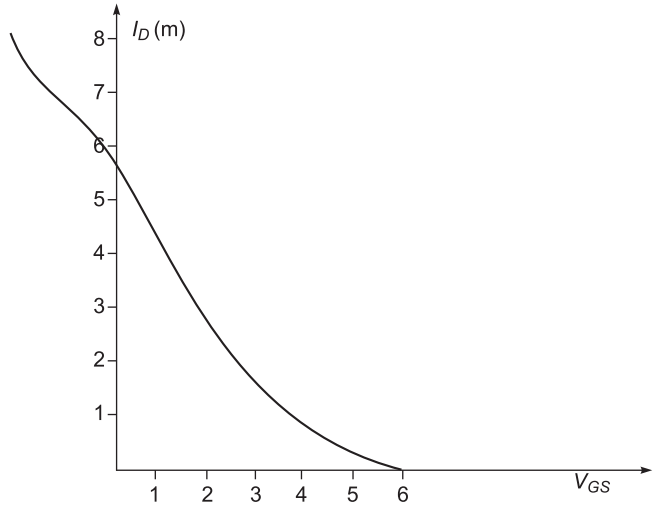


Fig. 4.19 Transfer characteristics of DMOSFET *P*-channel

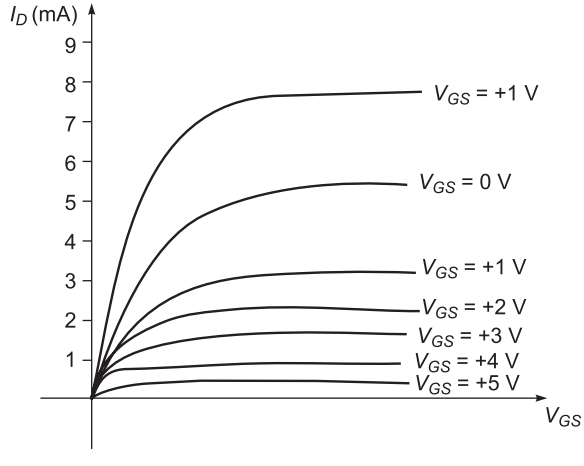


Fig. 4.20 Drain characteristics of DMOSFET *P*-channel depletion method

✦ The Shockley Equation

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{4.4}$$

applies for the depletion MOSFET.

✦ Symbols

The symbols for *n*-channel and *p*-channel depletion-type MOSFETs, and the gap between gate and channel are indicative of the insulation-layer of SiO_2.

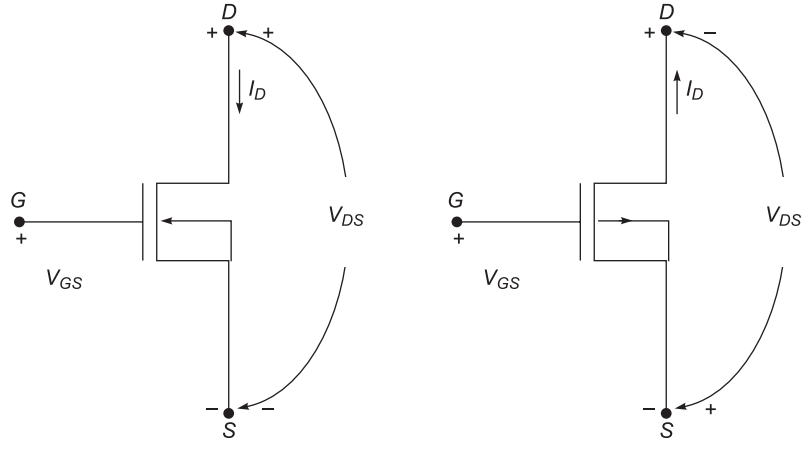


Fig. 4.21 Symbols for depletion-type MOSFET

✦ Enhancement-Type MOSFET

The construction is similar to that of depletion type except that there is no channel. Only the *P*-substrate as shown in Fig. 4.22.

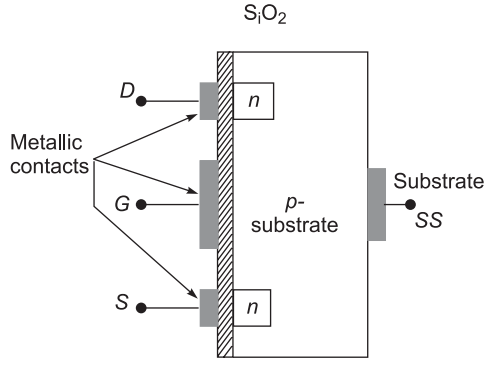


Fig. 4.22 Constructional details of enhancement MOSFET

In enhancement mode, the *p*-type MOSFET behaves in opposite manner as it is normally OFF when the gate-source voltage is 0 ($V_{GS} = 0$). When negative voltage is applied to the gate terminal then the channel becomes conductive and reaches ON state.

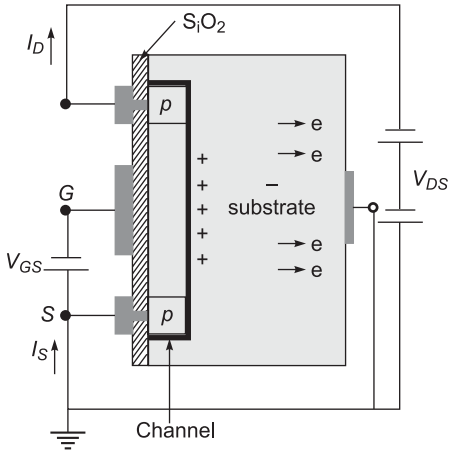


Fig. 4.23 *P*-channel enhancement type MOSFET

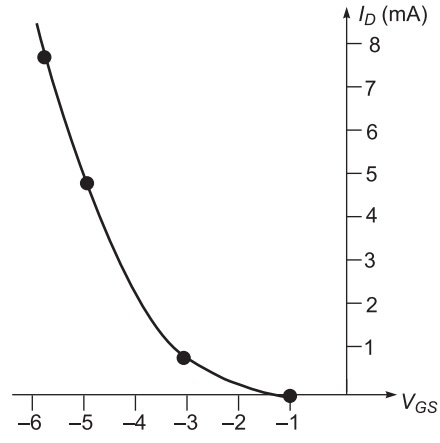


Fig. 4.24 Transfer characteristics of EMOSFET *P*-channel

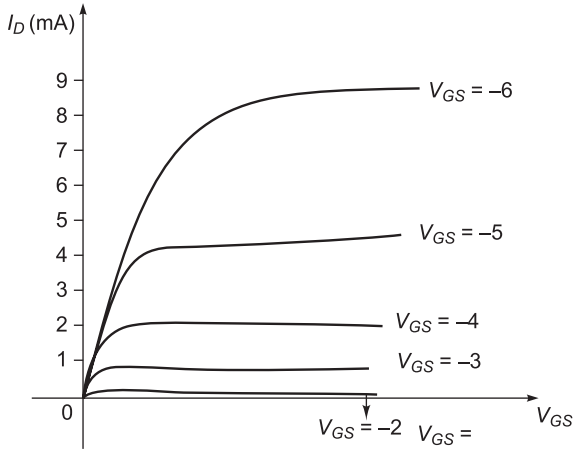


Fig. 4.25 Drain characteristics of EMOSFET *P*-channel

✦ Channel Formation and Operation

Let a positive voltage be applied between the gate and source as shown in the circuit diagram of Fig. 4.26.

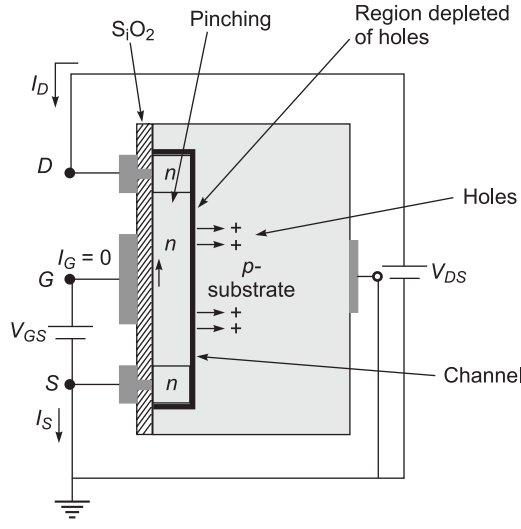


Fig. 4.26 Pinch-off $V_{GS} > V_p$, $V_{DS} > V_{DS(sat)}$

The positive gate pushes the holes in the region underneath the gate into the *P*-substrates, while the minority electrons are attracted more into the regions just below the SiO_2 layer. The result is the formation of an *N*-channel as shown in Fig. 4.26. As V_{GS} is increased, the concentration of electrons in the channel goes up, until $V_{GS} = V_T$, called *threshold voltage*. The channel begins to conduct with positive V_{DS} applied. There is a thin layer depleted of holes at the contact of *N*-channel and *P*-substrate.

Keeping V_{GS} constant above V_T, as V_{DS} is increased, a non-uniform depletion region widens at the *D*-end of the channel which becomes narrow as shown in Fig. 4.26. Further, as per KVL,

$$V_{GD} = V_{GS} - V_{DS}$$

At constant V_{GS}, increasing V_{DS} means V_{GD} reduces, i.e. gate becomes less positive than drain. For example, $V_{GS} = 8$ V, V_{DS} is increased from 2 to 5 V. V_{GD} reduces from 6 V to 3 V.

As a consequence, as in JFET, I_D reaches a saturation at a smaller value of $V_{DS(sat)}$. This is brought out in the drain characteristics of Fig. 4.27.

- It can be shown that the saturation level $V_{DS(sat)}$ is related to V_{GS} as

$$V_{DS(sat)} = V_{GS} - V_T \tag{4.5}$$

- For $V_{GS} > V_T$, it can be shown from the saturation region of the drain characteristics that I_D can be found from the analytical expression

$$I_D = k\,(V_{GS} - V_T)^2 \tag{4.6}$$

It is a nonlinear relationship which is obvious from Fig. 4.27 in which the characteristic spacing widens for equal increment of V_{GS}.

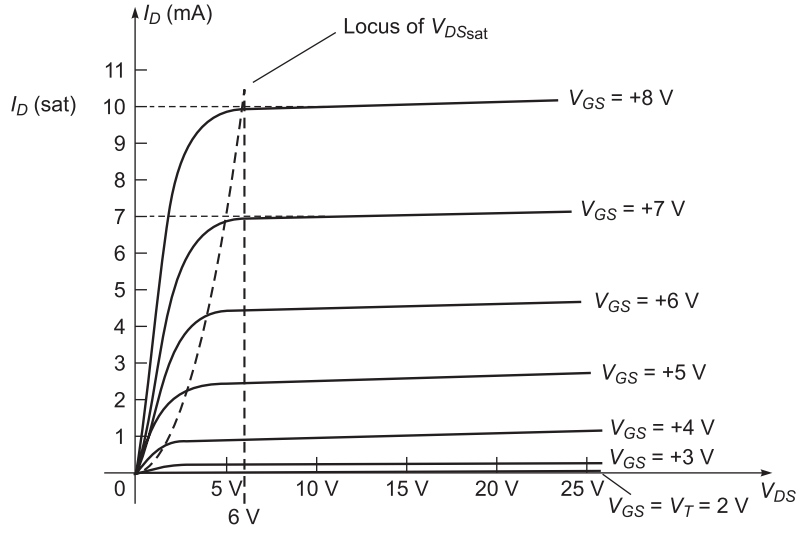


Fig. 4.27 Drain characteristics of EMOSFET

From any point on the drain characteristics, the constant k in Eq. (4.6) can found as

$$k = \frac{I_{D(\text{on})}}{(V_{GS(\text{on})} - V_T)^2} \text{ A/V}^2 \tag{4.7}$$

The expression (4.6) is indeed the transfer characteristic. A typical plot is drawn in Fig. 4.28. It can also be plotted from the drain characteristics by reading $I_{D(\text{sat})}$ against V_{GS}. V_T is also available there.

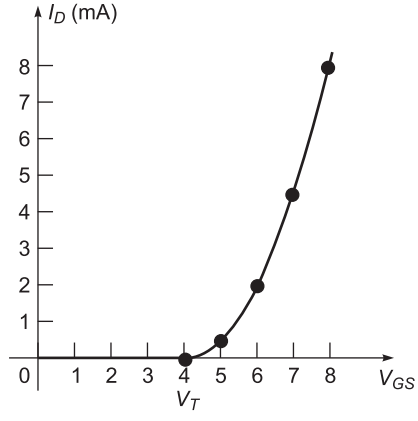


Fig. 4.28 Transfer characteristic of EMOSFET N-channel $k = 0.5 \times 10^{-3}$ A/V^2

✦ *P*-type EMOSFET

Substrate is *N*-type, Channel is *P*-type, V_{GS} negative, V_T negative, *D* – and S +.

✦ Symbol of EMOSFET (Fig. 4.29)

The channel is shown broken as it is formed on applying $V_{GS} > V_T$ for *N*-Channel.

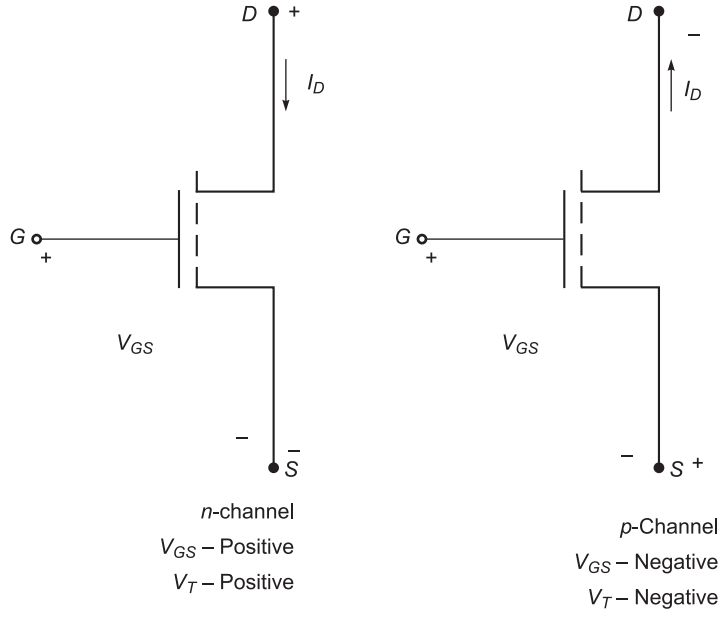


Fig. 4.29 Symbol of EMOSFET

EXAMPLE 4.7 For a depletion-type MOSFET,

I_D = 10 mA at V_{GS} = —1 V

Determine V_P if I_{DSS} = 15 mA.

Solution

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

$$10 = 15\left(1 + \frac{1}{V_P}\right)^2$$

Solving $V_P = -5.45$ V

EXAMPLE 4.8 For DMOSFET,

$I_D = 4.5$ mA, $V_{GS} = -2$ V

Determine I_{DSS} if $V_P = -5$ V.

Solution

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

$$4.5 = I_{DSS}\left(1 - \frac{-2}{-5}\right)^2$$

$$I_{DSS} = \frac{4.5}{(0.6)^2} = 12.5 \text{ mA}$$

EXAMPLE 4.9 For EMOSFET,

$V_T = 4$ V, $I_{D(on)} = 4$ mA, $V_{GS(on)} = 6$ V

Determine to write the general expression for I_D and find I_D for $V_{GS} = 8$ V.

Solution

$$k = \frac{I_{D(on)}}{(V_{GS(on)} - V_T)^2} = \frac{4 \times 10^{-3}}{(6-4)^2} = 1 \times 10^{-3} \text{ A/V}^2$$

Then

$$I_D = k(V_{GS} - V_T)^2$$

$$I_D = 1 \times 10^{-3}\,(8-4)^2 = 16 \text{ mA}$$

EXAMPLE 4.10 For EMOSFET,

$k = 0.45 \times 10^{-3}$ A/V^2, $I_{D(on)} = 3.5$ mA, $V_{GS(on)} = 4.5$ V

Determine V_T.

Solution

$$k = \frac{I_{D(on)}}{(V_{GS(on)} - V_T)^2}$$

$$0.45 = \frac{3.5}{(4.5 - V_T)^2}$$

$$(4.5 - V_T)^2 = \frac{3.5}{0.45}$$

$$4.5 - V_T = 2.79$$

$$V_T = 1.71 \text{ V}$$

4.4 FET CIRCUIT CONFIGURATIONS

As FETs are three-terminal devices, there are three possible configurations in which these can be connected.

✦ Common-Source (CS) Circuit

The input signal is connected to gate which is the control terminal. Output is taken from the drain across suitable impedance and source is the common terminal, which is grounded as shown in Fig. 4.30. This is the most useful and commonly used connection for voltage amplification.

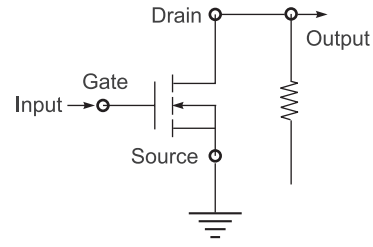


Fig. 4.30 Common-source connection

As in case of small signal models the capacitances are ignored, but at high frequency, the internal capacitances between each of the device's terminals can no longer be ignored. The JFET implementation of the common-source amplifier is similar to common-emitter amplifier. From Fig. 4.31, R_1 and R_2 makes the voltage divider circuit. C_1 and C_2 are the coupling capacitors that provide the ac grounds to the input signals and used to couple the input voltage to output voltage. C_S is the bypass capacitor. This external coupling and bypass capacitors are large enough that we can model them as short circuits for high frequencies. Common-source configuration is very important and is widely used. This configuration provides good impedance, good voltage gain and moderate output impedance. Equivalent circuit of common-source amplifier is drawn by capacitors with short circuits and by reducing the dc supply to zero.

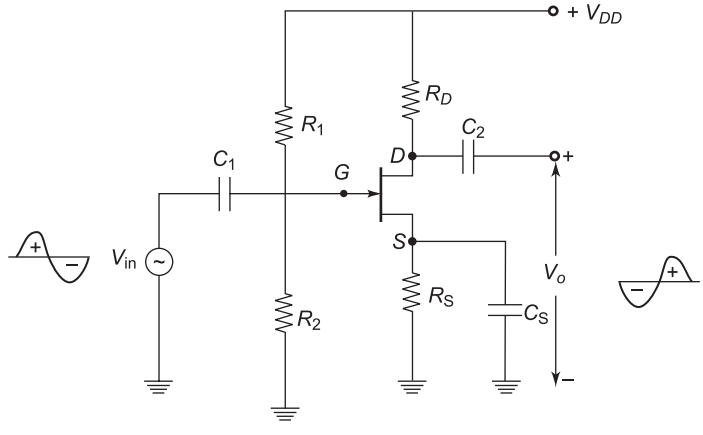


Fig. 4.31 Common-source amplifier at high frequency

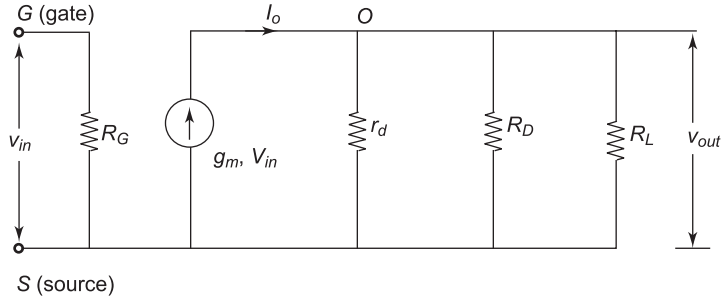


Fig. 4.32 ac equivalent circuit for common source amplifier

❑ *Common Source Amplifier Analysis*

(i) *Input resistance*

As $I_g = 0$, R_{gs} is infinite ideally or extremely high.

$$R_{\text{in}} = R_G \parallel R_{gs} \approx R_G$$

(ii) *Output resistance*

$$Z_d \approx r_d$$

Z_d is the output impedance of device

Circuit output impedance $R_D \parallel Z_d$

$$Z_{\text{out}} = R_D \parallel Z_d = R_D \parallel r_d$$

as $r_d >> R_D$

(iii) *Voltage gain*

$$V_{\text{out}} = I_D (r_d \parallel R_D \parallel R_L)$$

$$= -g_m V_{in} (r_d \parallel R_D \parallel R_L)$$

$$A_V = \frac{V_{\text{out}}}{V_{\text{in}}} = -g_m (r_d \parallel R_D \parallel R_L)$$

as $r_d >> R_D \parallel R_L$

$$A_V = -g_m (R_D \parallel R_L).$$

✦ Common-Gate (CG) Circuit

In this connection, the gate is the common terminal, the drain is the output terminal and source is where the input is connected as shown in Fig. 4.33. It hardly yields any gain and is not used in practice.

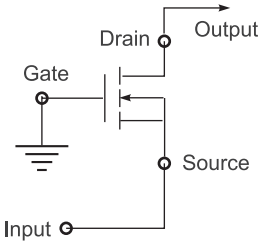


Fig. 4.33 Common-gate connection

✦ Common-Drain (CD) Circuit

The input signal is applied at the gate, the drain is the common grounded terminal, while output is taken from the source across an impedance. This connection is shown in Fig. 4.34. It has special applications and is known as *source-follower* as the output signal has the same phase as the input signal.

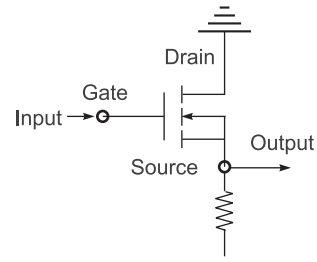


Fig. 4.34 Common-drain circuit

Note: Figures 4.30, 4.33 and 4.34 do not include biasing details, which will be explained in the following part of this chapter.

Just like common-source, common-drain configuration itself provides a high level of buffering and high input impedance. The actual input resistance of the FET is very high as it is a field effect device. This means that the source follower circuit is able to provide excellent performance as a buffer. Current gain for common-drain amplifier is high and the voltage gain is unity. Common-drain circuit is also known as source follower. In Fig. 4.35, the input signal is applied to gate using coupling capacitor C_1. The external load R_1 is connected to source through coupling capacitor C_2.When small ac signal is applied to gate, variations produced in gate source voltage causes variations in drain current. With increase in gate source voltage, the drain current also increases. Hence the voltage drop across source resistor and output voltage increases.

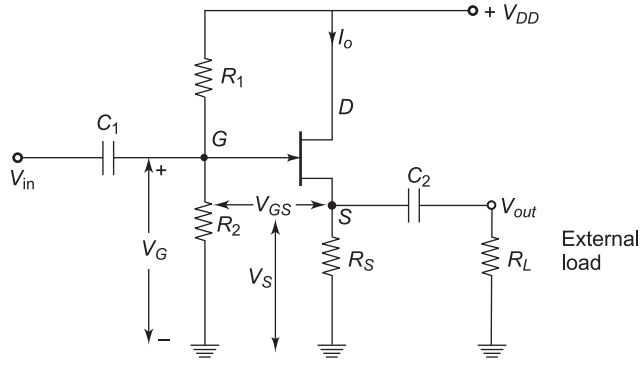


Fig. 4.35 Common-drain amplifier circuit

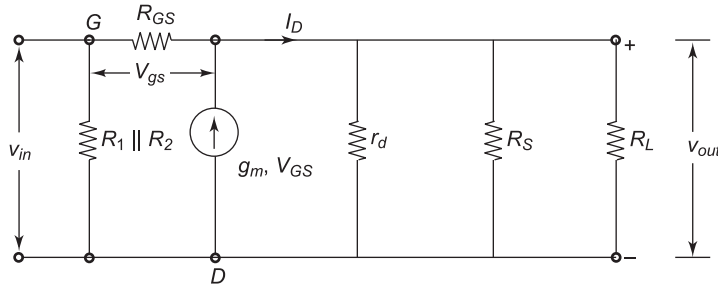


Fig. 4.36 ac equivalent circuit for common-drain amplifier

❑ *Common-drain Amplifier Circuit*

(i) *Voltage gain*

$$V_{GS} = V_{\text{in}} - V_{\text{out}}$$

$$R_G = R_1 \parallel R_2$$

$$V_{\text{out}} = I_D (r_d \parallel R_S \parallel R_L)$$

$$V_{\text{out}} = g_m V_{GS} (r_d \parallel R_S \parallel R_L)$$

$$V_{\text{in}} = V_{GS} + V_{\text{out}} = V_{gs} + g_m V_{gs} (r_d \parallel R_S \parallel R_L)$$

$$A_V = \frac{g_m (R_S \parallel R_L)}{1 + g_m (R_S \parallel R_L)} \quad \text{as } r_d >> R_S \parallel R_L$$

If $g_m (R_S \parallel R_L) >> 1$ then $A_V = 1$

(ii) *Input Impedance*

$$R_G = R_1 \parallel R_2$$

(iii) *Output Impedance*

As $$V_{gs} = V_{\text{out}} \frac{R_{gs}}{(R_S \parallel R_G) + R_{gs}}$$

Normally $R_{gs} >> R_S \parallel R_G$, $V_{gs} = V_{out}$

$$I_D = g_m V_{gs}$$

$$Z_S = \frac{V_{\text{out}}}{I_D} = \frac{V_{gs}}{g_m V_{gs}} = 1/g_m$$

as $r_d >> 1/g_m$, so neglected.

Output Impedance is

$$Z_{\text{out}} = R_S \parallel 1/g_m$$

EXAMPLE 4.11 For given g_m of 6 ms and drain resistance 1.7 kΩ, find value of ideal voltage gain.

Solution

$$A_V = -g_m . R_D$$
$$= -10.2$$

EXAMPLE 4.12 For Fig. 4.37, calculate the voltage gain. Given that:

g_m = 4 mA/V, r_d = 45 kΩ, R_D = 20 kΩ, and R_G = 100 MΩ

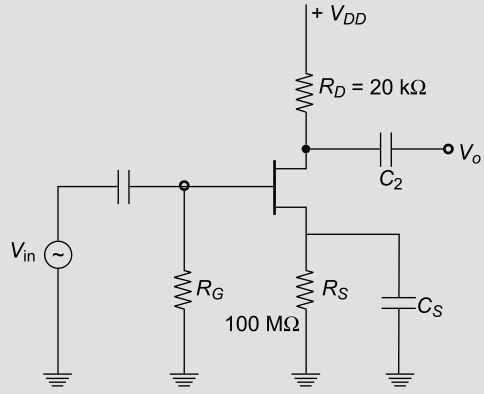


Fig. 4.37

Solution

$$A_V = -g_m r_L$$

$$r_L = R_D \parallel r_d = \frac{R_D \times r_d}{R + r_d} = \frac{20 \times 45}{20 + 45} = 13.8 \times 10^3\ \Omega$$

$$A_V = (-4 \times 10^{-3})\ (13.8 \times 10^3) = -55.2$$

$$R_i = R_G = 100\ \text{M}\Omega$$

$$R_D = R_i = 13.8 \times 10^3$$

EXAMPLE 4.13 For Fig. 4.38, find voltage gain of Amplifier, input and output resistance

Given that: g_m = 600 µs, Input Impedance = ∞, and Output Impedance = neglect

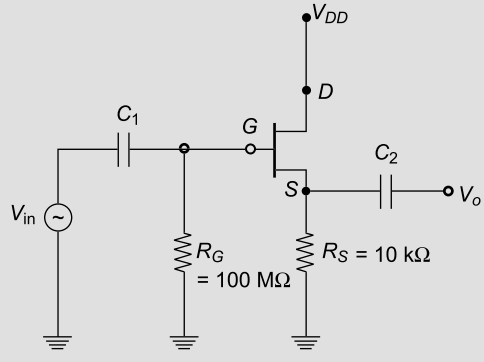


Fig. 4.38

Solution

(i) $A_V = \dfrac{R_S}{R_S + 1/g_m} = \dfrac{10 \times 10^3}{(10 \times 10^3) + 1666.6}$

(ii) Input resistance $R_i = R_G = 100$ mΩ

(iii) Output resistance = $R_D = 1/g_m = 166.6$ Ω

4.5 CMOS CIRCUITS

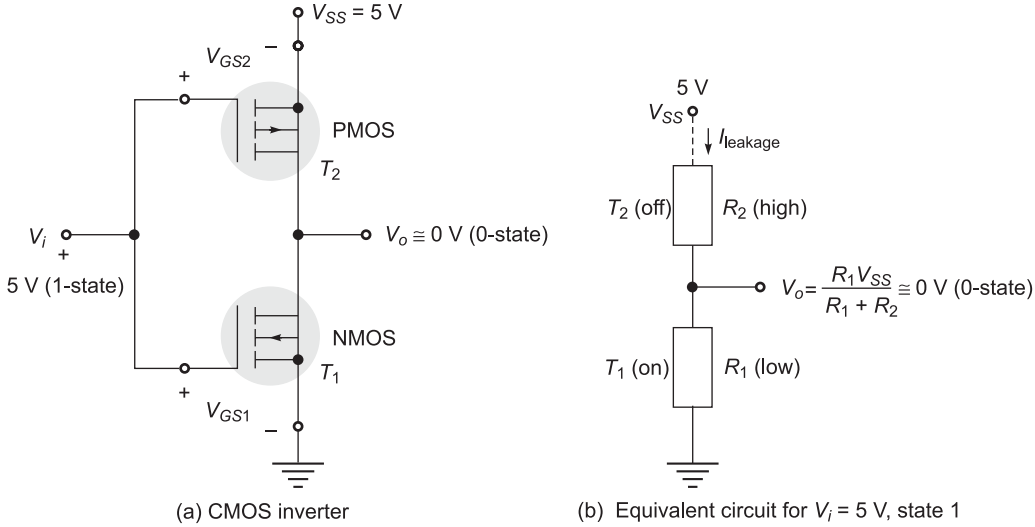
CMOS is the complimentary MOS wherein two enhancement MOSFETs, one *N*-type (*N*MOS) and other *P*-type (*P*MOS), are connected as a complimentary pair. The two gates are connected to form the input terminal and the two drains are connected to form the output terminals as shown in Fig. 4.20(a). The CMOS circuit offers two advantages:

1. The drain current is very low and flows mainly during transition from one state to the other (ON/OFF).
2. The power drawn in steady state is extremely small.

Because of these two advantages, it has gained great popularity in digital circuits. It also has certain applications in analog circuits.

Digital-Circuit Applications

The circuit for a CMOS digital inverter is drawn in Fig. 4.39(a). The source terminal of *P*MOS (T_2) is connected to V_{SS} = 5 V, while the source terminal of *N*MOS (T_1) is grounded.



(a) CMOS inverter (b) Equivalent circuit for V_i = 5 V, state 1

Fig. 4.39 CMOS circuits

✦ Operation

1. Input

$$V_i = 5 \text{ V} \Rightarrow \text{1-state}$$

$$V_{GS2} = 5 - 5 = 0 \text{ V}$$

T_2 is nonconducting, OFF (its V_T is negative), draws only leakage current, offers high resistance (R_2)

$$V_{GS1} = 5 \text{ V} > V_T$$

T_1 is conducting, ON; offers very low resistance (R_1). The circuit equivalent in this state is drawn in Fig. 4.39(b).

Output

$$V_o \approx 0 \Rightarrow \text{0-state}$$

It can be seen from the circuits of Fig. 4.39(b) that

$$V_o = \frac{R_1}{R_1 + R_2} V_{SS} \approx 0 \text{ V}$$

2. Input

$$V_i = 0 \text{ V} \Rightarrow \text{0-state}$$

$$V_{GS2} = -5 \text{ V}, T_2 \text{ conducting (low resistance)}$$

$$V_{GS1} = 0 \text{ V}; T_1 \text{ nonconducting (high resistance)}$$

Output

$$V_o \approx 5 \text{ V} \Rightarrow \text{1-state}$$

We thus see that the circuit acts as an inverter; 1-state input produces 0-state output and 0-state input produces 1-state output.

It is observed in this circuit that only one transistor is turned on in any of the output states. As the transistors are series connected, no current is drawn from the battery source in either of the two states. Current is drawn from the battery only during state transition (either way). CMOS circuits, therefore, draw extremely low power from the battery source and so their energy consumption is very small. This is the major attraction why CMOS is used in digital applications.

4.6 | FET BIASING

We shall consider only voltage-divider biasing as this is most commonly adopted. By examining the drain characteristics of the device, a Q-point is selected in the middle of the saturation region, which fixes V_{GSQ} and I_{DQ}. The biasing circuit resistors are to be selected for the device under dc conditions to operate at the Q-point.

4.6.1 Voltage Divider Biasing

The circuit is drawn in Fig. 4.40. It is the same for any FET.

As per voltage divider

$$V_G = \left(\frac{R_2}{R_1 + R_2}\right) V_{DD} \tag{4.8}$$

Then

$$V_{GS} = V_G - I_D R_S,\ I_S = I_D \tag{4.9}$$

$I_D R_S$ provides stabilising negative voltage feedback.

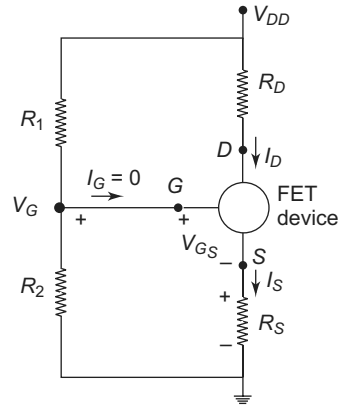


Fig. 4.40

Device transfer characteristic,

$$I_D = f(V_{GS})\Big|_{\text{Device Parameters}};\ \text{nonlinear} \tag{4.10}$$

As V_{GSQ} has been chosen, the choice of R_S and simultaneous solution of Eqs (4.8) and (4.9) yields I_{DQ}.

KVL for DS load yields

$$V_{DSQ} = V_{DQ} - I_D(R_S + R_D),\ R_D \text{ has to be selected} \tag{4.11}$$

Biasing analysis/design is then complete.

✦ Simultaneous Solution of Eqs. (4.9) and (4.10)

(i) Equation (4.10) is the transfer characteristic of FET (nonlinear function of V_{GS}). Equation (4.9) is a straight line, whose intersection with the transfer characteristic yields I_{DQ} and V_{GSQ}.

or

(ii) Substituting I_D from Eq. (4.9) in Eq. (4.10) leads to a quadratic equation in V_{GS} yielding two solutions from which the appropriate one is to be chosen.

✦ JFET

Refer Fig. 4.40.

JFET: $I_{DSS} = 10$ mA, $V_P = -6$ V

Circuit: $R_1 = 2.2$ MΩ, $R_2 = 280$ kΩ, $V_{DD} = 16$ V

$R_D = 2$ kΩ, $R_S = 1.5$ kΩ

To determine at Q-point

V_{GS}, I_D, V_{DS}, V_{DG}

Plot the transfer characteristic (Fig. 4.41).

$$I_D = 10\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{4.12}$$

$$V_{GS} = -3\text{ V},\ \frac{V_{GS}}{V_P} = \frac{1}{2},\ I_D = 2.5\text{ mA}$$

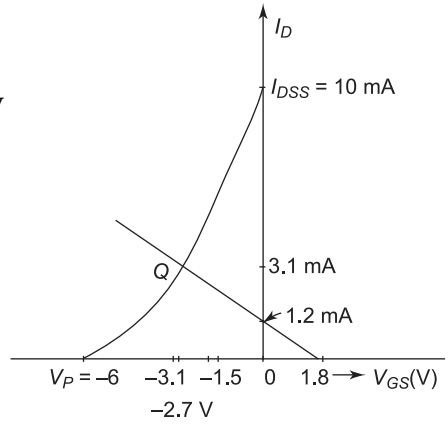


Fig. 4.41 Transfer characteristic

$$V_{GS} = 1.5 \text{ V}, \frac{V_{GS}}{V_P} = \frac{1}{4}, I_D = 5.265 \text{ mA}$$

Plot Eq. (4.7).

$$V_G = \frac{280}{2.2 \times 10^3 + 280} \times 16 = 1.8 \text{ V}$$

$$V_{GS} = 1.8 - 1.5 I_D \tag{4.13}$$

$$I_D = 0, V_{GS} = 1.8 \text{ V}$$

$$V_{GS} = 0, I_D = \frac{1.8}{1.5} = 1.2 \text{ mA}$$

At intersection, Q-point $\Rightarrow V_{GS} = -2.7$ V $I_D = 3.1$ mA

$$V_{DS} = 16 - (2 + 1.5) \times 3.1 = 5.15 \text{ V}$$

✦ Analytic Approach

From Eq. (4.13),

$$I_D = \frac{1.8 - V_{GS}}{1.5} = 1.2 - 0.67 V_{GS} \tag{4.14}$$

Substituting I_D in Eq. (4.12),

$$1.2 - 0.67 V_{GS} = 10 \left(1 - \frac{V_{GS}}{-6}\right)^2$$

Let $V_{GS} = x$

$$1.2 - 0.67x = 10\left(\frac{6 + x}{6}\right)^2 = \frac{10}{36} (6 + x)^2$$

$$4.32 - 2.41x = x^2 + 12x + 36$$

$$x^2 + 14.41x + 31.68 = 0$$

$x = -2.7, -11.7$ (rejected as more negative than V_P)

Then $V_{GS} = -2.7$ V

The same result as obtained by the graphical solution.

✦ DMOSFET

Refer Fig. 4.40.

DMOSFET parameters

$$I_{DSS} = 8 \text{ mA}, V_p = -4$$

Circuit data

$$V_{DD} = 16 \text{ V}$$

$$R_1 = 100 \text{ M}\Omega, R_2 = 10 \text{ M}\Omega,$$

$$R_D = 1.6 \text{ k}\Omega, R_S = 700 \ \Omega$$

To determine

$$I_{DQ},\ V_{GSQ},\ V_{DSQ}$$

Plotting transfer characteristic (Fig. 4.42),

$$I_D = I_{DSS}\left(1-\frac{V_{GS}}{V_P}\right)^2$$

$$V_{GS} = -2\text{ V};\ I_D = 8\left(1-\frac{1}{2}\right)^2 = 2\text{ mA}$$

$$V_{GS} = -1\text{ V},\ I_D = 8\left(1-\frac{1}{4}\right)^2 = 4.5\text{ mA}$$

$$V_{GS} = +1\text{ V},\ I_D = 8\left(1+\frac{1}{4}\right)^2 = 12.5\text{ mA}$$

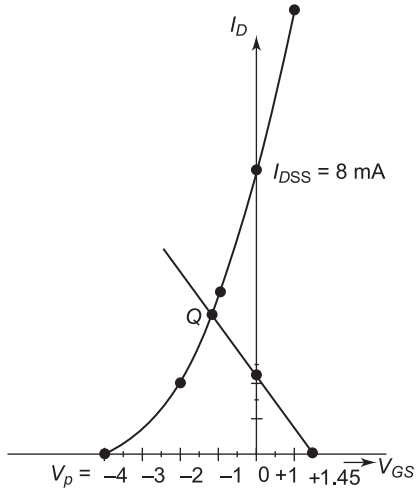


Fig. 4.42 Transfer characteristic

Plotting Eq. (4.8),

$$V_G = \frac{10}{100+10}\times 16 = 1.45\text{ V}$$

$$V_{GS} = V_G - 0.7\,I_D$$

$$I_D = 0 \qquad V_{GS} = V_G = 1.45\text{ V}$$

$$V_{GS} = 0 \qquad I_D = \frac{1.45}{0.7} = 2.07\text{ mA}$$

At intersection, Q-point

$$I_{DQ} = 3.8\text{ mA},\ V_{GSQ} = -1.2\text{ V}$$

From Eq. (4.10),

$$V_{DSQ} = 16 - (1.6 + 0.7)\times 3.8 = 7.26\text{ V}$$

✦ EMOSFET

Refer Fig. 4.40.

EMOSFET parameters

$$V_T = 4\text{ V},\ V_{GS(on)} = 8\text{ V},\ I_{D(on)} = 2.5\text{ mA}$$

Biasing circuit

$$V_{DD} = 35\text{ V},\ R_1 = 20\text{ M}\Omega,\ R_2 = 16\text{ M}\Omega$$

$$R_D = 2.5\text{ k}\Omega,\ R_S = 0.75\text{ k}\Omega$$

To determine: At Q-point

$$I_{DQ},\ V_{GSQ},\ V_{DS}$$

Plot transfer characteristic (Fig. 4.43).

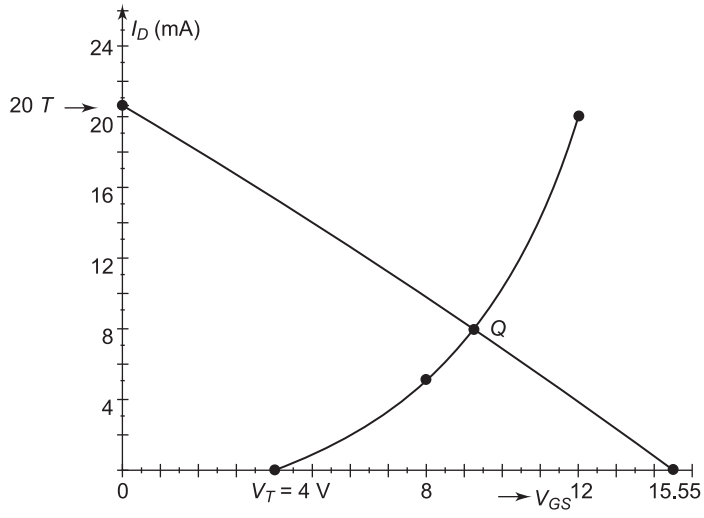


Fig. 4.43

$$k = \frac{2.5}{(8-4)^2} = 0.156 \text{ mA/V}^2$$

$$I_D = 0.156 (V_{GS} - 4)^2$$

$$V_{GS} = 8 \text{ V}, I_D = 0.156 \times 16 = 2.5 \text{ mA}$$

$$V_{GS} = 12 \text{ V}, I_D = 0.156 \times 64 = 10 \text{ mA}$$

Plot Eq. (4.8).

$$V_G = \frac{16}{20+16} \times 35 = 15.55 \text{ V}$$

$$V_{GS} = 15.55 - 0.75 I_D$$

$$V_{GS} = 0, I_D = 20.7 \text{ mA}$$

$$I_D = 0, \ V_{GS} = 15.55$$

At Q-point,

$$V_{GSQ} = 9.2 \text{ V}, I_{DQ} = 8.4 \text{ mA}$$

From Eq. (4.11),

$$V_{DS} = 35 - (2.5 + 0.7) \times 8.4 = 8.12 \text{ V}$$

✦ p-channel JFET

Refer Fig. 4.40 directions of I_D and I_S reverse, V_{DD} is negative, JFET parameters I_{DSS} = 10 mA, V_P = 6 V (positive).

Biasing circuit

$$V_{DD} = -20 \text{ V}, R_1 = 64 \text{ k}\Omega, R_2 = 20 \text{ k}\Omega$$

$$R_D = 2 \text{ k}\Omega. Rs = 1.2 \text{ k}\Omega$$

To determine at *Q*-point

$$V_{GSQ},\ I_{DQ}, V_{DS}$$

Shockley's equation

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

$$I_D = 10\left(1 - \frac{V_{GS}}{6}\right)^2$$

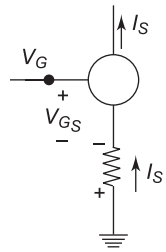


Fig. 4.44

Voltage divider

$$V_G = \frac{20}{64 + 20} \times (-20) = -4.76 \text{ V}$$

KVL for *GS* loop (adjoining Fig. 4.44)

$$V_G - V_{GS} + R_S I_D = 0;\ I_D - I_S; \text{ direction of currents } S \text{ to } D$$

$$-4.76 - V_{GS} + 1.2\ I_D = 0;\ I_D = I_S$$

or
$$I_D = \frac{4.76 + V_{GS}}{1.2} = 4 + 0.833 V_{GS} \qquad (4.15)$$

Let $V_{GS} = x$,

Substituting I_D in Eq. (4.15),

$$4 + 0.833x = \frac{10}{36}(6 - x)^2$$

$$14.4 + 3x = 36 - 12x + x^2$$

$$x^2 - 15x + 21.6 = 0 \Rightarrow x = 1.6 \text{ V}, 13.476 \text{ V (rejected)}$$

Then,

$$V_{GSQ} = 1.6 \text{ V}$$

$$I_{DQ} = 10\left(1 - \frac{1.6}{8}\right)^2 = 6.4 \text{ mA}$$

KVL *DS* loop,

$-20 + (2 - 1.2) \times 5.38 - V_{DS} = 0$

or
$$V_{DS} = -2.78 \text{ V}$$

SUMMARY

Transfer characteristics

➢ **JFET**

$I_D = I_S$, $I_G \approx 0$

N-channel $I_D = I_{DSS}\left(1-\frac{V_{GS}}{V_P}\right)^2$

$R_{in} > 100\ M\Omega$, $I_{DSS} = I_D$ at $V_{GS} = 0$

V_P = pinch-off voltage, negative value

➢ **DMOSFET**

$I_G = 0$, $I_D = I_S$

Depletion type, $I_D = I_{DSS}\left(1-\frac{V_{GS}}{V_P}\right)^2$

N-channel can conduct for $V_{GS} > 0$.

I_D rises very sharply

$R_{in} > 10^{10}\ \Omega$ (not used in this mode)

➢ **MOSFET**

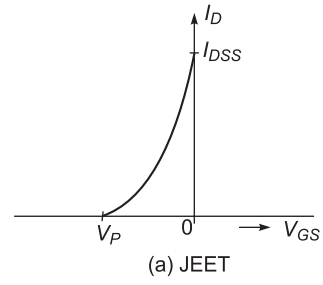
$I_D = k(V_{GS} = V_T)^2$

Enhancement type $V_T = V_{GS}$ (Threshold)

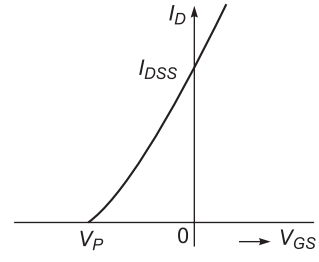
N-channel $k = \frac{I_{D(on)}}{(V_{GS(on)} - V_T)^2}$

$R_{in} > 10^{10}\ \Omega$

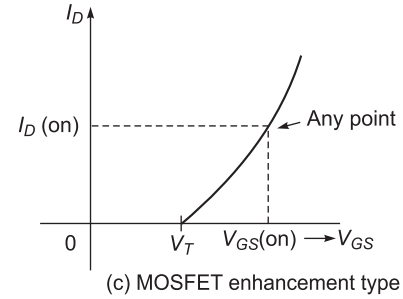
Remark for p channel V_{GS} and N-channel I_D reverses.



(a) JEET



(b) MOSFET depletion type



(c) MOSFET enhancement type

Fig. 4.45 Transfer characteristic of various types of FETs

EXERCISES

➥ Review Questions

1. Sketch an n-channel JFET. Explain how the pinch-off takes place for $V_{GS} = 0$. What is meant by I_{DSS}?
2. Draw the symbols for n-and p-channel JFETs. Indicate the polarity of voltages and direction of current.
3. Define the pinch-off voltage V_P.
4. Explain why $I_G \approx 0$ for a JFET.
5. What is meant by a field effect transistor? How it is different from BJT?

6. Explain the various regions of *n*-channel JFET. Comment in the space of saturation region drain current with respect to $V_{GS.}$
7. Why are *N*MOS devices preferred to *P*MOS devices?
8. Why is the input resistance of MOSFET larger than that of a JFET depletion type?
9. Sketch the constructional view of *p*-channel depletion-type MOSFET. Explain the form of the drain characteristic for $V_{GS} = 0$.
10. In what ways is the construction of depletion-type MOSFET similar to that of JFET and in what ways is it different?
11. Explain the construction difference between enhancement and depletion-type MOSFET.
12. What is the meaning and significance of V_T?
13. Draw the symbol of *N*- and *P*-type EMOSFETs. Indicate the polarity of voltages and direction of currents.
14. Sketch and compare the transfer characteristics of *N*-channel DMOSFET and EMOSFET.
15. Why small signal analysis is performed for FET?
16. List the advantages of MOSFET and BJT.
17. Justify the direction of flow of current for *p*-channel MOSFET when voltage is applied at the gate.
18. Define drain resistance and transconductance.
19. List the advantages and drawbacks of MOSFET.

➥ Problems

1. From the JFET characteristics of Fig. 4.5, determine I_D for $V_{CS} = -0.5$ V and $V_{GS} = -1.5$ V.
2. From the JFET characteristics, determine the device resistance at the initial part of the characteristics for $V_{GS} = 0$ and $V_{GS} = -1$ V.
3. Plot the transfer characteristic for the JFET characteristics of Fig. 4.8.
4. For the characteristics of Fig. 4.8, determine V_{GS} for $I_D = 5$ mA.
5. Given $I_{DSS} = 6$ mA, $V_P = -3$ V. Determine the drain current at $V_{GS} = -1, 0, 1$ and 2 V. What conclusions do you draw?
6. Given $I_D = 12$ mA and $V_{GS} = 1$ V. Determine V_P if $I_{DSS} = 8.5$ mA.
7. Given $I_D = 4.5$ mA at $V_{GS} = -2$ V. Determine I_{DSS} if $V_P = -5$ V.
8. Sketch the transfer and drain characteristic for *N*-channel depletion-type MOSFET with $I_{DSS} = 10$ mA and $V_P = -6$ V, range of $V_{GS} = -V_P$ to $V_{GS} = 1$.
   Hint. Take $V_{GS} = -6$ V, – 4 V, –2 V, 0 V and 1 V. Calculate I_D for each value of V_{CS} and then sketch.
9. Given $V_P = -4$ V, $I_D = 2.5$ mA at $V_{GS} = -2$ V. Calculate I_{DSS}. Also calculate I_D at $V_{GS} = +1$ V
10. Given $V_T = 4$ V, $I_{D(on)} = 4$ mA, $V_{GS(on)} = 6$ V. Determine k and then find V_{GS} for $I_D = 12$ mA.
11. Given $k = 0.4$ mA/V^2, $I_{D(on)} = 4$ mA, $V_{GS(on)} = 8$ V. Determine V_T.
12. A *P*-channel EMOSFET has $V_T = -6$ V and $k = 0.5$ mA/V^2. Determine I_D at $V_{GS} = -10$ V. Sketch the transfer characteristics.
13. The transfer characteristic of an *n*-channel EMOSFET is drawn in Fig. 4.46. Determine k and T. Write the general expression for I_D.
    $k = 0.5$ m A/V^2, $V_T = 2$ V.
14. For the fixed-bias circuit of Fig. 4.47, determine
    (a) I_{DQ} and V_{GSQ} (b) V_{DS}

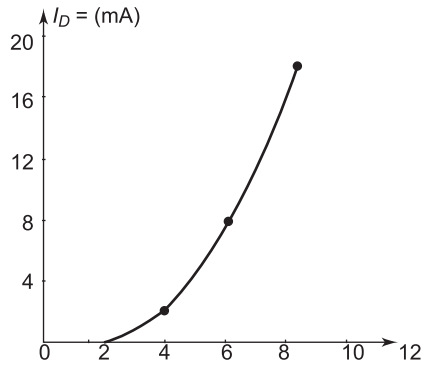


Fig. 4.46

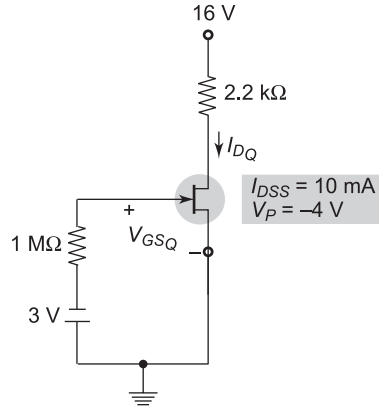


Fig. 4.47

15. For the self-bias circuit of Fig. 4.48, determine
(a) I_{DQ}, V_{GSQ} (b) V_{DS} (c) I_{DSS} = 119 mA

16. For the voltage-divider bias circuit of Fig. 4.49, determine
(a) V_{GSQ}, and I_{DQ} (b) V_{DS}

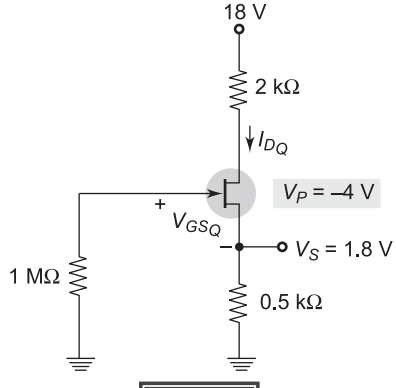


Fig. 4.48

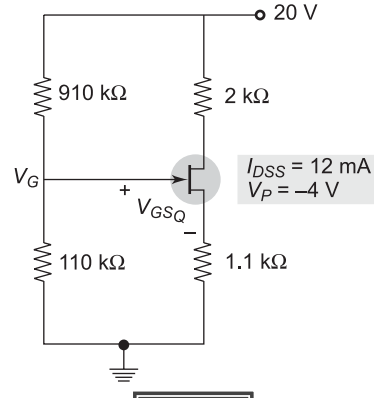


Fig. 4.49

17. For the fixed-bias configuration of Fig. 4.50, determine
V_{GSQ}, I_{DQ}, V_{DS}
Given I_{DSS} = 8 mA V_P = –5 V

18. The voltage-divider bias has the device parameters
V_T = 4 V, k = 2 mA/V^2
The biasing circuits are drawn in Fig. 4.51. Find V_{GSQ}, I_{DA}, V_{DS}.

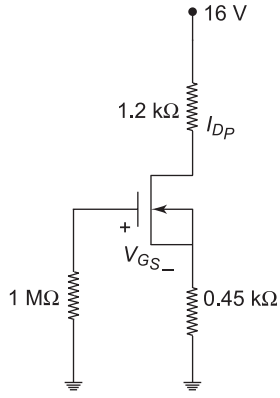


Fig. 4.50

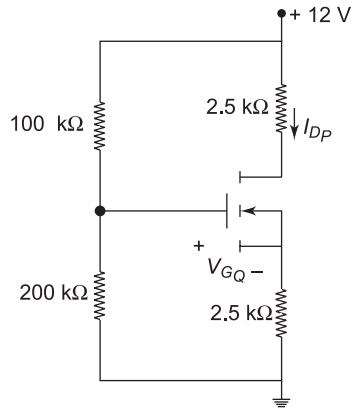


Fig. 4.51

➥ Multiple-Choice Questions

1. The best location for setting a *Q*-point on dc load line of an FET amplifier is at
(a) mid-point (b) saturation point (c) cut-off point (d) none of these

2. Which of the following techniques is used for biasing the enhancement-type MOSFET?
(a) Voltage-divider bias (b) Current-source bias
(c) Collector-feedback bias (d) Self-bias

3. A common-gate amplifier has
(a) high input resistance and high output resistance
(b) low input resistance and high output resistance
(c) low input resistance and low output resistance
(d) high input resistance and low output resistance

4. If properly biased, a JFET will act as a
(a) voltage-controlled voltage source (b) current-controlled voltage source
(c) voltage-controlled current source (d) current-controlled current source

5. Which of the following bias methods provide a solid *Q*-point in JFET amplifiers?
(a) Gate bias (b) Voltage-divider bias(c) Current bias (d) Self-bias

6. The voltage of a common source JFET amplifiers depends upon its
(a) transconductance (b) amplification factor
(c) internal load resistance (d) both (a) and (c)

7. The transconductance g_m of a JFET is equal to

(a) $\dfrac{-2I_{DSS}}{V_P}$ (b) $\dfrac{2}{|V_P|}(\sqrt{I_{DSS} \cdot I_D})$

(c) $\dfrac{-2\,I_{DSS}}{V_P}\left(1-\dfrac{V_{GS}}{V_P}\right)$ (d) $\dfrac{I_{DSS}}{V_P}\left(1-\dfrac{V_{GS}}{V_P}\right)$

8. A zero-gate bias channel resistance for a JFET is 750 Ω and the pinch-off voltage is 3 V. For a gate bias of 1.5 V and very low drain voltage, the device would have a resistance of
(a) 320 Ω (b) 1000 Ω (c) 3000 Ω (d) 1270 Ω

9. A transconductance amplifier has
(a) high input impedance and low output impedance
(b) low input impedance and high output impedance
(c) high input and output impedances
(d) low input and output impedances

10. The threshold voltage of an *n*-channel enhancement-mode MOSFET is 0.5 V. When the device is biased at a gate voltage of 3 V, pinch-off would occur at a drain voltage of
(a) 4.5 V (b) 3.5 V (c) 2.5 V (d) 1.5 V

11. If the resistance of the channel varies with input voltage then this effect is called
(a) saturation (b) polarization (c) cutoff (d) field effect

12. When $V_{GS} = 0$, then JFET is
(a) saturated (b) an analog device (c) an open switch (d) cutoff

13. When transistor cannot hold on current, then this region is called
(a) breakdown (b) saturation (c) depletion (d) pinch-off

14. The ratio of output current change to input voltage change is called
(a) transconductance (b) resistivity
(c) gain (d) impedance

15. When the JFET is no longer able to control the current, this point is called the:
(a) breakdown region (b) depletion region
(c) saturation point (d) pinch-off region

16. With a JFET, a ratio of output current change against an input voltage change is called:
(a) transconductance (b) siemens
(c) resistivity (d) gain

ANSWERS

✦ Problems

1. 6.125 mA, 3.125 mA
2. 0.22 kΩ, 1.33 kΩ
4. – 4.24 V
5. 2.67 mA, 6 mA, 10.67 mA, 16.6 mA; The device cannot be used for $V_{GS} > 0$
6. –5.31 V
7. 12.36 mA
9. 10 mA, 15.63 mA
10. $k = 1$ mA/V^2, 7.46 V
11. 4.84 V
12. 8 mA
14. (a) 1.11 mA, – 3 V (b) 13.56 V
17. –3.3 V, 7.3 mA, 4 V
18. –4.8 V, 1.28 mA, 5.6 V

✦ Multiple-Choice Questions

1. (a) **2.** (c) **3.** (b) **4.** (a) **5.** (c) **6.** (d) **7.** (c) **8.** (c) **9.** (c) **10.** (c)
11. (d) **12.** (a) **13.** (a) **14.** (a) **15.** (a) **16.** (a)

CHAPTER 5

Small-Signal Model of Transistors and Amplifiers

GOALS AND OBJECTIVES

- ❏ Introduction of BJT and amplifiers
- ❏ Analysis of BJT small-signal model—common base and common emitter configuration
- ❏ Hybrid equivalent circuit parameters of BJT
- ❏ Discussion of Norton's and Thevenin's Theorems
- ❏ Conversion formulas for the parameters of three transistor configurations
- ❏ Calculating the current gain and voltage gain of CE BJT amplifier
- ❏ FETs small-signal models and amplifier
- ❏ Discussion of the effect of source and load resistance and frequency response

5.1 INTRODUCTION

One of the important application of transistors is amplification of small-signal (ac signal). For this purpose, the transistor is biased at an operating point. The signal is then applied to the appropriate terminal of the transistor through a *coupling capacitor*, which acts as short circuit for the signal frequencies but blocks the biasing dc from disturbing the signal source. The output is then taken from the output terminal of the transistor through another coupling capacitor, so that the dc is blocked and only amplified signal is available at the output.

The transistor is carrying both dc and ac currents and voltages. The amplifier circuit analysis is rendered simple by separating dc and ac analysis. The dc analysis has been carried out in **chapters 3, 4**. The coupling capacitors act as open circuits. For ac analysis, dc sources are shorted out and coupling capacitors, bypass capacitor act as short circuit for ac frequencies.

To illustrate the procedure consider the example of a small-signal (ac) *BJT RC-coupled amplifier* of Fig. 5.1(a) wherein voltage divider bias is provided; biasing analysis is carried out by open circuiting all capacitors; coupling and bypass. For ac analysis, dc source is shorted out (grounded) and capacitor are shorted out resulting in the network of Fig. 5.1(b). Observe that R_E is shorted by the *bypass capacitor*. The transistor is now replaced by its smaller signal model at the operating point. This is the subject matter of this chapter for both BJT and FETs *RC*-coupled amplifier.

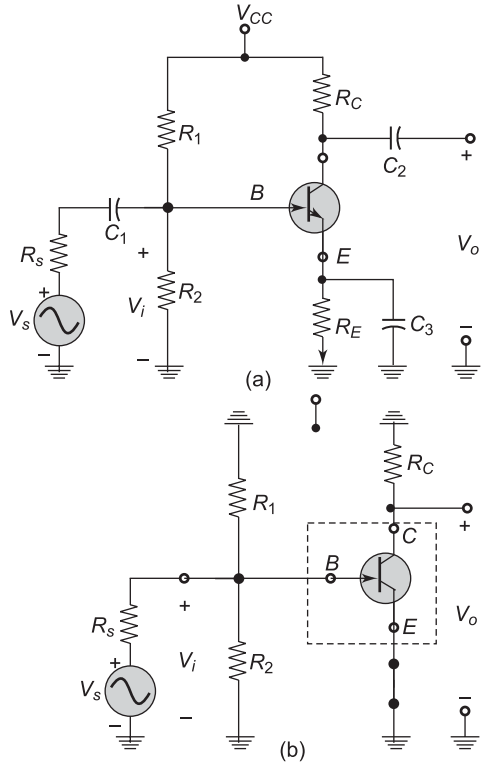


Fig. 5.1 (a) BJT RC-coupled amplifier (b) Small-signal circuit: dc grounded, capacitors shorted

5.2 BJT SMALL-SIGNAL MODEL

We will employ the r_e model.

5.2.1 Common-Base Configuration (CB)

The biased *NPN* transistor in sketched in Fig. 5.2(a). The forward-biased *BE* junction acts like a forward-biased diode as shown in Fig. 5.2(b). For small signals imposed on bias currents and voltages, the diode offer dynamic resistance r_e as shown in Fig. 5.2(c).

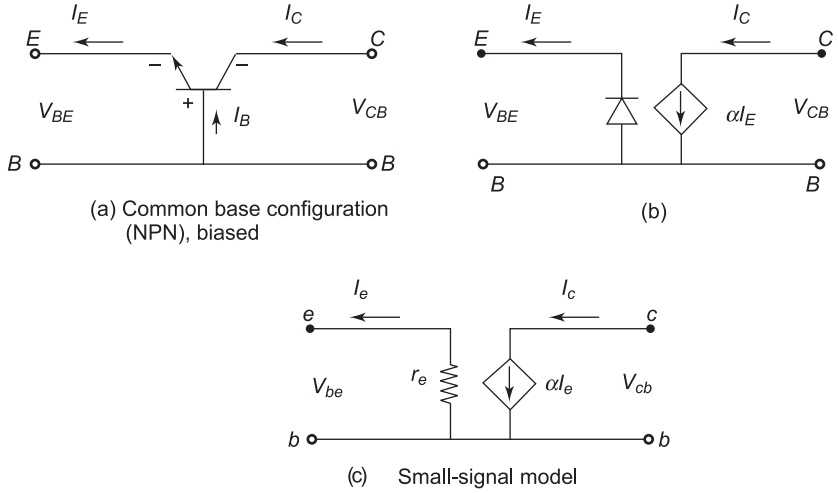


Fig. 5.2 CB configuration

As shown earlier,

$$r_e = \frac{26\text{ mV}}{I_E} = \frac{V_{be}}{I_e} \tag{5.1}$$

5.2.2 Common-Emitter (CE) Configuration

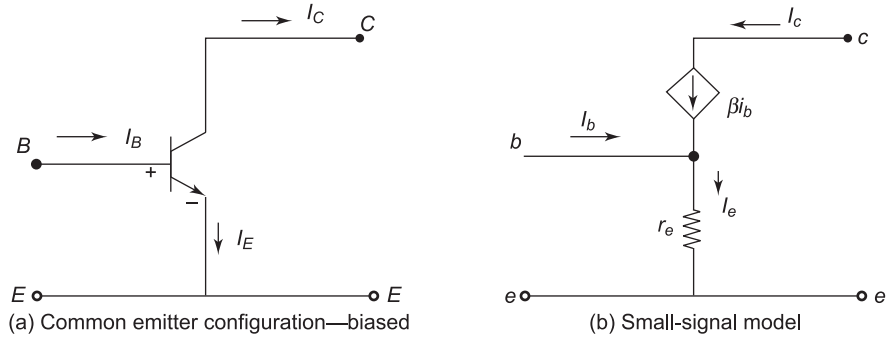


Fig. 5.3 CE configuration

The common-emitter configuration biased is shown in Fig. 5.3(a). The small-signal, forward-biased BE diode is replaced by dynamic resistance r_e carrying current I_e; $I_c = \beta I_b$. The small-signal model is drawn in Fig. 5.3(b).

Now,

$$I_e = (1 + \beta) I_b$$

So
$$V_{be} = r_e I_e = (1 + \beta) r_e I_b \approx (\beta r_e)\, I_b \qquad (5.2)$$

The circuit of Fig. 5.3(b) can be drawn as in Fig. 5.4.

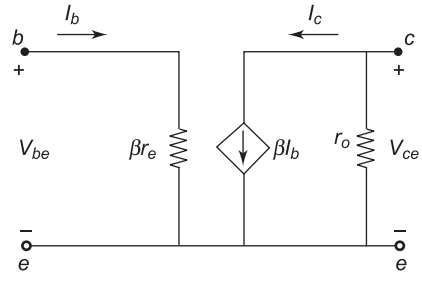


Fig. 5.4 Small-signal model—CE configuration

Now

$$\beta r_e = \beta \frac{V_{be}}{I_e} = \frac{I_c}{I_b} \cdot \frac{V_{be}}{I_e} = \frac{I_c}{I_e}\left(\frac{V_{be}}{I_b}\right)$$

$$\frac{I_c}{I_e} = \frac{I_c}{I_b + I_c} = \frac{\beta}{1 + \beta}$$

$$\beta r_e = \left(\frac{\beta}{1+\beta}\right)\frac{V_{be}}{I_b} = \left(\frac{\beta}{1+\beta}\right) r_\pi$$

$$\beta r_e \approx r_\pi; \textit{ dynamic function resistance} \qquad (5.3)$$

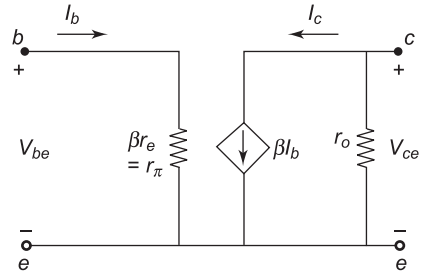


Fig. 5.5 r_π-β model

The slight upward slope of the collector characteristics contributes an output resistant at CE terminals.

$$r_o = \left.\frac{\Delta V_{CE}}{\Delta I_c}\right|_{\text{at } Q}; \text{ very high usually ignored}$$

It is shown in Figs 5.4 and 5.5.

Now
$$\beta I_b = \beta \frac{V_{be}}{\beta r_e} = \frac{V_{be}}{r_e} = g_m V_{be}$$

$$g_m = \frac{1}{r_e} = \textit{transconductance}, \text{ units S} \qquad (5.4)$$

βr_e-g_m model is shown in Fig. 5.6; r_o neglected.

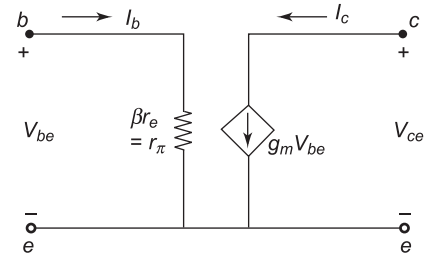


Fig. 5.6 βr_e-g_m model

The *circuit equivalent* of a BJT amplifier of Fig. 5.1(b) is drawn in Fig. 5.7. Voltage and current gain can be easily found by circuit analysis. One observation is made at a glance. The output voltage has a negative sign, which is indication of the fact that output voltage is *phase reversed* from the input voltage (source).

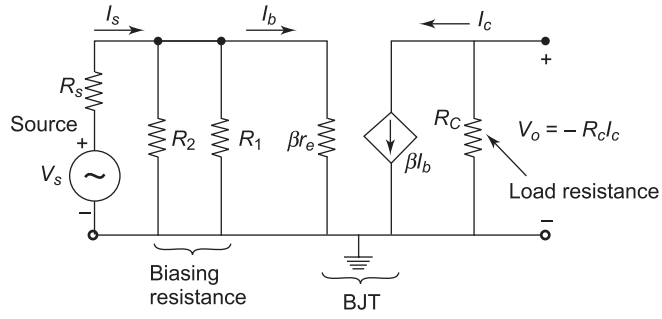


Fig. 5.7

5.3 | HYBRID EQUIVALENT OF BJT—SMALL-SIGNAL MODEL

It has been seen above that BJT is a two-port device. It can therefore be expressed in the form of *h*-parameters defined as per Fig. 5.8(a).

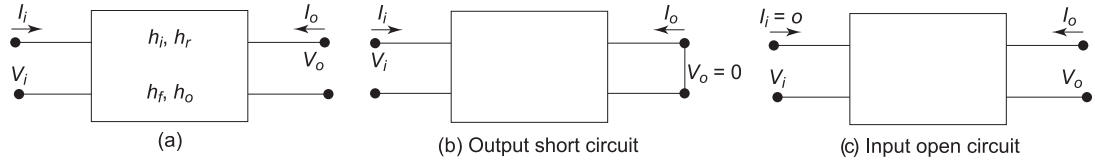


Fig. 5.8 Small-signal model of BJT

$$V_i = h_i I_i + h_r V_o \tag{5.5}$$

$$I_o = h_f I_i + h_o V_o \tag{5.6}$$

Output short circuited ($V_o = 0$), Fig. 5.8(b)

$$h_i = \frac{V_i}{I_i}, \text{ short circuit input impedance } (\Omega)$$

$$h_f = \frac{I_o}{I_i}, \text{ short circuit forward current transfer ratio (dimensionless)}$$

Input open circuited ($I_i = 0$), Fig. 5.8(c)

$$h_r = \frac{V_i}{V_o} = \text{open circuit reverse voltage ratio (dimensionless)}$$

$$h_o = \frac{I_o}{V_o} = \text{open circuit output admittance (S)}$$

The *h*-parameter circuit model is drawn in Fig. 5.9.

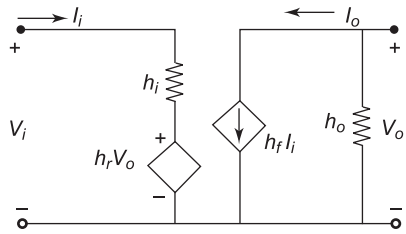


Fig. 5.9 Parameter circuit model

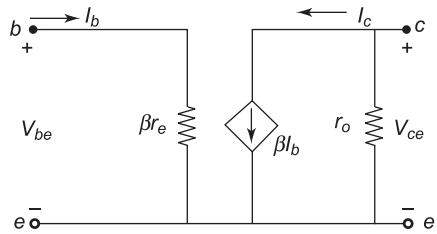


Fig. 5.10 CE small-signal model

The corresponding *CE* small-signal model is drawn in Fig. 5.10. By comparison,

$$h_i = \beta r_e$$
$$h_f = \beta$$
$$h_r = 0$$
$$h_o = r_o$$

In *CE* configuration, we write the *h*-parameters as

h_{ie}, h_{fe}, h_{re}, h_{oe}; suffix *e* (for CE is added)

In *CB* configuration, the *h*-parameters are

h_{ib}, h_{fb}, h_{rb}, h_{ob} ; suffix *b* (for CB is added)

Finally, in terms of *h*-parameters, for small-signal BJT in *CE* configuration, the circuit model is drawn in Fig. 5.11.

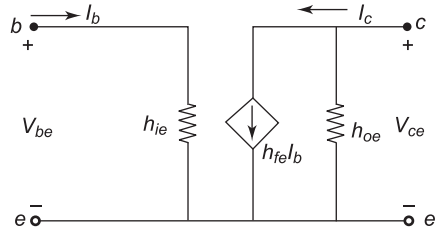


Fig. 5.11 Hybrid model of BJT (small signal)

5.3.1 Conversion Formulas for the Parameters of Three Transistor Configurations

The standard notations used for representing *h*-parameters in CB, CE and CC configurations are shown in Table 5.1 given below. Usually the datasheets of transistor represent the values of *h*-parameters for CE configurations; the conversion formulas are often used to derive the values for other two configurations. The approximate equations to interconvert the *h*-parameters are summarized in the tables given below. Table 5.2 represents the equations for deriving *h*-parameters of CE configuration from CB and CC configurations. Table 5.3 represents the equations for deriving *h*-parameters of CB configuration from CB and CC configurations. Table 5.4 represents the equations for deriving *h*-parameters of CC configuration from CE and CB configurations.

Table 5.1 *Standard notations for representing h-parameters in CB, CE and CC configurations*

S. No.	*h*-parameter	CB configuration	CE configuration	CC configuration
1.	h_{11}	h_{ib}	h_{ie}	h_{ic}
2.	h_{12}	h_{rb}	h_{re}	h_{rc}
3.	h_{21}	h_{fb}	h_{fe}	h_{fc}
4.	h_{22}	h_{ob}	h_{oe}	h_{oc}

Table 5.2 *Conversion formula for h-parameters of CE from CB and CC*

Parameter	CB configuration	CC configuration
h_{ie}	$\frac{h_{ib}}{1+h_{fe}}$	h_{ic}
h_{re}	$\frac{h_{ib}h_{ob}}{1+h_{fb}}-h_{rb}$	$1-h_{rc}$
h_{fe}	$-\frac{h_{fb}}{1+h_{fb}}$	$-(1+h_{fc})$
h_{oe}	$-\frac{h_{ob}}{1+h_{fb}}$	h_{oc}

Table 5.3 *Conversion formula for h-parameters of CB from CE and CC*

Parameter	CE	CC
h_{ib}	$\frac{h_{ie}}{1+h_{fe}}$	$\frac{-h_{ic}}{h_{fc}}$
h_{rb}	$\frac{h_{ie}h_{oe}}{1+h_{fe}}-h_{re}$	$h_{rc}-\frac{h_{ic}h_{oc}}{h_{fe}}-1$
h_{fb}	$-\frac{h_{fe}}{1+h_{fe}}$	$-\frac{1+h_{fc}}{h_{fc}}$
h_{ob}	$-\frac{h_{oe}}{1+h_{fe}}$	$-\frac{h_{oc}}{h_{fc}}$

Table 5.4 *Conversion formula for h-parameters of CC from CE and CB*

Parameter	CE	CB
h_{ic}	h_{ie}	$\frac{h_{ib}}{1+h_{fb}}$
h_{rc}	$1-h_{re}$	1
h_{fc}	$-(1+h_{fe})$	$\frac{-1}{1+h_{fb}}$
h_{oc}	h_{oe}	$\frac{h_{ob}}{1+h_{fb}}$

EXAMPLE 5.1 A BJT in *CB* configuration has I_E(dc) = 3.5 mA, α = 0.988, applied voltage 30 mV, load 2.5 kΩ. Determine r_e, Z_o (input impedance), I_e, V_o, A_v (voltage gain), A_i (current gain), and I_b.

Solution The small-signal circuit is drawn in Fig. 5.12.

$$r_e = \frac{26\text{ mV}}{I_E} = \frac{26}{3.5} = 7.43\ \Omega$$

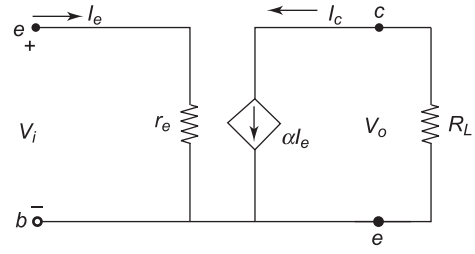


Fig. 5.12 CB small-signal model

Input impedance, $Z_i = r_e = 7.43\ \Omega$

$$I_i = I_e = \frac{V_i}{r_e} = \frac{30}{7.43} = 4\text{ mA}$$

$$V_o = -\alpha I_e R_L$$

$$= -0.988 \times 4 \times 2.5 = 9.88\text{ V}$$

Voltage gain, $A_V = \dfrac{V_o}{V_i} = \dfrac{-9.88}{30 \times 10^{-3}} = -329$

Current gain, $A_i = \dfrac{I_o}{I_i} = \dfrac{I_c}{I_e} = \alpha = 0.988$

$$I_c = 0.988\ I_e$$

$$I_e = I_c + I_b \quad \text{or} \quad I_b = I_e - I_c = (1 - 0.988)\ I_e$$

$$I_b = 0.012 I_e = 0.012 \times 4 = 0.048\text{ mA} \quad \text{or} \quad 48\ \mu\text{A}$$

EXAMPLE 5.2 A common-emitter amplifier has β = 100, I_E(dc) = 2 mA and r_o = 40 kΩ, input V_i = 100 mV. Determine $Z_i : I_b : A_i = I_o/I_i$ and A_v if R_L = 1.25 kΩ.

Solution CE amplifier (Fig. 5.13)

$$r_e = \frac{26\text{ mV}}{2} = 13\ \Omega$$

$$Z_i = \beta r_e = 100 \times 13 = 1.3\ \text{k}\Omega$$

$$V_i = 100\text{ mV}$$

$$I_i = I_b = \frac{100}{1300} = 0.077\text{ mA}$$

$$I_c = 100\ I_b = 7.7\text{ mA}$$

$$R = r_o \parallel R_L = \frac{40 \times 1.25}{40 + 1.25} = 1.21\ \text{k}\Omega$$

$$V_o = -I_c R_L = -7.7 \times 1.21 = -9.32\text{ V}$$

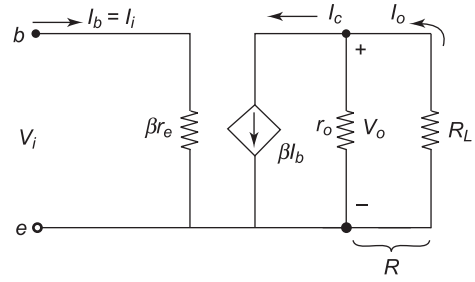


Fig. 5.13

$$I_o = -\frac{9.32}{1.25} = 7.46 \text{ mA}$$

$$A_i = \frac{I_o}{I_i} = \frac{7.46}{0.077} = 96.88 \text{ (slightly less than } \beta)$$

$$A_v = \frac{V_o}{V_i} = \frac{-9.32}{100 \times 10^{-3}} = -93.2$$

EXAMPLE 5.3 The *h*-parameters for CE configuration are h_{ie} = 2600 Ω, h_{fe} = 100, h_{re} = 0.02 × 10^{-2} and h_{oe} = 5 × 10^{-6} S. Find the *h*-parameters for CC configuration.

Solution

$$h_{ic} = h_{ie} = 2600\ \Omega$$

$$h_{fc} = -(1 + h_{fe}) = -101$$

$$h_{rc} = 1 - h_{re} = 1$$

$$h_{oc} = h_{oe} = 5 \times 10^{-6}\ \text{S}$$

5.3.2 Thevenin's and Norton's Theorems and Corollaries

Thevenin's Theorem states that any two terminal linear networks may be replaced by a voltage source equal to the open circuit voltage between the terminals in series with the output impedance seen at the port. The input circuit of a transistor and its equivalent circuit using Thevenin's theorem is shown in Fig. 5.14(a). In this figure, *V* is the open circuit voltage and *Z* is the impedance between the terminals. The value of *Z* may be calculated as

- Short-circuit all the independent voltage sources
- Open-circuit all the independent current sources

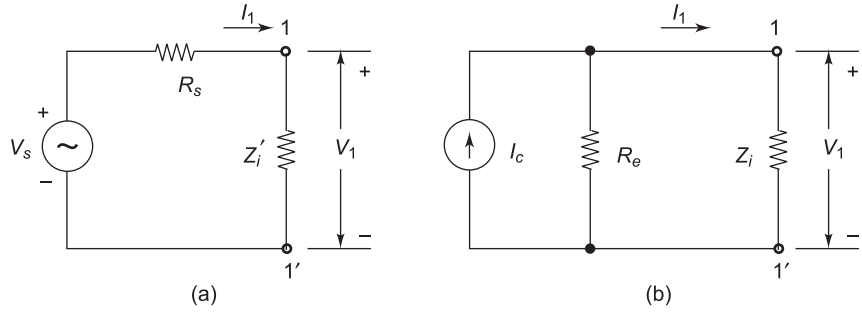


Fig. 5.14 Input circuit of a transistor amplifier using (a) Thevenin's Theorem, and (b) Norton's Theorem

Norton's Theorem states that any two-terminal linear network may be replaced by a current source equal to the short-circuited current between the terminals in parallel with the output

impedance seen at this port. The equivalent circuit of transistor using Norton's theorem is shown in Fig. 5.14(b). Here, I is the short circuit current and Z is the impedance between the terminals. Voltage source V in series with an impedance Z is equivalent to a current source $I = V/Z$ in parallel with the impedance Z.

✦ Corollaries

If V is the open circuit voltage, I the short circuit current, Z the impedance, and Y the admittance between two terminals in Fig. 5.15 then,

$$Z = V/I \text{ (this expression is used to calculate the value of impedance)}$$

$$I = V/Z = VY$$

$$V = IZ = I/Y \text{ (used to calculate voltage between two ports)}$$

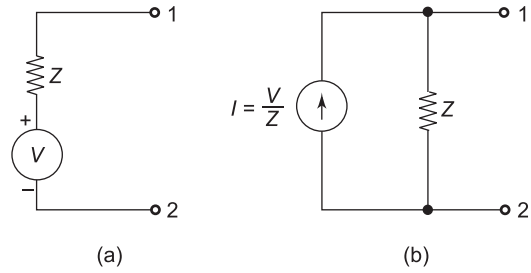


Fig. 5.15 (a) Thevenin's equivalent circuit, and (b) Norton's equivalent circuit

Thevenin's theorem, Norton's theorem and the corollaries are used to simplify the network for analysis.

5.4 SMALL-SIGNAL MODEL OF FETS AND AMPLIFIERS

The drain current of FET depends mainly on gate-to-source voltage related by the transfer characteristic JFET and DMOSFET. The relationship is the Shockley's equation

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{5.7}$$

At the Q-point, the linearized (small-signal) relationship is given by

$$g_m = \frac{dI_D}{dV_{GS}} = \textit{transconductance (S)} \tag{5.8}$$

In small-signal symbols,

$$I_d = g_m V_{gs} \tag{5.9}$$

By carrying out derivation of Eq. (5.7), it can be shown that

$$g_m = \frac{2I_{DSS}}{|V_P|}\left[1 - \frac{V_{GS}}{V_P}\right] \tag{5.10}$$

$|V_p|$ assures that g_m is a positive value.

At $V_{GS} = 0$, $$g_m = \frac{2I_{DSS}}{|V_P|} = g_{m0} \tag{5.11}$$

We can then write

$$g_m = g_{m0}\left[1 - \frac{V_{GS}}{V_P}\right] = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}} \tag{5.12}$$

❑ ***Output Resistance*** It is contributed by the slope of the drain characteristics at Q-point,

$$r_d = \left.\frac{dV_{DS}}{dI_D}\right|_{\text{at } V_{GS}} \tag{5.13}$$

or $$I_d = \frac{1}{r_d}V_{ds} \tag{5.14}$$

❑ ***Circuit Model*** Combining Eqs (5.9) and (5.14), we can write

$$I_d = g_m V_{gs} + \frac{1}{r_d}V_{ds} \tag{5.15}$$

From Eq. (5.18) and noting that $I_G = 0$, open circuit, circuit model is drawn in Fig. 5.16.

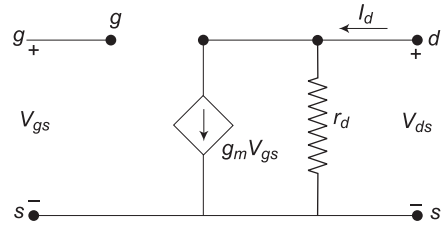


Fig. 5.16 Circuit model of FET-common source

✦ EMOSFET

Its transfer characteristics is

$$I_D = k\,(V_{GS} - V_T)^2 \tag{5.16}$$

Taking derivative

$$g_m = 2k\,(V_{GS} - V_T) \tag{5.17}$$

or $$g_m = 2\sqrt{kI_D} \tag{5.18}$$

EXAMPLE 5.4 For a JFET, calculate g_{m0}, other parameters are I_{DSS} = 15 mA ,V_p = −5 V

Solution

$$g_{m0} = \frac{2I_{DSS}}{|V_P|} = \frac{2 \times 15}{5} = 6 \text{ mS}$$

EXAMPLE 5.5 The parameters of a JFET are g_{m0} = 5 mS and V_p = −4 V. What is the device current at V_{GS} = 0?

Solution

$$g_{m0} = \frac{2I_{DSS}}{|V_P|}$$

$$5 = \frac{2I_{DSS}}{4}$$

$$I_{DSS} = 10 \text{ mA} = I_D(\text{at } V_{GS} = 0)$$

EXAMPLE 5.6 A JFET has parameters I_{DSS} = 12 mA, V_p = −6 V. It is biased at $I_D = I_{DSS}/4$. What is its g_m at the Q-point?

Solution

$$I_D = 12/4 = 3 \text{ mA}$$

$$g_{m0} = \frac{2I_{DSS}}{|V_P|} = \frac{2 \times 12}{6} = 4 \text{ mS}$$

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}} = 4\sqrt{\frac{3}{12}} = 2 \text{ mS}$$

EXAMPLE 5.7 A JFET has r_d = 100 kΩ and voltage gain of A_v = −200. What is the value of g_m?

Solution Refer common-source small-signal model of Fig. 5.16.

$$V_{ds} = -g_m V_{gs} r_d$$

Ideal gain, $$A_v = \frac{V_{ds}}{V_{gs}} = -g_m r_d = -200$$

$$g_m = \frac{200}{r_d} = \frac{200}{100} = 2 \text{ mS}$$

EXAMPLE 5.8 For the transfer characteristics of Fig. 5.17, determine:

(a) g_{m0} (b) g_m at $V_{GS} = -1.5$ V (c) g_m at $V_{GS} = -2.5$ V

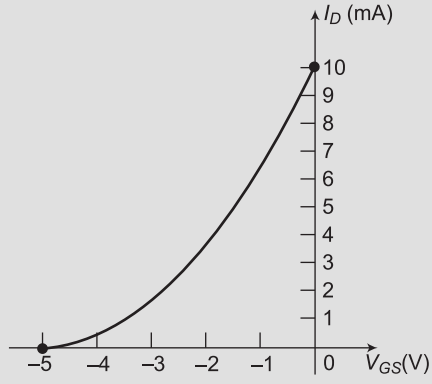


Fig. 5.17

Solution

(a) $I_{DSS} = 10 \text{ mA}. \; V_p = -5$

$$g_{m0} = \frac{2I_{DSS}}{|V_P|} = \frac{10}{5} = 2 \text{ mS}$$

(b) $V_{GS} = -1.5 \text{ V}$

$$g_m = g_{m0}\left(1 - \frac{V_{GS}}{V_p}\right)$$

$$= 2\left(1 - \frac{1.5}{5}\right) = 1.4 \text{ mS}$$

(c) $V_{GS} = -2.5$

$$g_m = 2\left(1 - \frac{2.5}{5}\right) = 1 \text{ mS}$$

Remark: g_m reduces with increases of V_{GS}.

EXAMPLE 5.9 For the JFET amplifier of Fig. 5.18(a), assume that the biasing circuit causes $V_{GSQ} = -3$ V. Determine: Z_i (input impedance) Z_0 (output impedance) and A_v (voltage gain).

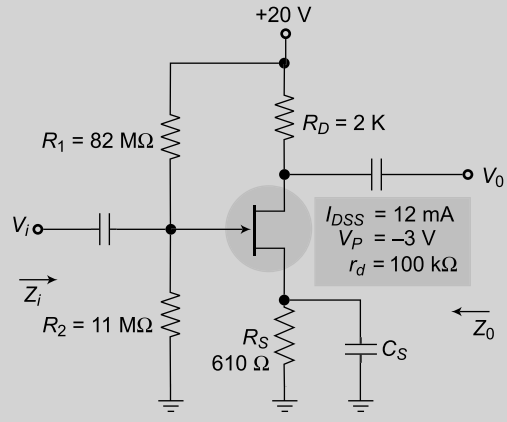


Fig. 5.18(a)

Solution

Steps

- Replace all capacitors by short circuit. R_S gets eliminated.
- Make +20 V to zero and connect to ground.
- Replace JFET by its small-signal model.

The small-signal circuit in drawn in Fig. 5.18(b).

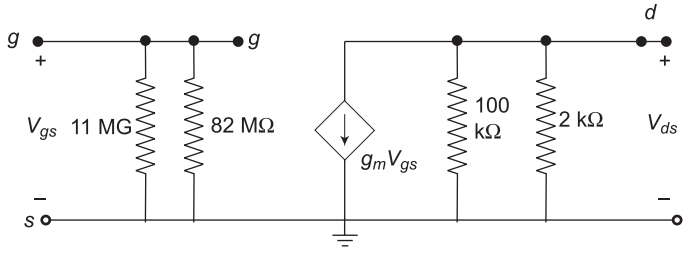


Fig. 5.18(b)

$$g_m = \frac{2I_{DSS}}{|V_p|}\left(1 - \frac{V_{GS}}{V_p}\right) = \frac{2 \times 12}{3}\left(1 - \frac{1.5}{3}\right) = 4 \text{ mS}$$

$$Z_i = \frac{82 \times 11}{82 + 11} = 9.7 \text{ MΩ}$$

Open circuit input dependent current source open.

$$Z_0 = \frac{2 \times 100}{2 + 100} = 1.96 \text{ kΩ}, \; r_d \text{ has very little effect}$$

$$V_0 = V_{ds} = -1.96 \, g_m V_{gs}$$

$$V_i = V_{gs}$$

$$A_v = \frac{V_0}{V_i} = -1.96 g_m = -1.96 \times 4 = -7.81$$

EXAMPLE 5.10

DMOSFET amplifier network is drawn in Fig. 5.19(a). The biasing (voltage divider) has been designed to place the Q-point at $V_{GSQ} = -1.2$ V and $I_{DQ} = 2.07$ mA.

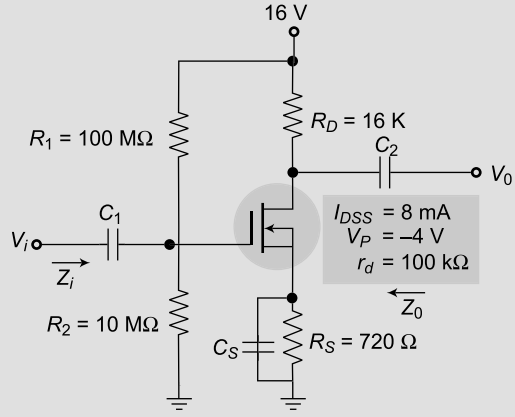


Fig. 5.19(a)

(a) Determine g_{m0} and g_m. (b) Sketch the small-signal equivalent circuit. (c) Calculate Z_i and Z_0. (d) Calculate A_v.

Solution

(a)
$$g_{m0} = \frac{2I_{DSS}}{|V_p|} = \frac{2 \times 8}{4} = 4 \text{ mS}$$

$$g_m = 4\left(1 - \frac{1.2}{4}\right) = 5.71 \text{ mS}$$

(b) Small-signal equivalent is drawn is Fig. 5.19(b).

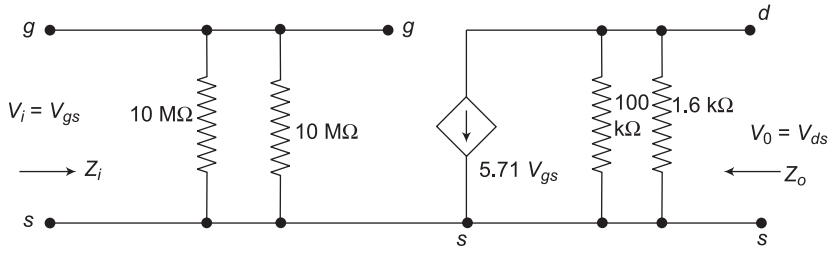


Fig. 5.19(b)

(c) $Z_i = 10 \| 100 = 9.1$ mΩ

(d) $Z_0 = 1.6 \| 100$ kΩ $= 1.57$ kΩ

(e) $A_v = -g_m (r_d \| R_D) = -5.71 \times 1.57 = -8.96$

5.4.1 Source Follower (Common Drain) Configuration

This corresponds to BJT's Emitter follower. The circuit is drawn in Fig. 5.20. Its small-signal circuit equivalent is drawn in Fig. 5.21(a). It is redrawn in Fig. 5.21(b) in modified form.

$$V_{gs} = V_i - V_0$$
$$V_0 = g_m V_{gs}(r_d \| R_s)$$
$$= g_m(r_d \| R_s)\,(V_i - V_0)$$
$$[1 + g_m(r_d \| R_s)]V_0 = g_m(r_d \| R_s)V_{gs}$$
$$A_v = \frac{V_0}{V_i} = \frac{g_m(r_d \| R_s)}{1 + g_m(r_d \| R_s)} \quad (5.19)$$

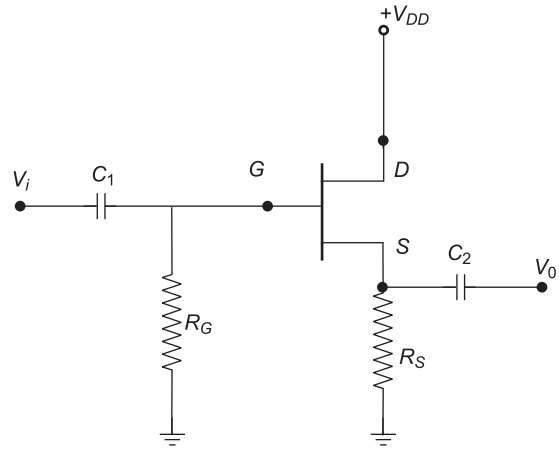


Fig. 5.20 Emitter follower circuit

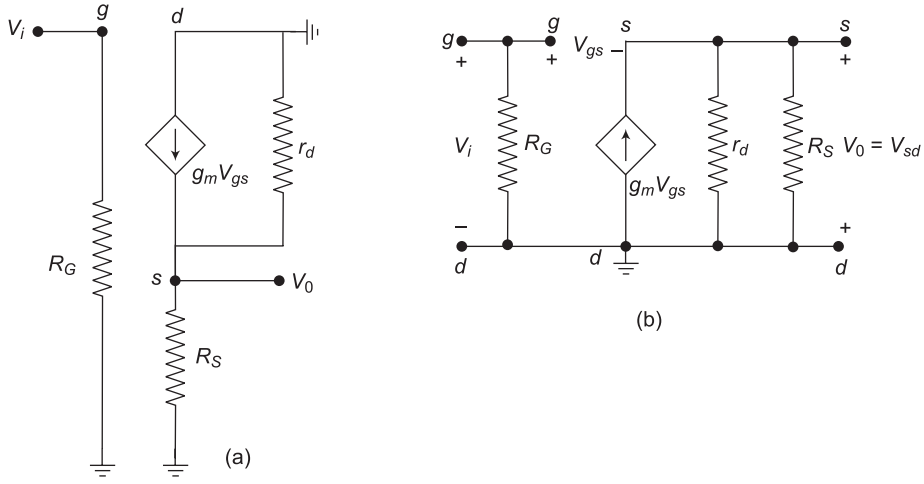


Fig. 5.21 (a) Small-signal circuit equivalent (b) modified form

✦ Observation

$A_v < 1$, A_v positive, no phase reversal as in common-source $Z_i = R_G$.

To find Z_0,

Open-circuit voltage = V_0

Short-circuit current (s-d shorted),

$$I_{sc} = g_m V_{gs}$$

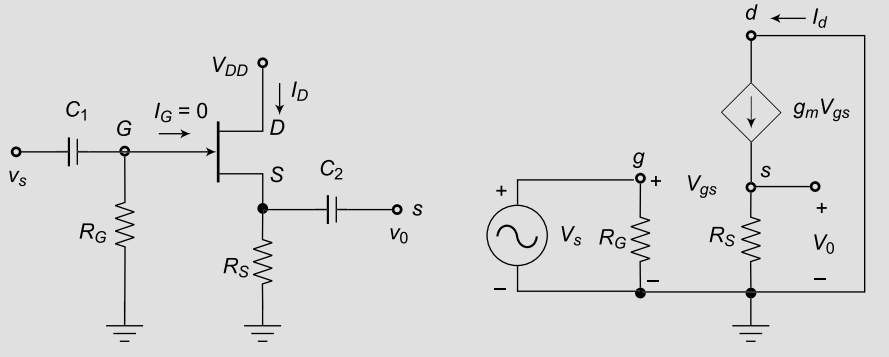
$$V_0 = \frac{g_m(r_d \parallel R_s)V_{gs}}{1 + g_m(r_d \parallel R_s)}$$

$$Z_0 = \frac{V_0}{I_{sc}} = \frac{g_m(r_d \parallel R_s)}{g_m[1 + g_m(r_d \parallel R_s)]} = \frac{1}{r_d \parallel R_s + g_m}$$

or
$$Z_0 = \frac{1}{\frac{1}{r_d} + \frac{1}{R_s} + g_m} \tag{5.20}$$

EXAMPLE 5.11 For the source-follower of Fig. 5.22, I_{DSS} = 16 mA, V_p = –4 V, R_G = 1 MΩ, R_s = 2.2 kΩ, V_{DD} = +9 V

Calculate the voltage gain, and output impedances.



(a) Source-follower (b) Small-signal model of source-follower

Fig. 5.22

Solution DC bias conditions

Assume $I_G = 0$

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{i}$$

$$V_{GS} = -R_S I_D \tag{ii}$$

Substituting values,

$$I_D = 16\left(1 + \frac{V_{GS}}{4}\right)^2 \tag{iii}$$

$$V_{GS} = -2.2\, I_D \tag{iv}$$

Substituting I_D from Eq. (iv) in Eq. (iii),

$$I_D = 16\left(1 - \frac{2.2 I_D}{4}\right)^2$$

or $I_D^2 - 3.84\, I_D + 3.31 = 0$

or $I_D = 1.3$ mA or 2.54mA

The second value is rejected as it gives V_{GS} more negative than V_P (pinch-off voltage). Thus,

$$I_D = 1.3 \text{ mA}$$

$$V_{GS} = -2.2 \times 1.305 = -2.86 \text{ V}$$

$$g_{m0} = -\frac{2I_{DSS}}{V_P} = 2 \times \frac{16}{4} = 8 \text{ mS}$$

$$g_m = g_{m0}\left(1 - \frac{V_{GS}}{V_P}\right) = 8\left(1 - \frac{286}{4}\right) = 2.28 \text{ mS}$$

$$r_m = \frac{1}{g_m} = 0.439 \text{ k}\Omega$$

$$A_V = \frac{2.2 \times 2.28}{1 + 2.2 \times 2.28} = 0.834 \ (< 1)$$

$$Z_i = R_G = 1 \text{ M}\Omega$$

$$Z_0 = R_S \,||\, r_m = 2.2 \,||\, 0.439 = 365\ \Omega \text{ [very low]}$$

Note: r_d is ignored.

✦ EMOSFET

The relationship between drain current and controlling voltage is

$$I_D = k(V_{GS} - V_T)^2 \tag{5.21}$$

Taking derivation wrt V_{GS},

$$g_m = 2k(V_{GS} - V_T) \tag{5.22}$$

g_m has to be evaluated at V_{GSQ} (Q-point)

EXAMPLE 5.12 Determine g_m if $V_T = 4$ V and $V_{GSQ} = 8$ V. Take $k = 3.5/\text{V}^2$

Solution

$$g_m = 2 \times 3.5(8 - 4) = 20 \text{ mS}$$

EXAMPLE 5.13 For the amplifier of Fig. 5.23, EMOSFET data is $V_T = 3$ V, $k = 0.4$ mA/V^2, $r_d = 40$ kΩ. Find the voltage gain A_v.

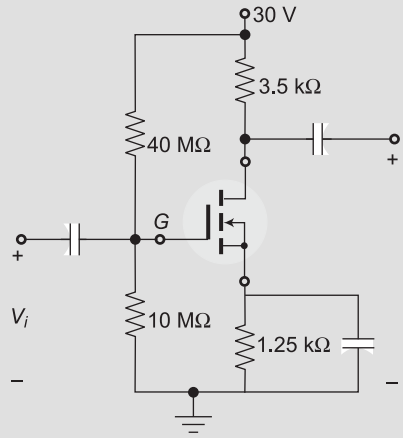


Fig. 5.23

Solution For ac signal analysis, we need V_{GSQ} for determining g_m. We carry out dc analysis by short circuiting all capacitors. Note that $I_G = 0$

$$V_G = 30 \times \frac{10}{10 + 40} = 6 \text{ V}$$

$$I_D = k(V_{GS} - V_T)^2 = 0.4(V_{GS} - 3)^2 \quad \text{(i)}$$

$$V_{GS} = V_G - I_S R_S = 6 - 1.25\, I_D \quad \textbf{(ii)}$$

$$V_{GS} = 0 \rightarrow I_D = \frac{6}{1.25} = 4.8 \text{ mA}$$

$$I_D = 0 \rightarrow V_{GS} = 6 \text{ V}$$

Equations (i) and (ii) are plotted in Fig. 5.24. Their intersection gives $V_{GSQ} = 4.7$ mV
We now find

$$g_m = 2k(V_{GSQ} - V_T)$$
$$= 2 \times 0.4\,(4.7 - 3) = 1.36 \text{ mS}$$

The ac signal circuit of the amplifier is drown in Fig. 5.25.

$$V_0 = -\, g_m V_{gs}\,(40 \| 3.5)$$

$$V_{gs} = V_i$$

$$V_0 = -\, 1.36\,(40 \| 3.5) V_i$$

$$A_v = \frac{V_0}{V_i} = -\, 4.38$$

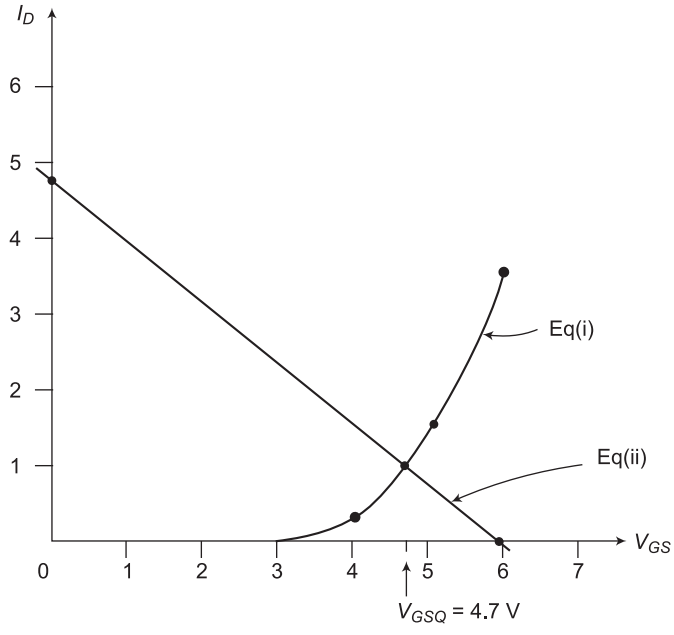


Fig. 5.24

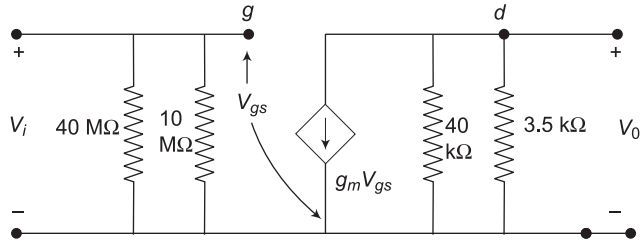


Fig. 5.25

✦ Effect of Source Resistance and Load Resistance

Treatment is similar to that of BJT based amplifier

Gain: Without load > with load

5.5 | CE BJT AMPLIFIER

A small-signal BJT amplifier with voltage divider bias is drawn in Fig. 5.26. In the range of frequencies called *midband*, the coupling capacitors act as short circuits ($1/C\omega \approx 0$). Thereby, the input and output signals are connected directly to the BJT. Also, the *bypass* capacitor acts

as a short across R_E. This is needed, otherwise R_E would provide negative feedback reducing the small-signal gain of the amplifier.

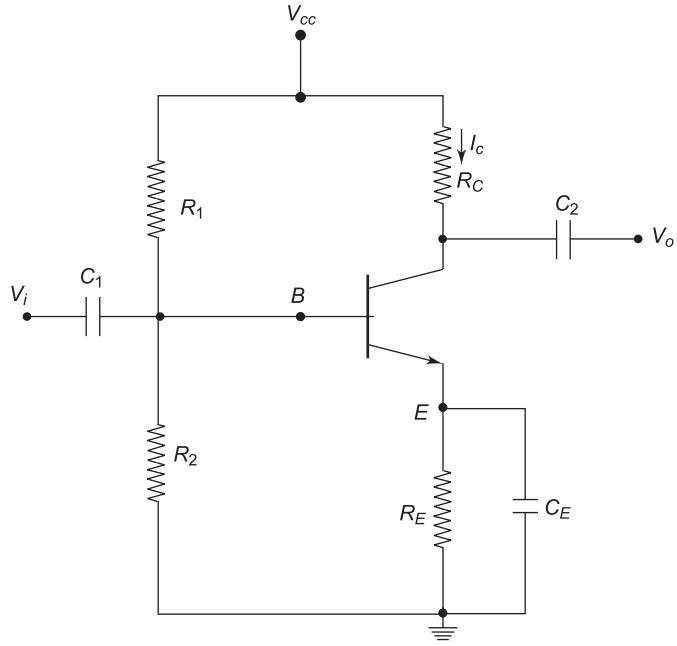


Fig. 5.26 CE amplifier

The small-signal circuit equivalent of the amplifier is drawn in Fig. 5.27.

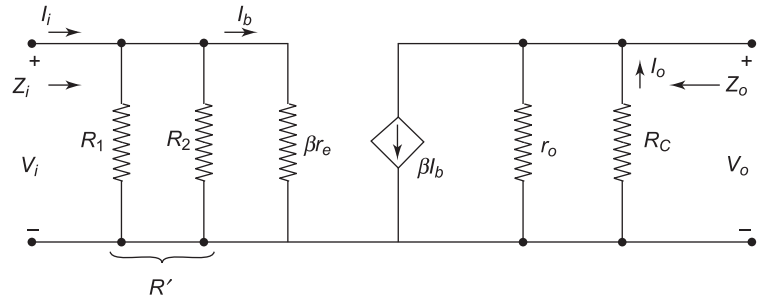


Fig. 5.27 Small-signal circuit equivalent

✦ Circuit Analysis

Input impedance

$$Z_i = R_1 \| R_2 \| \beta r_e \tag{5.23a}$$

Output impedance

$$Z_o = r_o \| R_C \approx R_C \text{ as } r_o \geq 10\, R_c \tag{5.23b}$$

✦ Voltage Gain

$$I_i = \frac{V_i}{R \parallel \beta r_e} \tag{5.23c}$$

Output impedance – Short-circuit input

The $Z_o = r_o \parallel R_c$

$$V_o = -\beta I_b Z_o = -\beta\left(\frac{V_i}{\beta r_e}\right) Z_o = -\left(\frac{V_i}{r_e}\right) Z_o$$

$$A_v = \frac{V_o}{V_i} = -\frac{Z_o}{r_e} = -\frac{r_o \parallel R_c}{r_e} \tag{5.24a}$$

or $$A_v = -\frac{R_c}{r_e} \text{ for } r_o \geq 10\, R_c \tag{5.24b}$$

✦ Current Gain

$$I_o = \beta I_b\left(\frac{r_o}{r_o + R_c}\right) \Rightarrow \frac{I_o}{I_b} = \frac{\beta r_o}{(r_o + R_c)}$$

$$I_b = \frac{V_i}{\beta r_e}$$

$$I_i = I_b + \frac{V_i}{R'} = I_b + \left(\frac{\beta r_e}{R'}\right) I_b = \left(\frac{R' + \beta r_e}{R'}\right) I_b$$

or $$\frac{I_b}{I_i} = \left(\frac{R'}{R' + \beta r_e}\right)$$

$$A_i = \frac{I_o}{I_i} = \frac{I_o}{I_b} \cdot \frac{I_b}{I_i} = \frac{R' \beta r_e}{(r_o + R_c)(R' + \beta r_e)} \tag{5.25}$$

❑ *To Sum Up*

- From the *RC*-coupled amplifier circuit, draw the small-signal circuit.
- From the biasing analysis, find I_E (dc).
- Compute $r_e = \dfrac{26 \text{ mA}}{I_E}$
- The transition β is known from manufacturing data.
- Z_i, Z_o, A_v and A_i are now computed for the small-signal circuit.

The expressions for these have been derived earlier.

EXAMPLE 5.14 For the BJT amplifier of Fig. 5.28, draw the small-signal models and determine there from Z_i, Z_o, A_v, A_i.

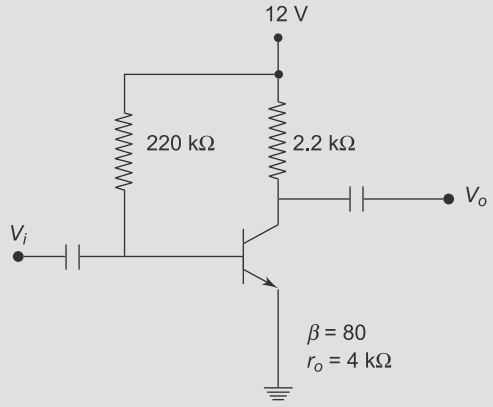


Fig. 5.28

Solution The dc analysis (capacitors short circuit)

$$I_B = \frac{12 - 0.7}{220} = 0.051 \text{ mA}$$

$$I_E \approx I_C = 80 \times 0.051 = 4.08 \text{ mA}$$

$$r_e = \frac{26}{4.08} = 6.37\ \Omega$$

$$\beta r_e = 80 \times 6.37$$

$$= 509.6\ \Omega = 0.51\ \text{k}\Omega$$

Small-signal model (Fig. 5.29)

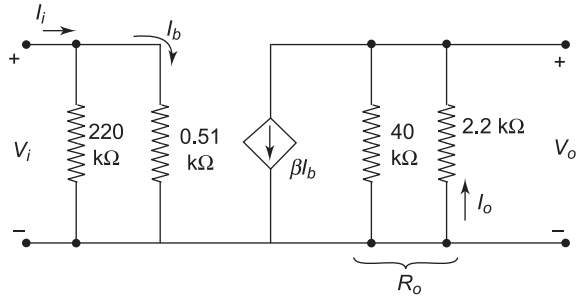


Fig. 5.29

$$R_o = \frac{2.2 \times 40}{2.2 + 40} = 2.085\ \text{k}\Omega$$

Z_i = 0.51 kΩ; 220 kΩ, can be ignored

Short input, βI_b open circuit

$$Z_o = R_o = 2.085 \text{ k}\Omega$$

Let $V_i = 1\text{V}$ (for convenience)

$$I_b = \frac{1}{0.51}, \beta I_b = \frac{80}{0.51} = 157 \text{ mA}$$

$$V_a = -\frac{80}{0.51} \times 2.085 = -327 \text{ V}$$

$$A_V = -327$$

Output current

$$I_o = 157 \times \frac{40}{2.2 + 40} = 148.8 \text{ mA}$$

Current gain

$$A_i = \frac{I_o}{I_b} = \frac{148.8}{1/0.51} = 75.9 \quad I_b \approx I_i$$

EXAMPLE 5.15

Draw the small-signal model of the amplifier circuit of Fig. 5.30. Determine Z_i, Z_o, A_v, A_i. The dc analysis yields I_E = 1.38 mA. All resistances are in kΩ.

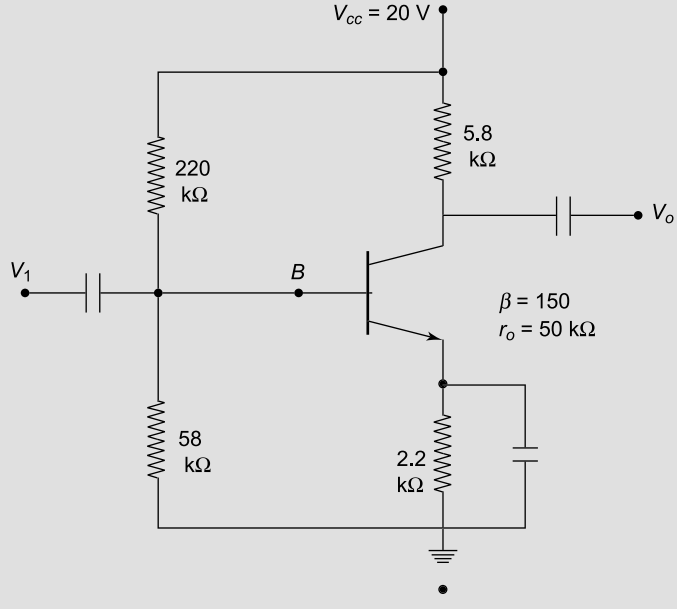


Fig. 5.30

Solution

$$r_e = \frac{26}{1.38} = 18.84\ \Omega$$

$$\beta r_e = 2.826\ \text{k}\Omega$$

The small-signal model is drawn Fig. 5.31.

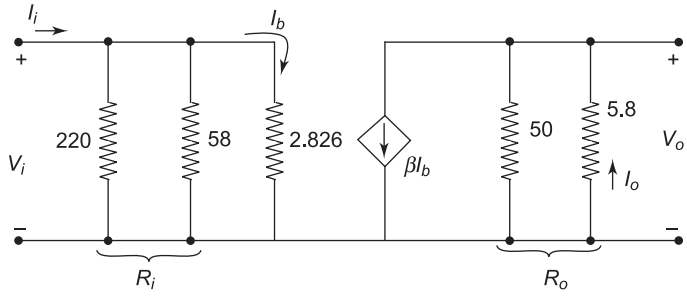


Fig. 5.31

Let $V_i = 1$ V (for convenience)

$$I_b = \frac{1}{2.826} = 0.354\ \text{mA}$$

$$\beta I_b = 150 \times 0.354 = 53.1\ \text{mA}$$

$$R_o = 50 \parallel 5.8 = 5.2\ \text{k}\Omega$$

$$V_o = -53.1 \times 5.2 = -276\ \text{V}$$

$$A_V = -276$$

$$R_i = 220 \parallel 58 \parallel 2.826 = 45.9 \parallel 2.826 = 2.66\ \text{k}\Omega$$

$$I_i = \frac{1}{2.66} = 0.376\ \text{mA}$$

$$I_o = 53.1 \times \frac{50}{50 + 5.8} = 47.6\ \text{mA}$$

$$A_i = \frac{47.6}{0.376} = 126.6$$

$$Z_i = R_i = 2.66\ \text{k}\Omega$$

$$Z_o = R_o = 5.2\ \text{k}\Omega$$

5.6 | EFFECT OF SOURCE AND LOAD RESISTANCE

A real voltage source which feeds the amplifier has a series resistance as shown Fig. 5.32(a). The series voltage drop reduces V_i fed to the amplifier.

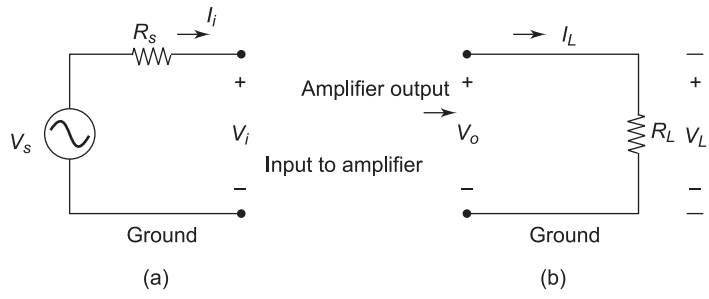


Fig. 5.32 Voltage source

The voltage amplifier as the output feeds a load, which is a shunt connector, as shown in Fig. 5.32(b). R_L is in shunt with $r_o \| R_c \approx R_c$. The output voltage

$$V_L = \beta I_b (R_c \| R_L) \text{ which is less than } V_0 = \beta I_b R_c \text{ as } R_c \| R_L < R_c$$

The overall affect of source and load resistance is to reduce the net voltage gain of the amplifier.

The ac circuit of a voltage-divider biased CE amplifier fed from a source and feeding a load is drawn in Fig. 5.33.

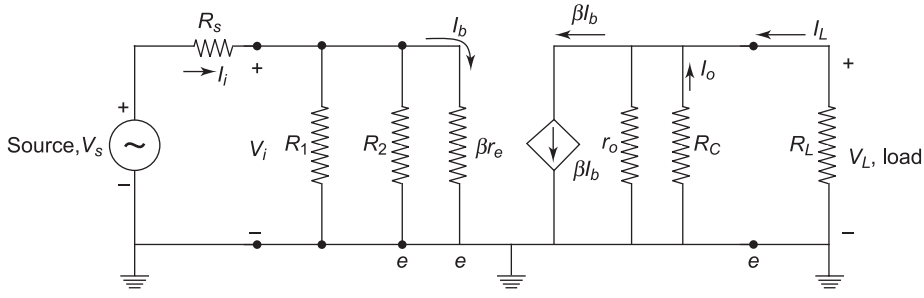


Fig. 5.33 ac circuit of voltage-divider biased CE amplifier

It is easy to see that $V_i < V_s$ and $I_L < \beta I_b$. The overall amplifier gain is

$$A_v(\text{overall}) = \frac{V_L}{V_s} < \frac{V_o}{V_i}, \; V_o \text{ is found by disconnecting } R_L$$

5.7 | FREQUENCY RESPONSE

So far, we have studied the BJT *RC*-coupled small signal amplifiers in the mid-band range of frequencies. In the mid-band, the capacitors are effective short circuits and the amplifier's mid-band gain is constant. Also, in the high-frequency end of mid-band, the transistor capacitances are in shunt but act as open circuits.

Let us now consider the frequency response of amplifiers at low frequencies end of mid-band and also at high-frequency end of mid-band.

The effect of coupling capacitors is *RC* series wherein the voltage drop in capacitive reactance ($1/2\pi fC$) increases as the signal frequency reduces; resulting in gain reduction. The bypass capacitor across (R_E or R_S) is a shunt effect (C in parallel with R). As the capacitive reactance increases with lower signal frequency, the effectiveness of bypassing reduces causing increase in negative voltage feedback and so gain reduction.

At high–frequency, the device capacitances—inter-terminal, stray (wiring, etc.) and load capacitance—are effectively in shunt (C in parallel with R). At high frequencies, ($1/2\pi fC$) decreases and so, the shunt effect reduces the gain. A typical frequency response is plotted to scale in log frequency in Fig. 5.34.

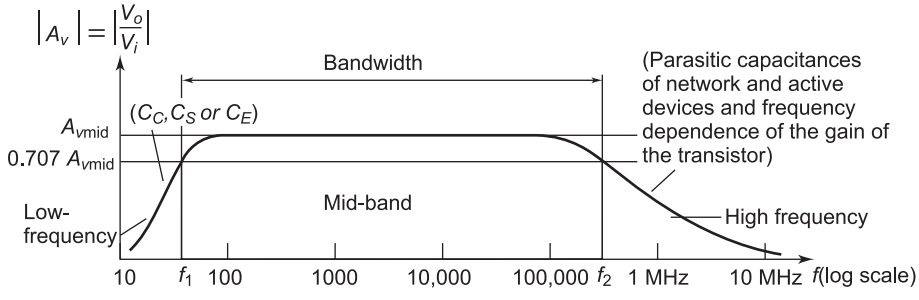


Fig. 5.34 Typical frequency response of RC-coupled amplifier

As f_1 is very small, BW (bandwidth) $\approx f_2$ (5.26a)

✦ Decibel (dB)

It is 1/10th of bel. It is defined as gain

$$G_{\text{dB}} = 10 \log \frac{P_2}{P_1} \tag{5.26b}$$

where P_2 and P_1 are two equipment powers being compared.

It is advantageous to take P_1 as reference power. It is taken as

1 mW, for 600 Ω

Thus, $$G_{\text{dB}m} = \left.\frac{P}{1\,\text{mW}}\right|_{\text{at } 600\Omega} \tag{5.27}$$

For ratio of voltage (voltage gain, amplification), for a given R

$$P_2 = V_2^2/R,\ P_1 = V_1^2/R$$

Then $$G_{\text{dB}} = 10 \log \left(\frac{V_2}{V_1}\right)^2 = 20 \log \frac{V_2}{V_1} \tag{5.28}$$

At lower 3-dB frequency, the gain reduces by $(1/\sqrt{2})$. In dB scale

$$\text{dB}\left(\frac{1}{\sqrt{2}}\right) = 20\log\frac{1}{\sqrt{2}} = -3\text{ dB} \tag{5.29}$$

Thus, at lower 3-dB frequency, gain drops by 3 dB.

5.8 | LOW-FREQUENCY RESPONSE

The frequency response of C_s (input side) and C_c (output side) can be simulated by the simple circuit of Fig. 5.35(a). The circuit is redrawn in frequency domain in Fig. 5.35(b). By voltage division,

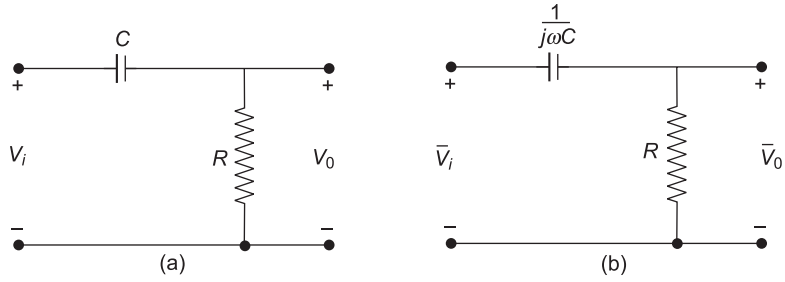


Fig. 5.35 (a) RC circuit (b) RC circuit in frequency domain

$$\bar{V}_0 = \left(\frac{R}{R+\frac{1}{j\omega c}}\right)\bar{V}$$

The circuit gain (complex) is

$$\bar{A}_v = \frac{\bar{V}_0}{\bar{V}_i} = \frac{1}{1-j(1/\omega CR)}$$

or

$$\bar{A}_v = \frac{1}{(1-j(\omega_1/\omega))} \tag{5.30}$$

where $\omega_1 = \dfrac{1}{CR}$ or $f_1 = \dfrac{1}{2\pi CR}$; *break frequency* (5.31)

We can write

$$\bar{A}_V = \frac{1}{1-j(f_1/f)} \tag{5.32}$$

$$dB\ \left|\bar{A}_v\right| = 20\log\frac{1}{\sqrt{1+(f_1/f)^2}}$$

For $f >> f_1$, dB $|\bar{A}_v| = 20 \log 1 = 0$ dB, asymptote

For $f << f_1$, dB $|\bar{A}_v| = -20 \log(f_1/f)$, asymptote slope + 20 dB/decade as f decreases

At $f = f_1$, dB $|\bar{A}_v| = 20 \log \frac{1}{\sqrt{2}} = -3$ dB

The asymptote and corrected dB log f (*Bode plot*) is drawn in Fig. 5.36.

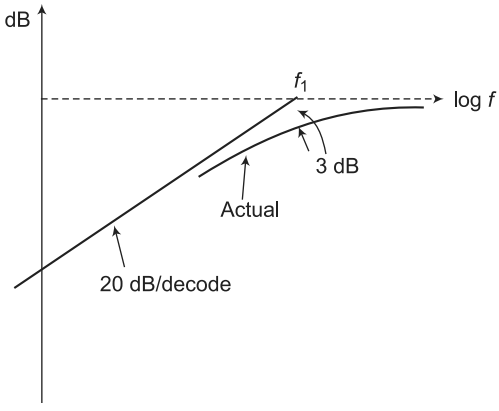


Fig. 5.36 Bode plot

✦ Low Frequency Response BJT Amplifier

From small-signal equivalents; input, output capacitances considered.

❑ ***Input Side (C_s)*** The input side small-signal circuit retaining C_s is drawn in Fig. 5.37. If follows that

$$\bar{V}_a = \left[\frac{R_i}{(R_s + R_i) - jX_c}\right] V_s \tag{5.33a}$$

Break frequency from Eq. 5.33,

$$f_{Li} = \frac{1}{2\pi(R_s + R_i)C} \tag{5.33b}$$

❑ ***Output Side (C_c)*** The output side small-signal circuit retaining C_c is drawn in Fig. 5.38. The break frequency

$$f_{L0} = \frac{1}{2\pi(R_0 + R_L)C_c} \text{ as } r_0 \gg R_c \tag{5.34}$$

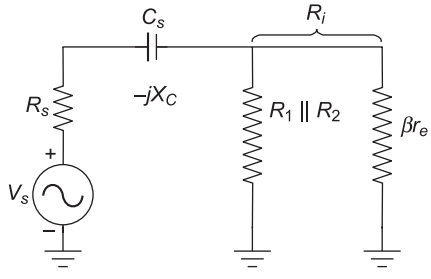


Fig. 5.37 Input side

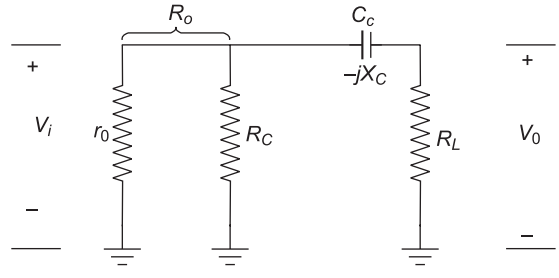


Fig. 5.38 Output side

❑ ***Emitter Circuit (R_E || C_E)*** If can the shown by certain assumptions (see Fig. 5.39) that

$$f_{LE} = \frac{1}{2\pi R_e C_E} \qquad (5.35)$$

where $R_e = R_E \| \left(\frac{R'_s}{\beta} + r_e \right)$

$R'_s = R_s \| R_1 \| R_2$

If low frequency and there are three break frequencies (not coincident). The slope of Bode plot beyond lowest break frequency is then –3 × 20 = –60 dB/decade. Highest of the three is the break frequency f_1.

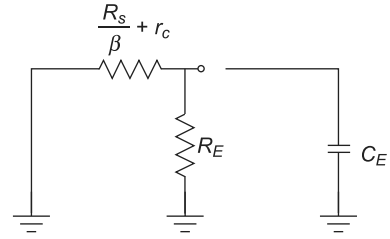


Fig. 5.39 Emitter circuit

✦ Low Frequency Response JFET Amplifier

If can be similarly shown that [see Figs 5.40 (a), (b) and (c)]

$$f_{Li} = \frac{1}{2\pi(R_s + R_i)C_G} \quad (C_G = \text{Grid side capacitance}) \qquad (5.36)$$

$$f_{Lo} = \frac{1}{2\pi(R_o + R_L)C_c} \qquad (5.37)$$

$$R_0 = R_D \| Z_d$$

$$f_{Ls} = \frac{1}{2\pi R_{\text{eq}} C_s} \qquad (5.38)$$

$$R_{\text{eq}} = R_s \| \frac{1}{g_m}; \; r_d \text{ ignored}$$

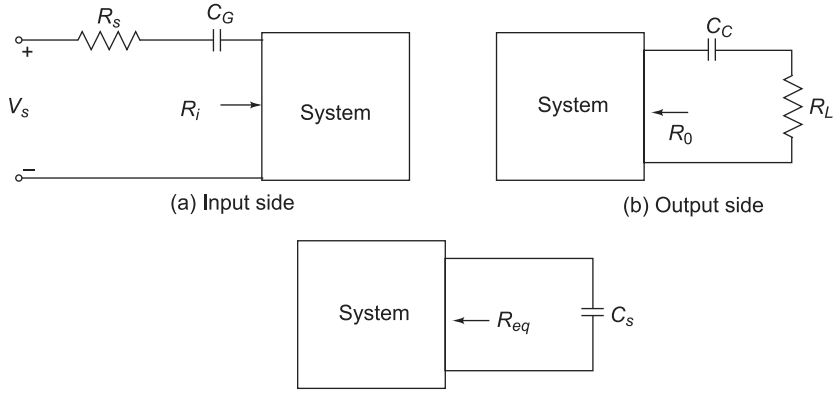


Fig. 5.40

5.9 HIGH-FREQUENCY RESPONSE

The effect of device capacitances, stray capacitance and load capacitance can be lumped as shunt capacitance C_i in input side and C_0 on output side. Of course, coupling and bypass capacitors are more effective short circuit. Thus, the input and output circuits have the parameters as of Fig. 5.41 from which it follows that

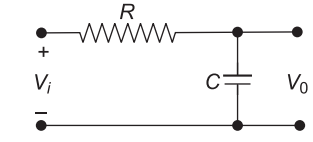


Fig. 5.41 High-frequency response

$$\bar{V}_0 = \left(\frac{-jX_C}{R - jX_C}\right)\bar{V}_i \qquad (5.39a)$$

$$\frac{-jX_C}{R - jX_C} = \frac{1}{1 + j\frac{R}{X_c}} = \frac{1}{1 + j\omega R} = \frac{1}{1 + j(f/f_2)} \qquad (5.39b)$$

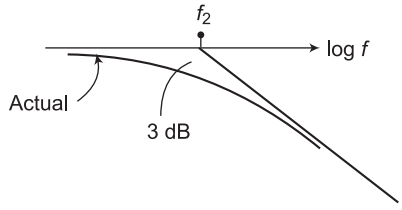


Fig. 5.42 Asymptote slope

Break frequency, $f_2 = \dfrac{1}{2\pi RC}$ (5.40)

Asymptote slope = –20 dB/decade as shown in Fig. 5.42.

✦ High-Frequency Response

❑ ***BJT*** The ac circuit of high frequency is shown in Fig. 5.43, the coupling and bypass capacitors are shorts.

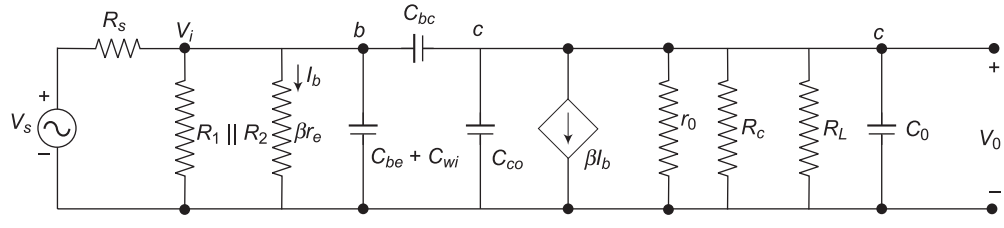


Fig. 5.43 BJT CE amplifier ac circuit at high frequencies

Various capacitances are

- Device capacitances C_{be}, C_{bc}, C_{ce}
- Stray wiring capacitances C_{wi} (input side) C_{wo} (output side)
- Load capacitances, C_L

The capacitances C_{bc} is converted to C_{mi} input side and C_{mo} output side by application of Miller theorem for simplification of circuit configuration.

Net input capacitances, $C_i = C_{wi} + C_{bc} + C_{mi}$

Net output side capacitances, $C_0 = C_{w0} + C_{ce} + C_{mo}$, C_{ce} is of negligible orders.
The resistances are combined by Thevenine theorem as

$$R_{THi} = R_s \parallel R_1 \parallel R_2 \parallel \beta r_e \tag{5.41}$$

$$R_{H0} = r_0 \parallel R_c \parallel R_L \tag{5.42}$$

The Thevenin equivalent circuit is drawn in Fig. 5.44.

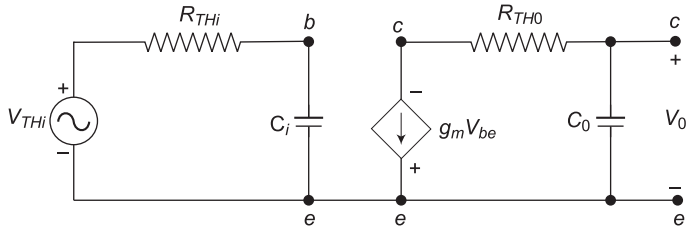


Fig. 5.44 Thevenin equivalent circuit

The break frequencies are then

$$f_{Hi} = \frac{1}{2\pi R_{THi} C_i} \tag{5.43}$$

and

$$f_{H0} = \frac{1}{2\pi R_{TH0} C_0} \tag{5.44}$$

❑ ***FET*** If can be similarly shown that

$$f_{Hi} = \frac{1}{2\pi R_{THi} C_i} \tag{5.45}$$

where $R_{THi} = R_s \parallel R_G$

$C_i = C_{wi} + C_{gs} + C_{mi}$

and

$$f_{H0} = \frac{1}{2\pi R_{TH0} C_i} \tag{5.46}$$

where $R_{TH0} = R_s \parallel R_D \parallel r_d$

$C_0 = C_{w0} + C_{ds} + C_{m0}$, C_{ds} in of negligible order C_{mi} and C_{m0} arise from C_{gd}.

EXAMPLE 5.16 The circuit analysis of a BJT RC coupled amplifier is drawn in Fig. 5.45. The data given

$V_{CC} = 20$ V
$R_1 = 220$ kΩ. $R_2 = 58$ Ω
$R_E = 2.2$ kΩ. $R_c = 5.8$ kΩ
$\beta = 150$. $R_s = 1$ kΩ

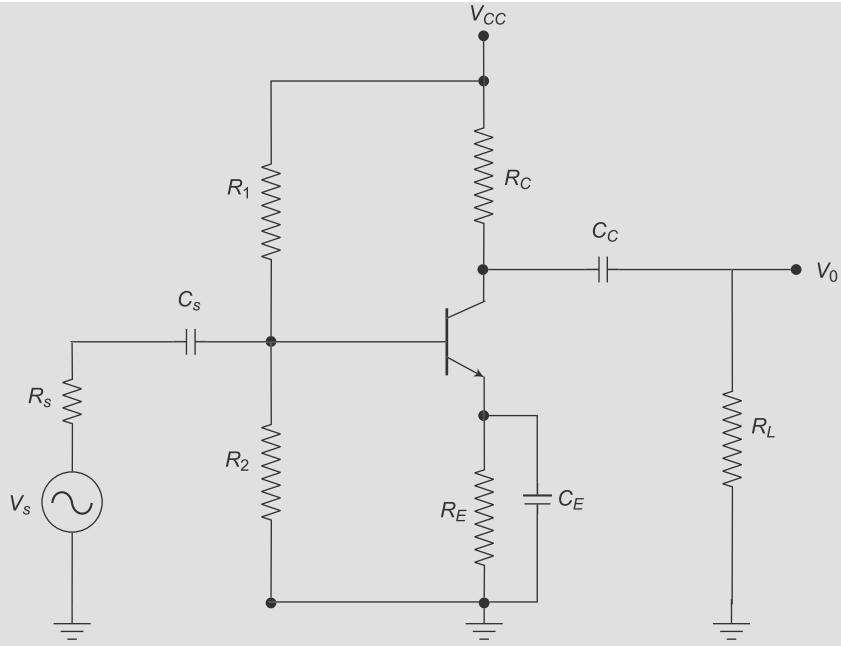


Fig. 5.45

Solution *DC analysis yields*

$$I_E = 1.38 \text{ mA}$$

$$r_e = \frac{26}{1.38} = 18.84\ \Omega$$

$$\beta r_e = 150 \times 18.84 = 2.864\ \text{k}\Omega$$

$$C_s = 10\ \mu\text{F},\ C_c = 1\ \mu\text{F},\ C_E = 20\ \mu\text{F}$$

Mid-band gain

$$A_v = \frac{V_0}{V_i} = \frac{-R_c \parallel R_L}{r_e} = -\frac{\frac{5.8 \times 2.2}{5.8 + 2.2}}{18.84 \times 10^{-3}} = -84.7$$

$$Z_i = R_1 \parallel R_L \parallel \beta r_e = \frac{1}{\frac{1}{220} + \frac{1}{5.8} + \frac{1}{2.864}} = 1.9\ \text{k}\Omega$$

$$\frac{V_i}{V_s} = \frac{1.9}{1 + 1.9} = 0.655$$

$$A_v = \frac{V_0}{V_i} = -84.7 \times 0.655 = -55.5$$

Low-frequency breaks

$$f_{Ls} = \frac{1}{2\pi(R_s + R_i)C_s} = \frac{10^6}{2\pi(2+1.9)\times 10^3 \times 10} = 5.5 \text{ Hz}$$

$$f_{L0} = \frac{1}{2\pi(R_c + R_L)C_c} = \frac{10^6}{2\pi(5.8+2.2)\times 10^3 \times 1} = 20 \text{ Hz}$$

$$f_{LE} = \frac{1}{2\pi R_e C_E} = \frac{10^6}{2\pi \times 24.5 \times 20} = 325 \text{ Hz}$$

R_e is compiled below

$$R_e = R_E \| \left(\frac{R_s}{\beta} + r_e\right), R_s' = R_s \| R_1 \| R_2 = \frac{1}{\frac{1}{1}+\frac{1}{220}+\frac{1}{5.8}} = 0.85 \text{ k}\Omega$$

$$= \frac{2.2\times 10^3 \times 24.5}{2.2\times 10^3 + 24.5}, \frac{R_s'}{\beta} = \frac{0.85}{150} = 5.7\ \Omega, \frac{R_s'}{\beta} + r_e = 5.7 + 18.84 = 24.5\ \Omega$$

$$= 24.23\ \Omega$$

Therefore, for bandwidth calculations f_1 = 325 Hz (highest of the three break frequencies)

EXAMPLE 5.17 The circuit of a DMOSFET is drawn in Fig. 5.46. The data is an in Example 5.6 repeated below.

R_1 = 100 MΩ, R_2 = 10 MΩ, R_D = 1.6 kΩ, R_S = 720 Ω, V_{DD} = 16 V

As found in Example 5.6, g_m = 5.7 mS

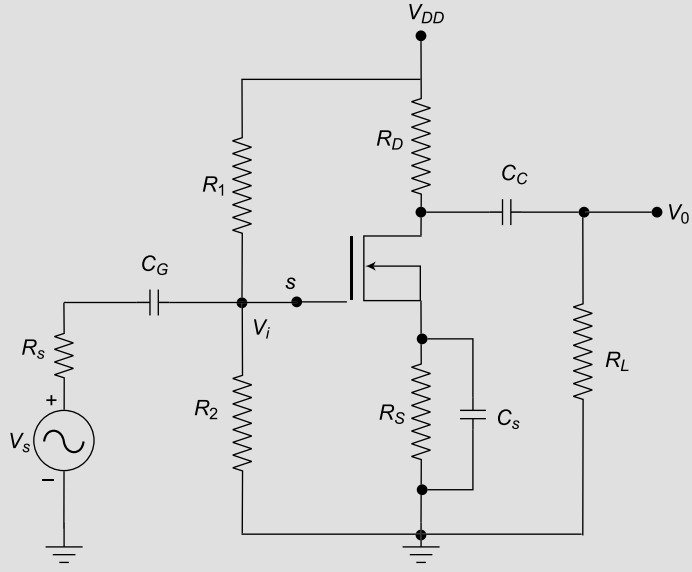


Fig. 5.46

Additional data:

R_S = 10 kΩ, $r_d = \infty$, R_1 = 2.2 kΩ, C_G = 0.01 μF, C_c = 0.5 μF, C_s = 2 μF

Determine the corner frequencies in the low frequency end.

Solution

$$R_1 \| R_2 = \frac{100 \times 10}{110} = 9.1 \text{ M}\Omega$$

$$f_{LG} = \frac{1}{2\pi(R + R_1 \| R_2)C_G} = \frac{10^6}{2\pi \times 9.1 \times 10^6 \times 0.01} = 1.75 \text{ Hz}$$

$$f_{L0} = \frac{1}{2\pi(R_D + \text{R}_1)C_C} = \frac{10^6}{2\pi\,(1.6 + 2.2) \times 10^3 \times 0.51} = 83.7 \text{ Hz}$$

$$f_{LS} = \frac{1}{2\pi R_{eq} C_s} = \frac{10^6}{2\pi \times 0.141 \times 10^3 \times 2} = 564 \text{ Hz}$$

$$R_{\text{eq}} = R_s \| \frac{1}{g_m} = 0.72 \| \frac{1}{5.7} = 0.72 \| 0.175 = 0.141 \text{ k}\Omega$$

Thus, low-frequency break is $f_1 = f_{LS} = 564$ Hz

SUMMARY

➢ BJT and amplifiers are introduced. Analysis of different configurations are carried out. Current and voltage gain are calculated. Frequency response is discussed.

EXERCISES

➥ Review Questions

1. Derive the expressions of 'voltage gain', 'current gain', 'input impedance' and 'output impedance', for a BJT using low frequency *h*-parameter model for (a) *CE* configuration, (b) *CB* configuration and (c) *CC* configuration.
2. Compare the performance of a BJT as an amplifier in *CE*, *CB* and *CC* configuration.
3. Derive the expressions of 'voltage gain', 'current gain', 'input impedance' and 'output impedance'. For a BJT using the approximation *h*-parameter model for (a) *CE* configuration, (b) *CB* configuration, and (c) *CC* configuration.
4. What is an amplifier? What are the various types of amplifiers?
5. Draw the circuit diagram of a *CE* amplifier and explain its working.
6. Draw the ac equivalent of a *CE* amplifier with fixed-bias using *h*-parameter model and 'r_e' model and derive the equations for input impedance, voltage gain and current gain.
7. Draw the ac equivalent circuit of a *CE* amplifier with 'voltage divider' bias using *h*-parameters model and r_e model and also derive the equations for 'input impedance' and 'voltage gain'.
8. Draw the small-signal model of BJT for low frequency and high-frequency regions.

9. Justify the validity of approximate hybrid model applicable in low-frequency regions.
10. Draw the r_e model for a BJT in *CE* configurations and derive the equations for voltage gain and current gain.
11. A *CE* amplifier is drawn by a voltage source of internal resistance $R_s = 1000\ \Omega$ and the load impedance is a resistance 'R_L' = 1200 Ω. The *h*-parameters are $h_{ie} = 1.2\ \text{k}\Omega$, $h_{re} = 2 \times 10^{-4}$, $h_{fe} = 60$ and $h_{oe} = 25\ \mu\text{A/V}$. Compute the current gain A_I, input resistance R_i, voltage gain A_v and output resistance R_o using exact analysis.
12. A *CE* amplifier stage uses an *NPN* transistor having β_o = 125 and is biased at I_{CQ} = 1 mA. If $R_s = 300\ \Omega$, $R_c = 1.2\ \text{k}\Omega$, determine A_v and R_i.
13. For any single-stage amplifier, prove that, $\left[R_i = \dfrac{h_i}{1 - h_r A_v} \right]$
14. Draw the ac equivalent circuit of *CE* and *CC* amplifiers subject to $R_L = 0$. Show that the input impedances of the two circuits are identical.
15. Which of the three BJT configurations has (a) highest R_i, (b) lowest R_i, (c) highest R_o, (d) lowest R_o, (e) highest A_v, and (f) lowest A_v?
16. What is meant by-frequency response of a voltage amplifier?
17. Explain the meaning of a break frequency as half-power frequency.
18. What is meant by-bandwidth of an amplifier? What is its significance?
19. What are low-frequency breaks and their censes?
20. What are the causes of high-frequency breaks?
21. How much is the dB gain drop at break frequency?
22. Draw the BJT amplifier's ac circuit showing all elements. Identify the capacitors which effect the low-frequency response and those which effect the high-frequency response. Explain why.
23. Repeat Question 22 for an FET amplifier.
24. How is capacitance connecting *b* and *c* in BJT amplifier ac circuit for finding high-frequency break.
25. Repeat Question 24 for FET amplifier for capacitance connecting *d* and *s*.
26. Draw the small-signal model of common base configuration.
27. How do you find the value of r_e base-emitter resistance?
28. Draw the small-signal model of common-emitter configuration. What is the meaning of each circuit element? What is output resistance r_o?
29. Outline the steps in a determining the ac gain of a BJT based CE amplifier. Why do we need the biasing (dc) analysis of the circuit?
30. What is an *RC*-coupled amplifier?
31. In ac analysis of an *RC* coupled amplifier, the coupling capacitors and by-pass capacitors can be assumed to the short circuit in mid band frequencies. Elaborate.
32. In a CE-amplifier, the output is phase reversed. Elaborate.
33. What is an emitter follower? How is it different from a CE amplifier? What are its applications?
34. What is g_m, transconductance?
35. Write the expression for g_m for JFET, DMOSFET and EMOSFE. Elaborate the symbols used.
36. What contributors the output resistance r_d?
37. Draw the common source circuit model FETs.
38. Draw the circuit of a source follower with feedback bias.
39. Draw the ac circuit of the source follower. What are its properties in comparison to common-source circuit?

➥ Problems

1. For the CE amplifier of Fig. 5.47, determine:
   (a) I_B, I_c and r_e (b) Z_i and Z_o (c) A_v and A_i

Given: $\beta = 100$, neglect r_o.

2. For the CE amplifier of Fig. 5.48, BJT has $\beta = 110$ and $r_o = \infty$

(a) Determine r_e, Given $I_E = 7.36$ µA

(b) Calculate Z_i and Z_o

(c) Find A_v and A_i

Hint: R_E is shorted

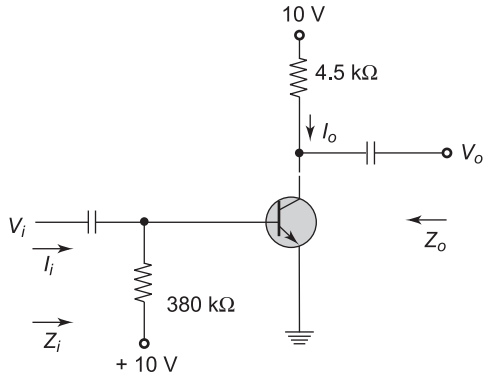


Fig. 5.47

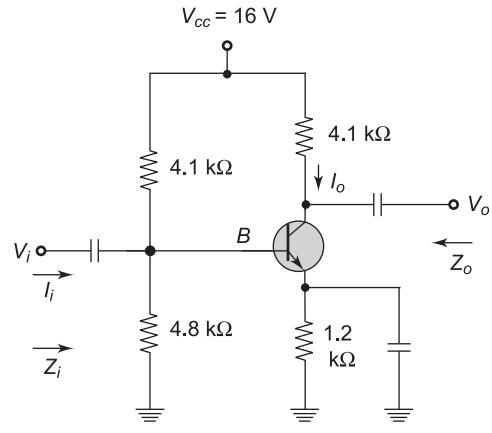


Fig. 5.48

3. For the BJT amplifier circuit of Fig. 5.49,

(a) Determine the break frequencies in the low-frequency end due to C_{c1}, C_{c2} and C_E-what is the cut-off frequency?

(b) Calculate the mid-band voltage gain.

Data: $C_{C1} = 1$ µF $= C_{C2}$, $C_E = 50$ µF, $\beta = 100$

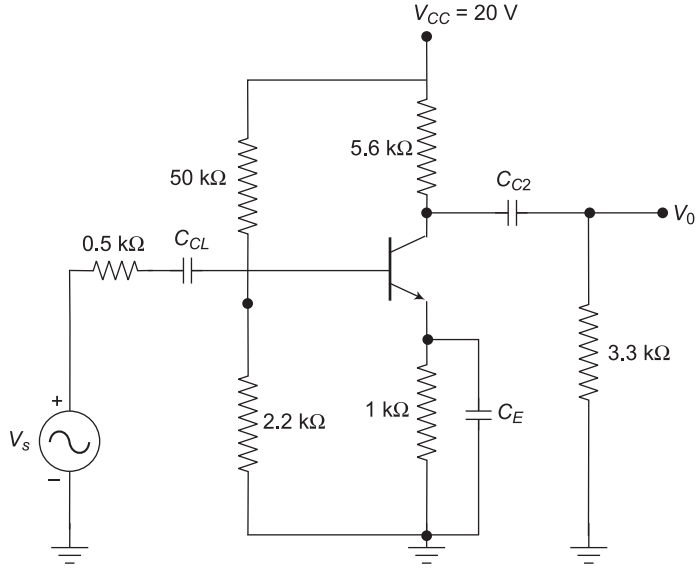


Fig. 5.49

4. For the FET amplifier circuit of Fig. 5.50,
   (a) Find the mid-frequency voltage gain
   (b) Find the lower cut-off frequency
5. Calculate the size of the emitter resistance bypass capacitor to provide a low-frequency 3 dB point of 100 Hz for a CE BJT amplifier.
   Given: $R_E = 1.5\ k\Omega$, $\beta = 100$, $r_e = 10\ \Omega$, $R_s = 1\ k\Omega$, $R_B = \infty$.
6. For the BJT amplifier circuit of Fig. 5.51, determine:
   (a) The lower half-power frequency due to (i) C_{C1} (ii) C_E
   (b) Hence, determine the overall lower cut-off frequency of the amplifier circuit.

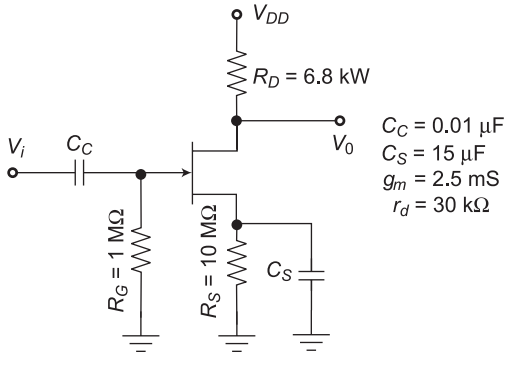


Fig. 5.50

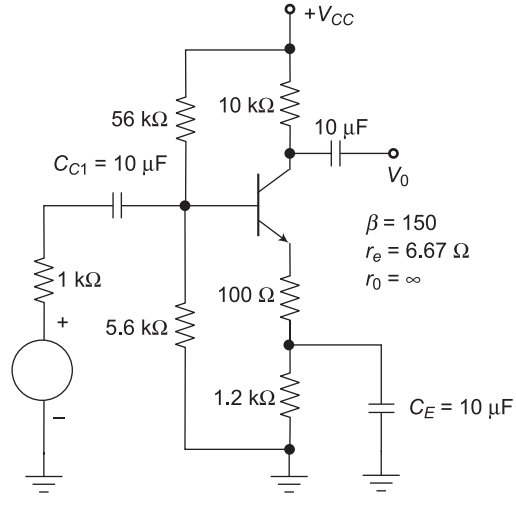


Fig. 5.51

7. For the FET amplifier circuit of Fig. 5.52,
   (a) Determine the lower 3 dB cut-off frequencies due to capacitor C_{C1} and C_S. Then find the effective cut-off frequency.
   (b) Find the mid-frequency gain.

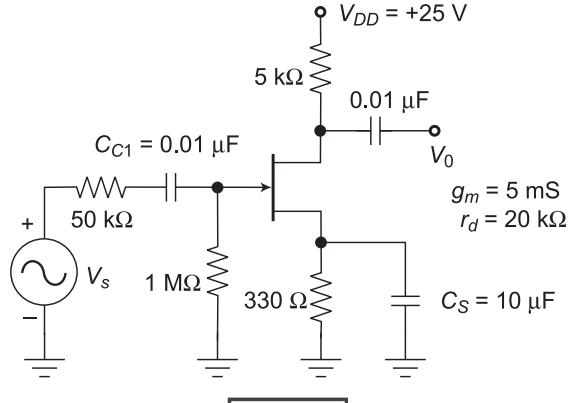


Fig. 5.52

8. The amplifier in Fig. 5.53 is to have a lower cut-off frequency of 20 Hz. Assuming I_C = 1.18 mA and C_E, C_{C2} to be large, draw an approximate circuit model and specify C_{C1} (β = 100).

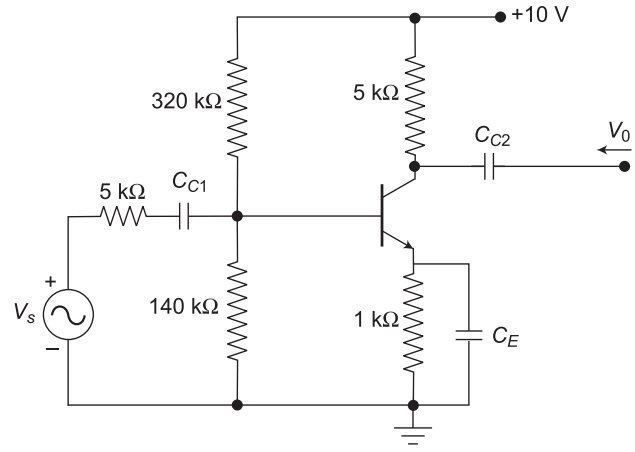


Fig. 5.53

9. For the JFET drain characteristics drawn in Fig. 5.54,
(a) Find r_d for V_{GS} = 0
(b) Find g_{m0} for V_{DS} = 10 V

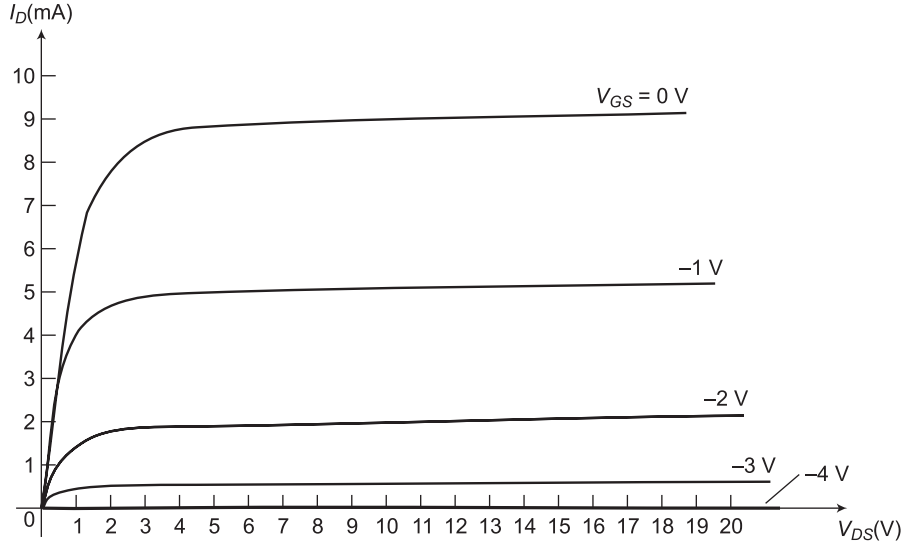


Fig. 5.54

10. A JFET has parameters I_{DSS} = 10 mA, V_p = – 6 V. If is biased at V_{GSQ} = V_p/4. Determine $g_{m.}$

11. A JFET has r_d = 100 kΩ and its ideal voltage gain is –250. What is the value of g_m?

12. A JFET fixed bias amplifier circuit is drawn in Fig. 5.55. JFET data: I_{DSS} = 12 mA, V_p = –4 V, r_d = 25 kΩ. Determine Z_i, Z_0 and A_v.

13. A JFET amplifier with voltage divider bias is shown in Fig. 5.56 JFET parameters are
I_{DSS} = 12 mA, V_p = – 4 V, r_d = 100 kΩ
Determine Z_i, Z_0 and A_v

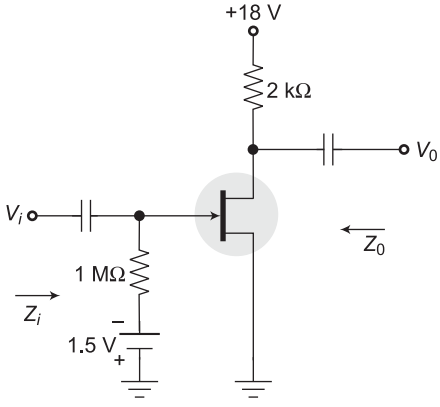


Fig. 5.55

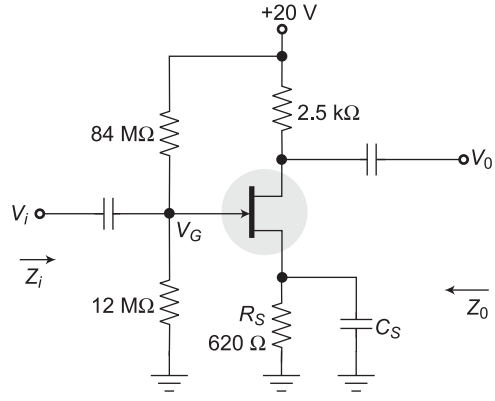


Fig. 5.56

14. The DMOFET amplifier of Fig. 5.57 has g_m = 6 mS, r_d = 30 kΩ. Determine Z_i, Z_0 and A_v.

15. For EMOSFET drain feedback bias amplifier determine Z_i, Z_0, A_v
EMFOST data
V_T = 3.5 V, $k = 0.3 \times 10^{-3}$ A/V^2, r_d = 100 kΩ

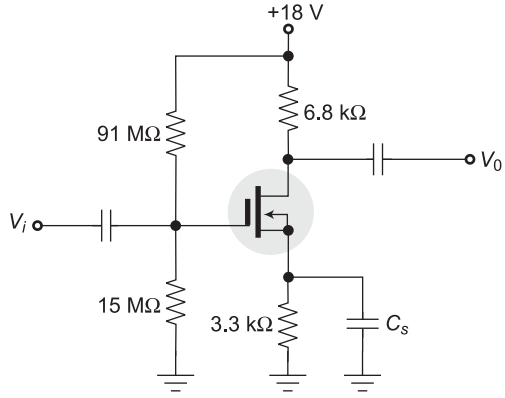


Fig. 5.57

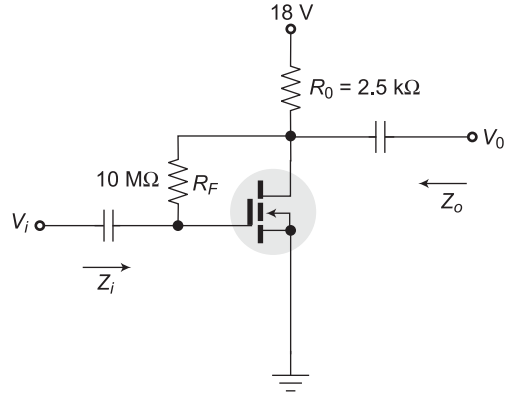


Fig. 5.58

Hint: ac circuit

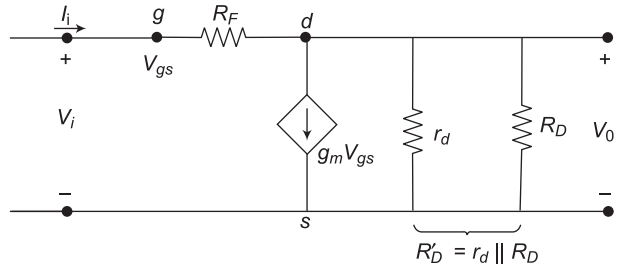


Fig. 5.59

At node *d*

$$I_i = \frac{V_i - V_0}{R_F} = V_i g_m + \frac{V_0}{R_0} \quad \text{(i)}$$

$$V_i\left[\frac{1}{R_F} - g_m\right] = V_0\left[\frac{1}{R_F} + \frac{1}{R'_D}\right]$$

$$A_v = \frac{V_0}{V_i} = \frac{\left[\frac{1}{R_F} - g_m\right]}{\left[\frac{1}{R_F} + \frac{1}{R'_g}\right]} \quad \text{(ii)}$$

Substituting V_0 from Eq. (ii) in Eq. (i)

$$I_i = V_i\left\{g_m + \frac{1}{R'_D}\frac{\left[\frac{1}{R_F} - g_m\right]}{\left[\frac{1}{R_F} + \frac{1}{R'_D}\right]}\right\}$$

$$Z_i = \frac{V_i}{I_i} = \frac{R_F + R'_D}{1 + g_m R'_D}; \text{ upon simplification}$$

$$Z_0 = R_F \parallel R'_D$$

16. Given $h_{fe} = 50$ and $h_{ie} = 0.83$ kΩ. Find the current gain (h_{fb}) and input impedance (h_{ib}) for a transistor in CB configuration.

➥ Multiple-Choice Questions

1. The voltage gain of a common-base amplifier is
(a) zero (b) unity (c) less than unity (d) greater than unity

2. For a common-base transistor amplifier having input resistance (R_i) and output resistance (R_o), which of following statements holds good?
(a) R_i is high, R_o is low (b) R_i is low, R_o is high
(c) R_i and R_o are both medium (d) None of these

3. When an emitter-bypass capacitor in a common-emitter amplifier is removed, its ..... is considerably reduced.
(a) output load resistance (b) voltage gain
(c) collector-current (d) emitter current

4. Determine the quiescent levels of I_C and V_{CE} for the network in Fig. 5.60.

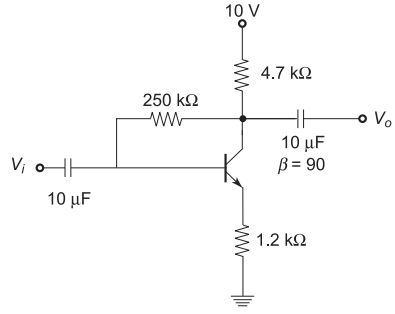


Fig. 5.60

(a) 1.07 mA, 3.69 V (b) 1.07 A, 3.69V (c) 1 mA, 4.5 V (d) 2 mA, 5 mV

5. The dc load line of a transistor circuit
(a) is a graph between I_C and I_B (b) is a curved line
(c) is a graph between I_C and V_{CE} (d) does not contain the locating point

6. When a BJT is employed as an amplifier, it operates
(a) in saturation (b) in cut-off
(c) well into saturation (d) over the active region.

7. The negative part of the output signal in a transistor circuit-starts clipping, if Q-point of the circuit moves.
(a) towards the cut-off point (b) towards the saturation point
(c) towards the center of the load line (d) none of the above

8. Thermal runaway will take place if the quiescent point is such that
(a) $V_{CE} > ½\ V_{CC}$ (b) $V_{CE} < 2V_{CC}$
(c) $V_{CE} < ½\ V_{CC}$ (d) None of these

9. The stability factor of a BJT circuit is defined as

(a) $S = \frac{dI_E}{dI_B}$ (b) $S = \frac{dI_C}{dI_B}$ (c) $S = \frac{dI_C}{dI_E}$ (d) $S = \frac{dI_C}{dI_{CO}}$

10. Low-frequency hybrid-TT model of transistor parameter transconductance g_m is

(a) $g_m = \frac{I_C}{25}$ (b) $g_m = \frac{I_C}{2}$ (c) $g_m = I_C$ (d) $g_m = \frac{I_C}{100}$

ANSWERS

✦ Review Questions

11. A_I = –58.25, R_i = 1.186 kΩ, A_v = –58.937, R_o = 51.162 kΩ, –43.78, 3.425 kΩ

✦ Problems

1. (a) 24 μA, 2.4 mA, 10.7 Ω; (b) Z_i = 1.067 kΩ, Z_o = 4.5 kΩ; (c) A_v = –4.8, A_i = –99.74
2. (a) 0.81 mA, 31.7 Ω; (b) 1.92 kΩ; (c) –129, –0.225
3. (a) 77.6 Hz, 17.9 Hz, 112.4 Hz, f_2 = 112.4 Hz; (b) 66.5
4. –13.9, 15.9 Hz **5.** 81.45 μF
6. 3.27 Hz, 15.38 Hz, 15.38 Hz
7. (a) $f_{c_{c1}}$ = 15.16 Hz, $f_{c_{c2}}$ = does not exist, f_{c_s} = 97.64 Hz, f_L = 97.64 Hz
8. 1.13 μF **9.** (a) ≈ 50 kΩ; (b) 7.6 mS
10. 2.5 ms **11.** 2.5 ms
12. 1 MΩ, 2 kΩ, –6 13. 10.5 MΩ, 2.43 kΩ, –10.2
14. 12.88 MΩ, 5.54 kΩ, 33.24 **15.** 1.524 MΩ, 2.44 kΩ, –4.36
16. –0.98, 16.27 Ω

✦ Multiple-Choice Questions

1. (d) **2.** (b) **3.** (b) **4.** (a) **5.** (c) **6.** (d) **7.** (b) **8.** (c) **9.** (d) **10.** (a)

CHAPTER 6

Op-Amp

GOALS AND OBJECTIVES

- Introduction of op-amp's and its characteristics
- Block diagram of basic op-amp circuit and its architecture
- Detailed explanation of Schmitt trigger configuration
- Operation of constant-gain multiplier
- Analysis of basic logarithmic amplifier using diode and transistor
- Drawing of inverting and non-inverting configurations
- Illustration of inverting summing amplifier and op-amp differential gain configurations

6.1 INTRODUCTION

Op-Amp is the standard acronym for operational amplifier. This name comes from the early days of amplifier design when it was used in analog computers, with some added external components, when the fundamental amplifier was used to perform various mathematical operations like addition, subtraction, integration, etc.

It is one of the fundamental building blocks of linear design which consists of two input terminals out of which one inverts the phase of the signal, and the other conserve the phase and one output terminal. Figure 6.1 given above shows the standard symbol of an op-amp. Two power supply terminals are not shown here, obviously which are required for operation. Op-amps are often divided into two types: (i) ones that use both a positive and negative power supply and (ii) others use only a single, usually positive power supply. Since all op-amps have only two supply pins, this distinction is often unnecessary.

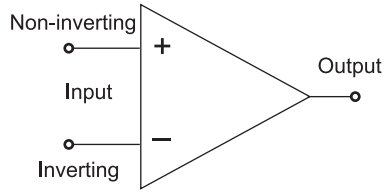
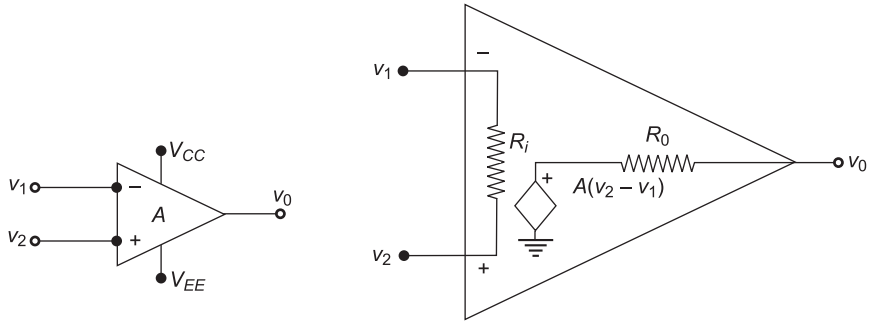


Fig. 6.1 Standard symbol for Op-Amp

An op-amp is perhaps the most important and versatile analog IC; it is applied in analog signal processing and analog filtering. The symbol of an op-amp is drawn in Fig. 6.2(a). It has two input terminals, inverting (–) and non-inverting (+), and the output is a single terminal. The voltages are measured w.r.t. ground reference, which may not be always shown. The simplified circuit model of an op-amp is drawn in Fig. 6.2(b). It has a gain of A, input resistance R_i and output resistance R_o. An ideal op-amp has $A = \infty$, $R_{in} = \infty$ and $R_o = 0$.



(a) Symbol, power supply need not be indicated

(b) Circuit model

Fig. 6.2 op-amp – IC 741

Infinite gain means

$$v_o = A(v_2 - v_1)$$

or $$v_2 - v_1 = \frac{v_o}{A = \infty} = 0$$

or $$v_2 = v_1 \qquad (6.1)$$

It means that inverting and non-inverting terminals are at the same potential.

OP-AMP Architecture

The block diagram of an op-amp is drawn in Fig. 6.3. Various blocks are described under.

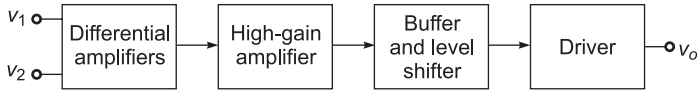


Fig. 6.3 Block diagram of an operational amplifier

✦ Differential Amplifier (DA)

It amplifies the difference of inputs, $v_d = (v_2 - v_1)$, where the differential mode gain is A_d. If the same input $v_c = \dfrac{v_2 + v_1}{2}$ is applied to both the terminals, the amplification is labelled A_c. As the difference of common-mode signals is zero, A_c should be zero but in a practical differential amplifier, it is a small value.

The ratio of these gains is called

$$\text{Common-Mode Rejection Ratio (CMRR)} = \frac{A_d}{A_c}; \tag{6.2}$$

Being very large it is measured in dB.

We will not deal with the circuit of a DA (Differential Amplifier) except to state that it is two BJT [or MOSFET] amplifiers connected in opposition so as to amplify the difference of two input signals.

- It is a high-gain amplifier (direct coupled, highly stabilised).
- The buffer is emitter follower for matching the load. If the output is nonzero for zero input, the level shifts make it zero.
- Driver is power amplifier.

The most popular op-amp and the industry standard is the **IC 741 op-amp**. It is an analog (linear) IC. Its power supply voltages are $+V_{CC}$ and $-V_{EE}$ with $V_{CC} = V_{EE} = 15$ V. The range of output is $-10 \le v_0 \le +10$ V. Further, it is short-circuit protected. Typical parameters for the 741 op-amp are

$A = 200{,}000$ $\quad R_{in} = 2\ \text{M}\Omega$, $\quad R_0 = 75\ \Omega$

$CMRR = 80\text{-}100$ dB $\quad 100\ \text{dB} \rightarrow 10^5$

The output of the op-amp is

$$v_o = A_d\,(v_2 - v_1) + A_c \left(\frac{v_2 + v_1}{2}\right)$$

or

$$v_o = A_d\,V_d + A_c\,V_c = A_d V_d \left[1 + \frac{1}{CMRR} \cdot \frac{V_c}{V_d}\right] \tag{6.3}$$

As CMRR is very high (10^5),

$$V_o = A_d\,V_d = A(v_2 - v_1) \tag{6.4}$$

Thus, the op-amp gain is $A = A_d$; common-mode signal is of no consequence.

Other important parameters of the op-amp are

- **Input offset voltage** Differential voltage (V_{os}) needed to make $v_o = 0$
  Typical value $V_{os} = 1$ mV
- **Bias current** $I_B = \dfrac{I_{B1} + I_{B2}}{2}$ to make $v_o = 0$, I_{B1} and I_{B2} are base currents of the two transistors

  Typical value is 80 mA.

✦ Frequency Response

Some *RC* couplings are provided in an op-amp circuit so as to reduce the gain at high frequencies. Otherwise, a positive feedback occurs from stray capacitances causing high-frequency oscillations. The op-amp acts like a low-pass filter with a break frequency at $f_h = 10$ Hz. The typical frequency response (Bode plot) is drawn in Fig. 6.4.

From the Bode plot,

$$20 \log \frac{f_T}{f_h} = \text{dB(A)} = 20 \log A,$$

Therefore,

$$\frac{f_T}{f_h} = A$$

or $Af_h = f_T$ = unity gain (0 dB) frequency = constant

Af_h = gain-bandwidth product

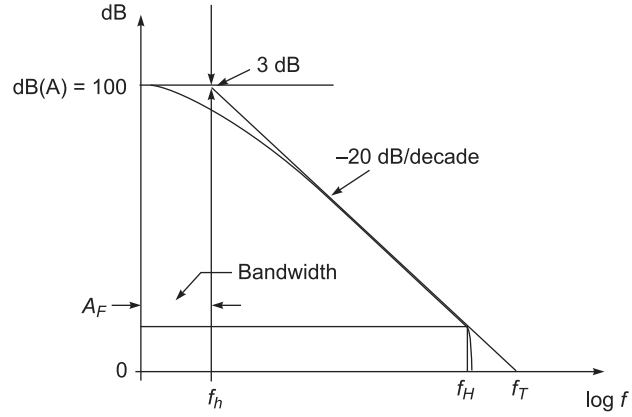


Fig. 6.4 Op-amp frequency response

If negative feedback around op-amp it reduces the gain to A_F, then

$$A_F f_H = f_T \tag{6.5a}$$

or $$f_H = \frac{f_T}{A_F} \gg f_h \tag{6.5b}$$

Due to reduction in gain caused by feedback, the bandwidth increases from f_h to f_H.

✦ Slew Rate

Because of the presence of capacitances, which is *RC* combination, it can charge at a limited rate, and the maximum rate of change of output of the op-amp is limited to

$$\text{Slew rate } S = \left.\frac{dv_o}{dt}\right|_{\max} \tag{6.5c}$$

Its typical value is 0.5 V/μs as shown in Fig. 6.5.

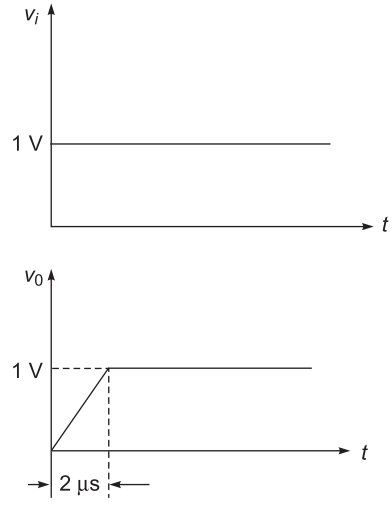


Fig. 6.5 Illustrating effect of slew rate

For a sinusoidal signal,

$$V_o = V_m \sin \omega$$

$$S = \left.\frac{dv_o}{dt}\right|_{\max} = V_m \omega$$

or $$S = 2\pi f V_m \tag{6.6}$$

It is a combination of frequency and peak value of output.

6.2 | BASIC OP-AMP CIRCUITS

✦ Inverting Amplifier

The op-amp circuit with voltage feedback is shown in Fig. 6.6.

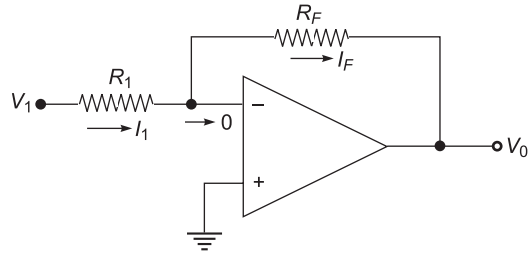


Fig. 6.6 Inverting amplifier

✦ Virtual Ground

As the (+) terminal is grounded, the (–) terminal is virtually at ground potential.

- Current into op-amp terminal is always zero as $R_{in} = \infty$

  Therefore,

$$I_F = I_1 \tag{6.7a}$$

$$\frac{0 - V_o}{R_F} = \frac{V_1}{R_1} \tag{6.7b}$$

The gain is

$$\frac{V_o}{V_1} = -\frac{R_F}{R_1} \text{ or } V_0 = -\left(\frac{R_F}{R_1}\right) V_1 \tag{6.8}$$

We find that output voltage is the negative of the input voltage, i.e. it is inverted.

✦ Non-inverting Amplifier (Fig. 6.7)

As the two input terminals of op-amp must be at same potential,

$$V(A) = V_1$$

Applying KCL at A,

$$\frac{V_1}{R_1} + \frac{V_1 - V_0}{R_F} = 0 \tag{6.9a}$$

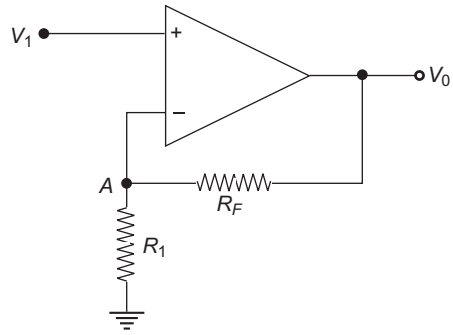


Fig. 6.7 Non-inverting amplifier

Reorganising, we get

$$\frac{V_0}{V_1} = 1 + \frac{R_F}{R_1} \text{ or } V_0 = +\left(1 + \frac{R_F}{R_1}\right) V_1 \tag{6.9b}$$

As V_0 has the same sign as V_1, the amplifier is non-inverting.

Note: In analysis of op-amp circuits, we note the following:

- Inverting and non-inverting terminals are at the same potential.
- The current into these terminals is zero.

✦ Transfer Characteristic of OP-AMP

An op-amp is linear with high gain over a very small value of $V_i = \in$ as shown in Fig. 6.8 beyond which it saturates, $V_0 = V_{CC}$.

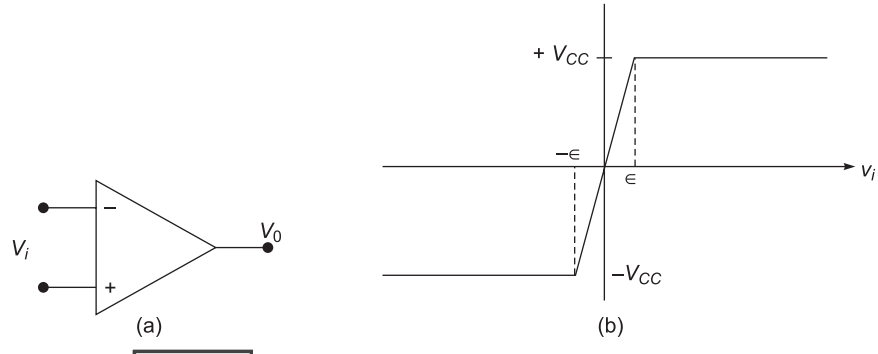


Fig. 6.8 (a) Open-loop op-amp (b) Transfer characteristic

Applications of op-amp are linear or nonlinear according to the region of its operation.

✦ Linear Applications of OP-AMP

❑ ***Summer Circuit*** The op-amp circuit is drawn in Fig. 6.9.

$$I_1 + I_2 + I_3 = I_F \tag{6.10a}$$

$$\frac{v_1}{R_1} + \frac{v_2}{R_2} + \frac{v_3}{R_3} = -\frac{v_0}{R_F} \tag{6.10b}$$

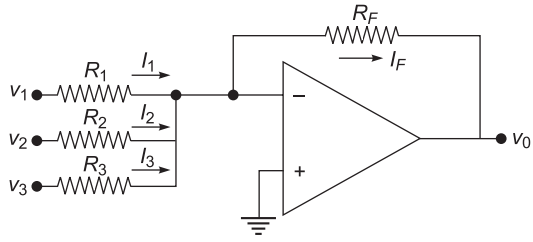


Fig. 6.9 Summer circuit

Reorganising,

$$v_0 = -\left\{\left(\frac{R_F}{R_1}\right)v_1 + \left(\frac{R_F}{R_1}\right)v_2 + \left(\frac{R_F}{R_3}\right)v_3\right\} \tag{6.11a}$$

For $R_1 = R_2 = R_3 = R_F$

$$v_0 = -(v_1 + v_2 + v_3) \tag{6.11b}$$

$$= -(\text{sum of voltages})$$

❑ ***Subtractor*** The op-amp circuit is shown in Fig. 6.10.

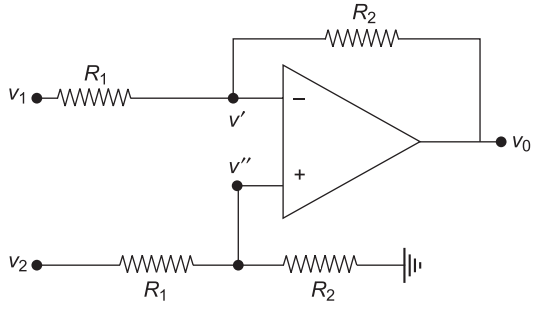


Fig. 6.10 Subtractor

$$v' = \left(\frac{R_2}{R_1 + R_2}\right)v_2 \tag{6.12a}$$

$$\frac{v_1 - v'}{R_1} = \frac{v' - v_0}{R_2} \tag{6.12b}$$

or $$\frac{v_0}{R_2} = v'\left(\frac{1}{R_1} + \frac{1}{R_2}\right) - \frac{v_1}{R_1} \tag{6.12c}$$

Substituting v' from Eq. (6.12a) in Eq. (6.12c),

$$\frac{v_0}{R_2} = \left[\frac{R_1 + R_2}{R_1 R_2} \cdot \frac{R_2}{R_1 + R_2}\right]v_2 - \frac{v_1}{R_1} = \frac{v_2 - v_1}{R_1}$$

or $$v_0 = \frac{R_2}{R_1}(v_2 - v_1) \quad \text{or} \quad v_0 = v_2 - v_1 \text{ for } R_1 = R_2 \tag{6.13}$$

✦ Source Converters

❑ ***Voltage Follower*** Refer op-amp circuit of Fig. 6.11. As terminal voltages must be equal, so

$$v_0 = v_s$$

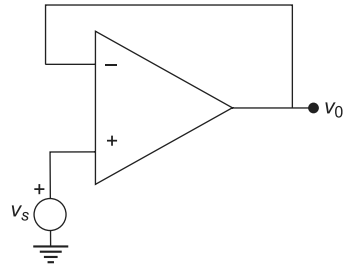


Fig. 6.11 Op-amp circuit

❑ ***Current-to-voltage Converter*** (Transresistance amplifier, Fig. 6.12)

$$V_0 = -R_F I_i$$

The disadvantage is that the bias current I_B gets added to I_i, introducing a small error.

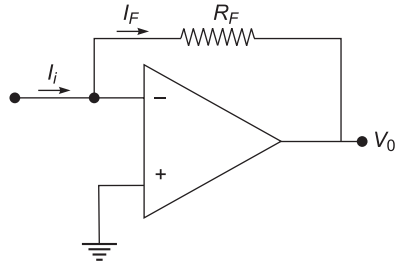


Fig. 6.12 Transresistance amplifier

❑ ***Voltage-to-current Converter*** (Transconductance amplifier)

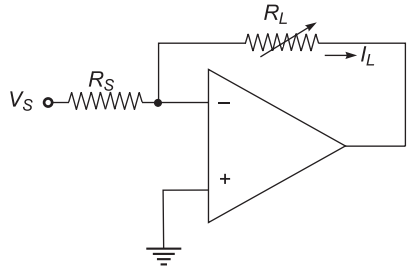


Fig. 6.13 Voltage-to-current converter

Op-amp current of Fig. 6.13,

$$I_L = -\left(\frac{V_S}{R_S}\right)$$

For a grounded load, the circuit is drawn in Fig. 6.14.

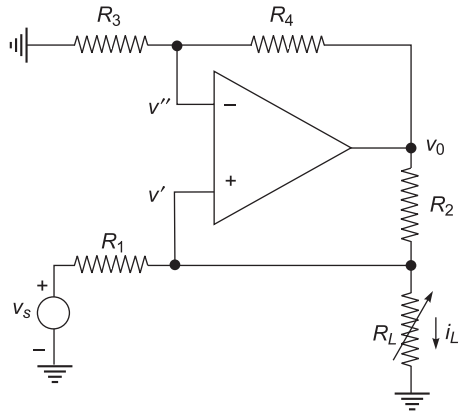


Fig. 6.14 Circuit for grounded load

Prove that $i_L = \dfrac{v_s}{R_1}$

Proof

$$v' = \left(\frac{R_3}{R_3 + R_4}\right) v_0$$

$$\frac{v_s - v'}{R_1} + \frac{v_0 - v'}{R_2} = i_L$$

$$\frac{v_s}{R_1} - \left(\frac{1}{R_1} + \frac{1}{R_2}\right) v' + \frac{v_0}{R_2} = i_L$$

$$\frac{v_s}{R_1} - \left[\left(\frac{R_1 + R_2}{R_1 R_2}\right) \frac{R_3}{R_3 + R_4} - \frac{1}{R_2}\right] v_0 = i_L \qquad (\therefore \quad R_1 = R_2 = R_3 = R_4)$$

Let $R_4 = R_3 = R_2 = R_1$

Hence, $$\frac{V_S}{R_1} = i_L$$

❑ ***Constant Voltage Source (Fig. 6.15)***

$$\frac{V_S}{R_S} = -\frac{V_0}{R_F}$$

or $$V_0 = -\left(\frac{R_F}{R_S}\right) V_S$$

Independent of load resistance

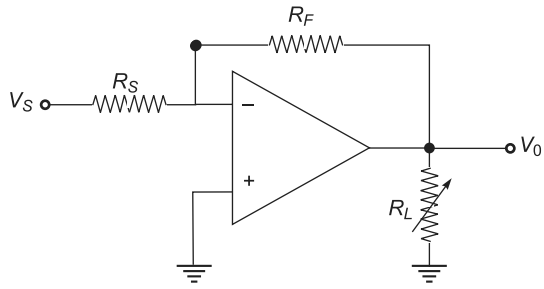


Fig. 6.15 Constant voltage source

✦ Integrator

The output is fed back through a capacitor as in Fig. 6.16.

$$i_1 = i_F = \frac{v_1}{R_1}$$

$$v_0 = -\frac{1}{C_F}\int i_F\, dt = -\frac{1}{R_1 C_F}\int v_1\, dt$$

Thus, v_0 is a scaled version of the integral of v_1.

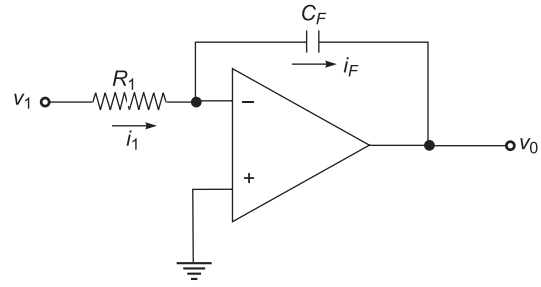


Fig. 6.16 Integrator

The dc offset current and input bias current on continuous integration cause the output to reach saturation limit; well before that the integrator must be recycled. We will not discuss this in detail any further.

✦ Differentiator

The input is through capacitance C_1 as shown in Fig. 6.17.

$$i_1 = i_F$$

$$C_1 \frac{dv_1}{dt} = -\frac{v_0}{R_F}$$

$$v_0 = -R_F C_1 \frac{dv}{dt}\text{; scaled differentiation}$$

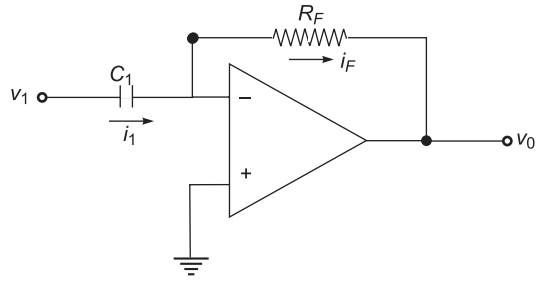


Fig. 6.17 Differentiator

Let $v_1 = \sin \omega t$,

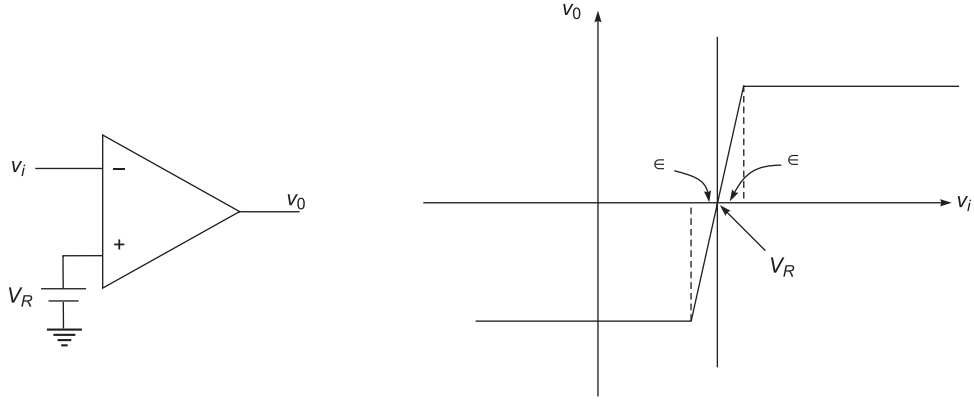
Then

$$v_0 = -(R_F C_1)\, \omega \cos \omega t$$

The output amplitude increases linearly with signal frequency. Therefore, high-frequency noise would get amplified. Differentiation is avoided in op-amp circuits.

❑ ***Comparator*** The saturation characteristic of an op-amp in open loop (see Fig. 6.18) is made use of in determining if a signal is more or less than a certain value (V_R). The op-amp circuit and the output are shown in Fig. 6.18(a) and (b). If the + terminal is grounded ($V_R = 0$), the circuit becomes zero crossing detector.

If v_i is sinusoidal, the waveform of v_0 (reader to make sketch) will be rectangular, assuming $\in = 0$.



(a) (b) Input-output, $\in$ is μV as gain A is very high

Fig. 6.18 Comparator

EXAMPLE 6.1 For the op-amp circuit of Fig. 6.19, find the output and closed-loop gain.

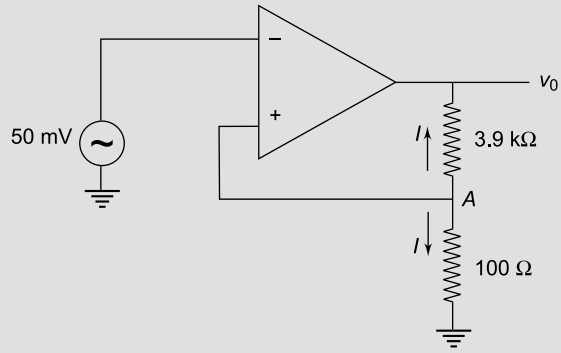


Fig. 6.19

Solution

$$V_A = 50 \text{ mV}$$

$$I(100\ \Omega) = \frac{50}{100} = 0.5 \text{ mA}$$

As no current flows into the + terminal,

$$I(3.9 \text{ k}) = 0.5 \text{ mA}$$

$$V_0 = V_A + 3.9 \times 0.5 \times 10^3 = 50 + 1950$$

$$= 2000 \text{ mV}$$

Gain, $$A_F = \frac{2000}{50} = 40$$

EXAMPLE 6.2 In Fig. 6.20, R_L is the transducer resistance which can vary from 1 kΩ to 10 kΩ. Determine the range of variation of the output voltage.

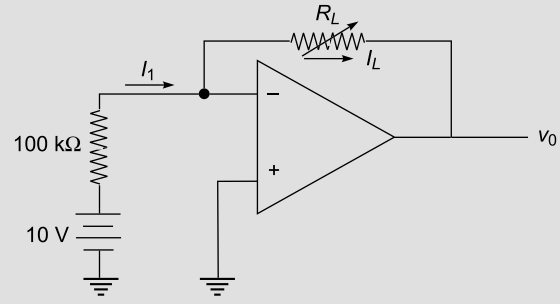


Fig. 6.20

Solution

$$I_1 = I_L$$

$$\frac{10}{100} = -\frac{V_0}{R_L}$$

or $$V_0 = -\frac{R_L}{10}$$

$$\left.\begin{array}{ll} R_L = 1\,\text{k}\Omega & V_0 = -0.1\,\text{V} \\ R_L = 10\,\text{k}\Omega & V_0 = -1\,\text{V} \end{array}\right\} \text{range}$$

EXAMPLE 6.3 For the op-amp circuit of Fig. 6.21, find V_0.

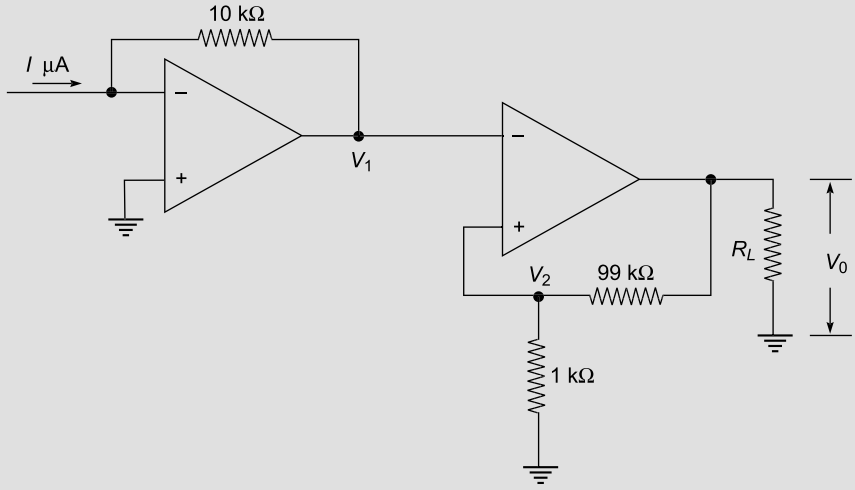


Fig. 6.21

Solution

$$V_1 = -1 \times 10^{-3} \times 10 = -0.1\ \text{V},\ V_2 = V_1$$

Applying KCL at the node (2),

$$\frac{V_1}{1} + \frac{V_1 - V_0}{99} = 0$$

$$V_1\left(\frac{1}{1} + \frac{1}{99}\right) = \frac{V_0}{99}$$

$V_0 = (99 + 1)\ V_1 = 100 \times -0.01 = -1\text{V}$; independent of R_L

EXAMPLE 6.4

For the op-amp circuit of Fig. 6.22, determine v_p, v_n, v_0, and also power supplied/absorbed by the 4 V source.

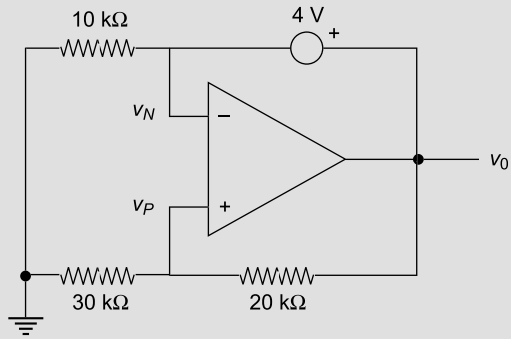


Fig. 6.22

Solution

$$v_p = \left(\frac{30}{30+20}\right) v_0 = 0.6\, v_0$$

$$v_N = v_p = 0.6\, v_0$$

$$v_N + 4 = v_0$$

$$0.6\, v_0 + 4 = v_0$$

or $$v_0 = 10 \text{ V}$$

$$v_p = 6 \text{ V} = v_N$$

$$i\,(10k) = \frac{6}{10} = 0.6 \text{ mA}$$

i(4 V source) = 0.6 mA entering + terminal

Power absorbed, $P = 4 \times 0.6 = 2.4$ mW

EXAMPLE 6.5

For the op-amp circuit of Fig. 6.23, find v_0 if v_s = 9 V. All reisitances are in kΩ.

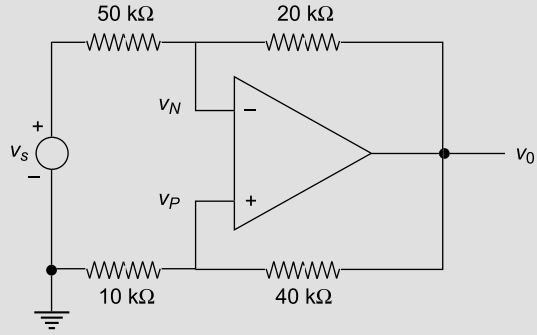


Fig. 6.23

Solution

$$v_p = \left(\frac{10}{10+40}\right) v_0 - 0.2\, v_0$$

$$v_N = 0.2\, v_0$$

Apply KCL at v_N node

$$\frac{v_s - v_N}{50} = \frac{v_N - v_0}{20}$$

$$\frac{9 - 0.2v_0}{50} = \frac{0.2\, v_0 - v_0}{20}$$

Solving, we get

$$v_0 = -5 \text{ V}$$

EXAMPLE 6.6 An op-amp has a slew rate of 0.8 V/μs. What is the maximum amplitude of undistorted sine wave that the op-amp can produce at a frequency of 40 kHz? What is the maximum frequency of the sine wave that op-amp can reproduce if the amplitude is 3 V?

Solution

$$S = 2\pi f V_m$$

(a) $$0.8 \times 10^6 = 2\pi \times 40 \times 10^3\, V_m$$

or $$V_m = 3.18 \text{ V}$$

(b) $$0.8 \times 10^6 = 2\pi f \times 3$$

or $$f = 42.44 \text{ kHz}$$

EXAMPLE 6.7 Figure 6.24 shows the effect of bias current ($I_{B1} = I_{B2} = I_B$) on the output for an inverting op-amp circuit. What value of R_2 will make $V_o = 0$?

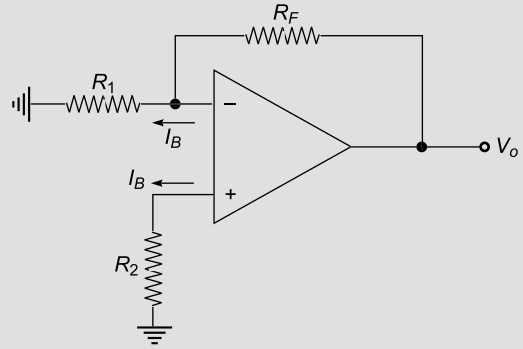


Fig. 6.24

Solution

$$V_+ = R_2 I_B$$

At the –terminal,

$$I_B = \frac{V_+}{R_1} + \frac{V_+ - V_o}{R_F}$$

For $V_o = 0$

$$R_2 = R_1 \parallel R_F$$

6.3 SCHMITT TRIGGER

The Schmitt trigger configuration is shown in Fig. 6.25 where the fraction of the output voltage is fed back to the non-inverting terminal of the comparator is

$$\beta = \frac{R_1}{R_1 + R_2}$$

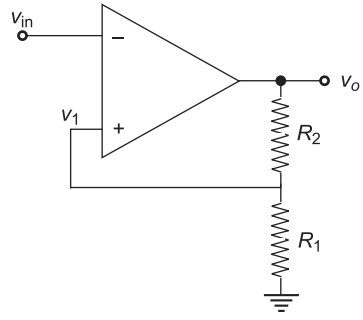


Fig. 6.25 Schmitt-trigger configuration

When the comparator output is $v_o = +V_o$,

$$v_1 = +\beta V_o$$

and the triggering (transition of state) will occur at $v_{in} > +\beta V_o$.

When the comparator output is $v_o = -V_o$.

$$v_1 = -\beta V_o$$

and so the state transition will occur for $v_{in} < -\beta V_o$.

The comparator thus has two trip points: the Upper Trip Point ($UTP = +\beta V_0$) and Lower trip Point ($LTP = -\beta V_o$). The operations around UTP and LTP are indicated in Fig. 6.26(a) and 6.26(b) and the overall operation of the trigger is shown in Fig. 6.26(c), which clearly brings into focus the hysteresis phenomenon wherein the hysteresis voltage V_H is given as

$$V_H = \text{UTP} - \text{LTP} = 2\beta V_o.$$

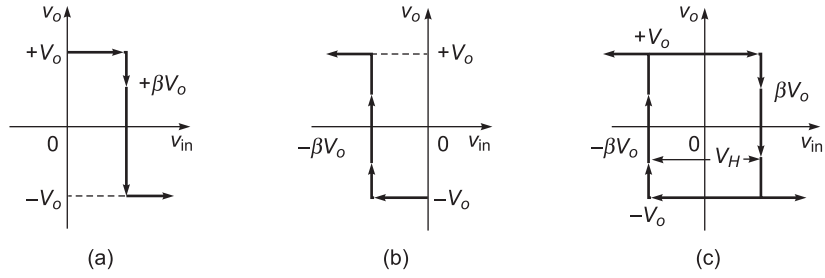


Fig. 6.26 Transfer characteristics of Schmitt trigger

When the input signal is above $+\beta V_o$, the output is $+V_o$. Thus, the Schmitt trigger prevents erroneous operation for any noise level less than V_H. Similar noise immunity is present at input signal less than $-\beta V_o$ with output $-V_o$.

The trip voltages can be shifted by modifying the basic circuit of Fig. 6.25 by providing a fixed bias voltage at the lower end of the feedback circuit (resistive voltage divider) as shown in Fig. 6.27(a).

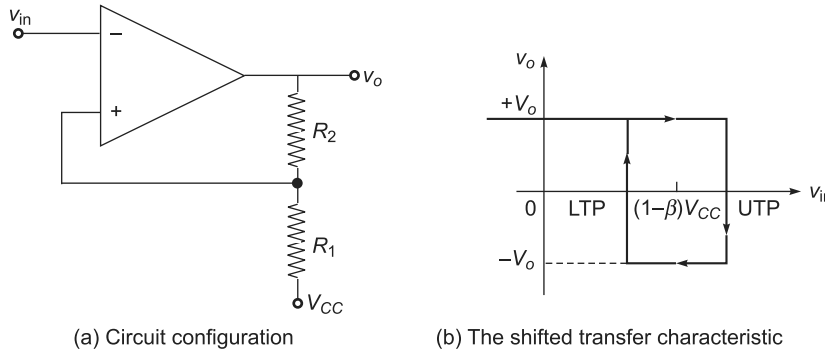


Fig. 6.27 Schmitt trigger with fixed bias voltage

It can be easily shown in this case that

$$\text{UTP} = (1 - \beta)\, V_{CC} + \beta V_o \tag{6.14}$$

$$\text{LTP} = (1 - \beta) V_{CC} - \beta V_o \tag{6.15}$$

with hysteresis voltage, V_H remains the same. The shifted transfer characteristic is drawn in Fig. 6.27(b).

One of the most important uses of a Schmitt trigger is to convert a slowly varying voltage signal into a square waveform with sharp edge transitions. This use is illustrated in Fig. 6.28, wherein the input waveform can be arbitrary except that its amplitude is sufficiently large for the signal to go beyond both the hysteresis limits (range V_H). The output in general would be asymmetric.

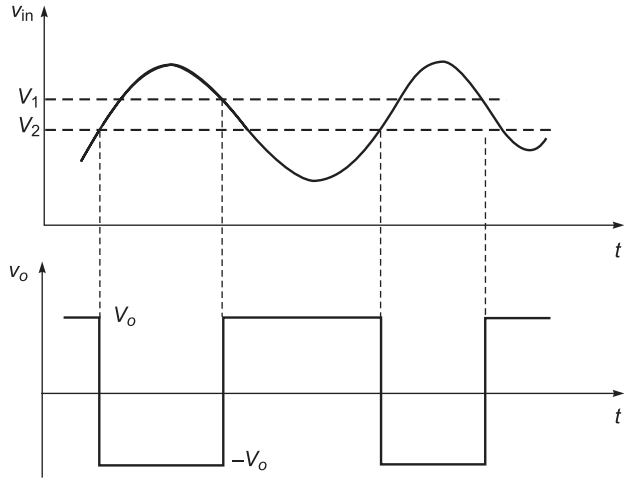


Fig. 6.28 Input and output waveforms of a Schmitt trigger

6.4 CONSTANT-GAIN MULTIPLIER

One of the most common op-amp circuits is the inverting constant-gain multiplier, which provides a precise gain or amplification. Figure 6.29 shows a standard circuit connection, with the resulting gain given by.

$$A = \frac{-R_f}{R_1}$$

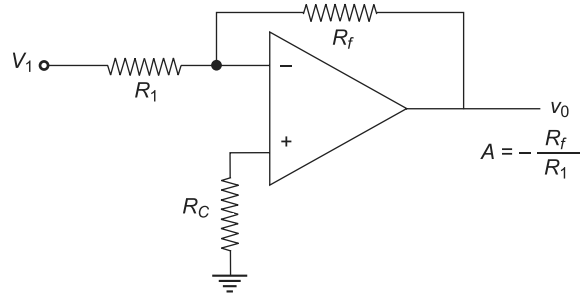


Fig. 6.29 Constant-gain multiplier

6.5 BASIC LOGARITHMIC AMPLIFIER

With the background of diode and current equations, let us study the basic logarithmic amplifier circuit using an op-amp. The fundamental log amplifier is formed by placing a diode or a transistor in the negative feedback path of the op-amp.

6.5.1 Basic Log Amplifier using Diode

The circuit diagram of a basic log amplifier using a diode is shown in Fig. 6.30.

The diode D is used in the negative feedback path. The node A is grounded, hence node B is at virtual ground. Hence, $V_B = 0$.

$$\therefore \qquad I = \frac{V_{\text{in}}}{R} \tag{6.16}$$

As the op-amp input current is zero,

$$I = I_f = \text{diode current} \tag{6.17}$$

Now I_f is the current through the diode and voltage across the diode is $V_B - V_o$, i.e. $-V_o$

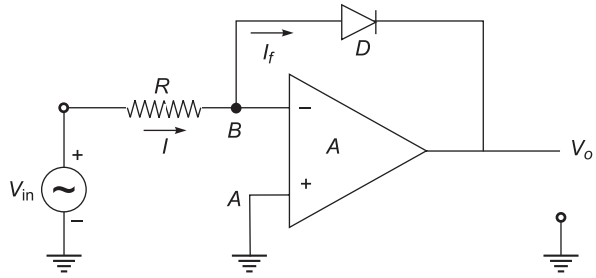


Fig. 6.30 Basic log amplifier

From diode equation,

$$I_f = I_0 \left[e^{V/\eta V_T} - 1 \right] \tag{6.18}$$

we can write the expression for V as,

$$V = \eta V_T \ln \left[\frac{I_f}{I_o} \right]$$

Using this with $V = -V_o$ for the circuit,

$$-V_o = \eta V_T \ln \left[\frac{I_f}{I_o} \right] \tag{6.19}$$

Substituting $I_f = I = \dfrac{V_{\text{in}}}{R}$

$$\therefore \qquad V_o = -\eta V_T \ln \left[\frac{V_{\text{in}}}{R I_o} \right] \tag{6.20}$$

As $I_o R$ is constant dc voltage, let us denote it as V_{ref}.

$$\therefore \qquad \boxed{V_o = -\eta V_T \ln \left[\frac{V_{\text{in}}}{V_{\text{ref}}} \right]} \tag{6.21}$$

where $V_{ref} = I_o R$

Thus, the output voltage V_o is a function of logarithm of the input voltage V_{in}. The circuit gives the logarithm to base, i.e. natural logarithm. But the same circuit can be used to find out logarithm values to base 10 by proper scaling as,

$$\log_{10} X = 0.4343 \ln (X) \tag{6.22}$$

6.5.2 Basic Log Amplifier using Transistor

The basic log amplifier can be obtained by using a transistor as a diode in the negative feedback path of an op-amp, as shown in Fig. 6.31.

The node B is at virtual ground, hence $V_B = 0$.

$$\therefore \quad I = \frac{V_{in} - V_B}{R} = \frac{V_{in}}{R} \tag{6.23}$$

As the op-amp input current is zero,

$$I = I_C = \text{collector current} \tag{6.24}$$

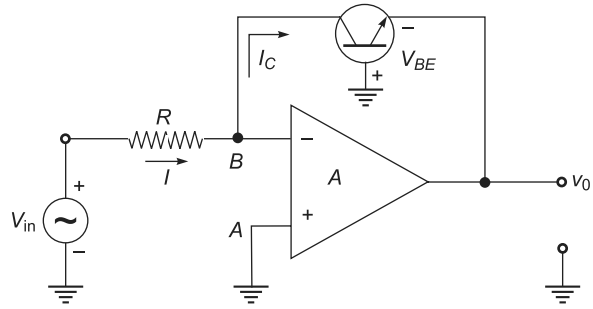


Fig. 6.31 Basic log amplifier

The voltage $V_{CB} = 0$ as the collector is at virtual ground and base is grounded. Hence, we can write the equation of I_C as,

$$V_{BE} = V_T \ln \left(\frac{I_C}{I_s} \right) \tag{6.25}$$

Applying to the output side, we get,

$$V_o + V_{BE} = 0 \tag{6.26}$$

$$\therefore \quad V_{BE} = -V_o \tag{6.27}$$

and $$I_C = I = \frac{V_{in}}{R}$$

Substituting in Eq. (6.25),

$$-V_o = V_T \ln \left(\frac{V_{in}}{R I_s} \right) \tag{6.28}$$

Let $V_{ref} = R I_s$,

$$\therefore \qquad \boxed{V_o = -V_T \ln\left[\frac{V_{in}}{V_{ref}}\right]} \qquad (6.29)$$

The equation is similar to Eq. (6.21), which gives the output, proportional to the logarithm of the input voltage V_{in}.

SUMMARY

➢ This chapter has described OP-AMP, one of the most important and versatile analog IC, various types of OP-AMPs with their characteristics are discussed.

EXERCISES

➥ Review Questions

1. Describe an op-amp; and the various stages of an op-amp circuit.
2. Draw a simple current equivalent of op-amp. Derive from this an ideal op-amp. Justify idealisation op-amp in current analysis.
3. What is CMRR? Why does it have a high value?
4. Sketch the frequency response of an op-amp open-loop mode.
5. What is meant by gain-bandwidth product?
6. How does negative feedback around an op-amp affect the bandwidth?
7. What is meant by slew rate of op-amp? How does it affect its performances?
8. What is meant by virtual ground?
9. State some of the applications of op-amp.
10. Sketch the transfer characteristic of op-amp.
11. How is an op-amp used as a comparator? How is it used as a zero crossing detector?

➥ Problems

1. In the op-amp circuit of Fig. 6.32, show that
   $v_0 = (v_1 + v_2) - (v_2 + v_4)$
   If all resistances are equal, is it a subtractor circuit?

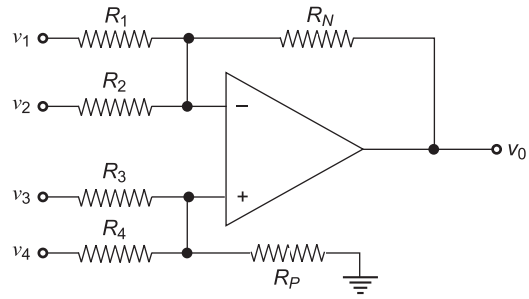


Fig. 6.32

2. For the op-amp circuit of Fig. 6.33, determine V_0.

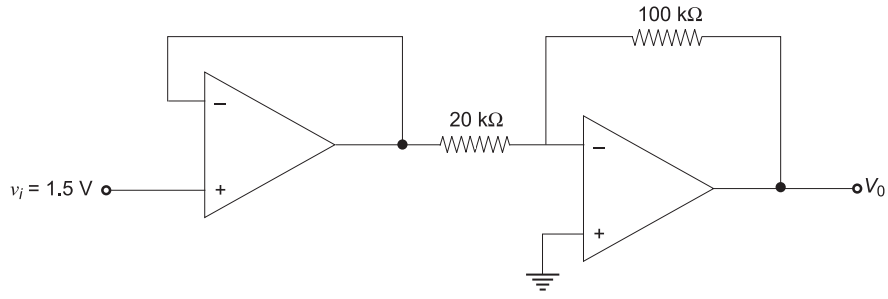


Fig. 6.33

3. For the op-amp circuit of Fig. 6.34, find V_0.

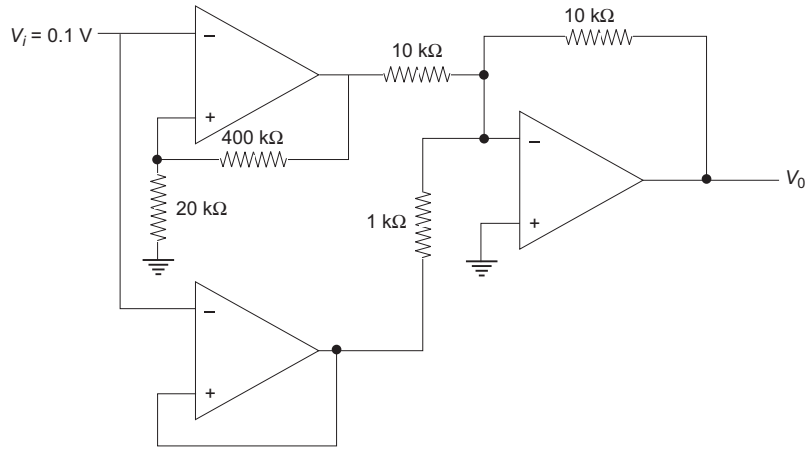


Fig. 6.34

4. In the op-amp circuit of Fig. 6.35, the gain A is finite but $R_{in} = \infty$ and $R_0 = 0$. Calculate the gain.

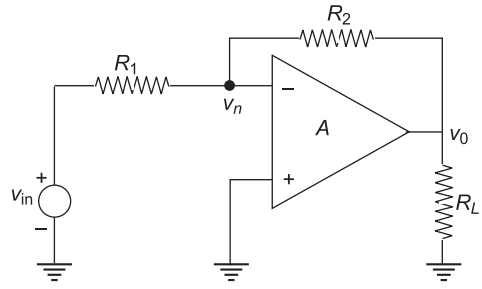


Fig. 6.35

5. In Problem 3, the bandwidth of the op-amp is f_h. Determine the bandwidth of the feedback amplifier.

6. For the op-amp circuit of Fig. 6.36, determine the values V_o.
7. For the op-amp circuit Fig. 6.37, find the value of V_o.

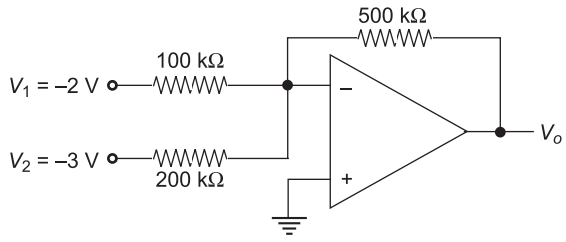


Fig. 6.36

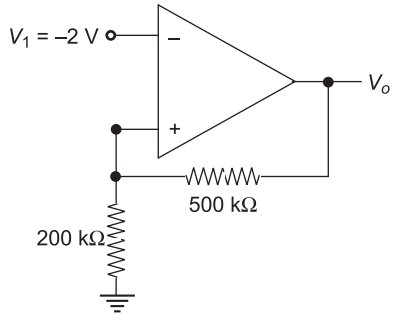


Fig. 6.37

8. For the instrumentation amplifier shown in Fig. 6.38, verify that

$$V_o = \left(1 + \frac{R_2}{R_1} + \frac{2R_2}{R}\right)(V_2 - V_1)$$

Can the gain can be adjusted by varying R?

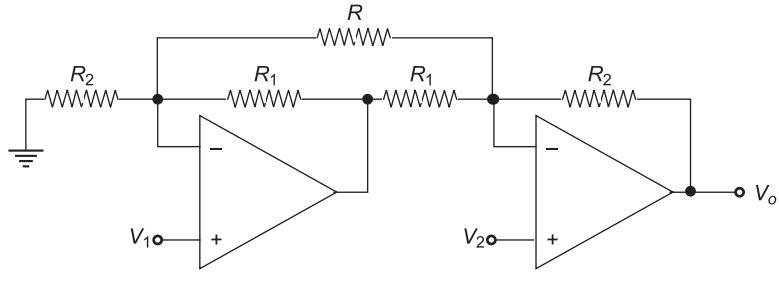
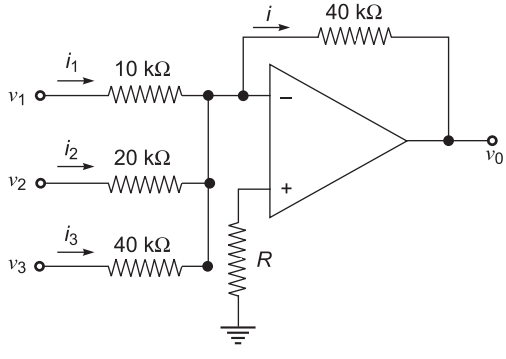


Fig. 6.38

9. For the op-amp circuit of Fig. 6.39, sketch waveforms i_1, i_2, i_3 for the given waveforms of v_1, v_2 and v_3. Hence, sketch the waveform v_o.



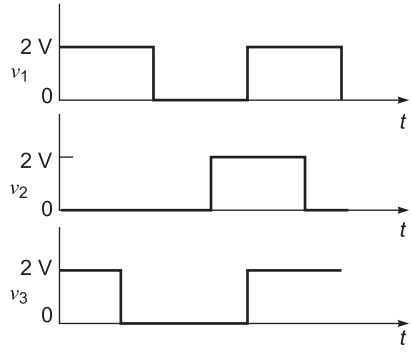


Fig. 6.39

10. For the op-amp circuit of Fig. 6.40, find $v_o(t)$. The capacitor initially uncharged.

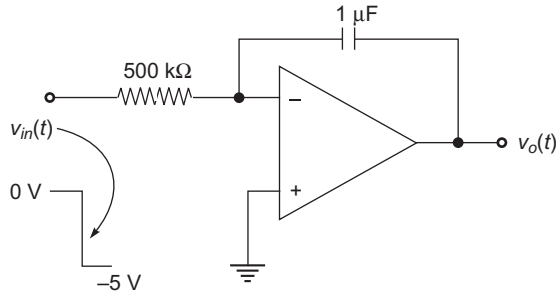


Fig. 6.40

11. CMRR of differential amplifier given in Fig. 6.41 is

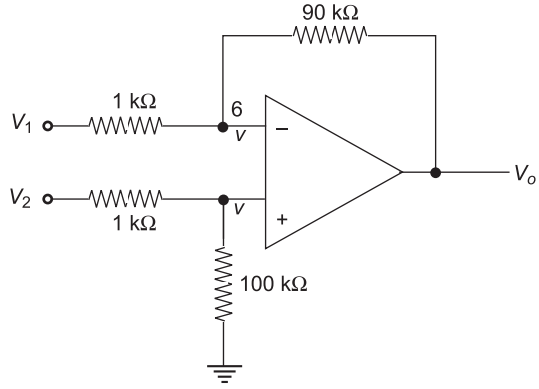


Fig. 6.41

➥ Multiple-Choice Questions

1. An ideal op-amp has
(a) zero R_o (b) infinite R_i (c) infinite Av (d) all of these

2. An op-amp has
(a) very small input resistance and very large output resistance
(b) very small input resistance and very small output resistance
(c) very large input resistance and very small output resistance
(d) very large input resistance and very large output resistance

3. The two input terminals of an op-amp are known as
(a) positive and negative (b) differential and non-differential
(c) inverting and non-inverting (d) high and low

4. The voltage gain of an open-circuit ideal op-amp is
(a) infinity (b) around 300 (c) 0 (d) around 10,000

5. The input impedance of an ideal op-amp is
(a) 0 (b) 10 kΩ (c) infinity (d) none

6. A circuit (Fig. 6.42) using an op-amp has
(a) current-series feedback
(b) current-shunt feedback
(c) voltage-shunt feedback
(d) voltage-series feedback

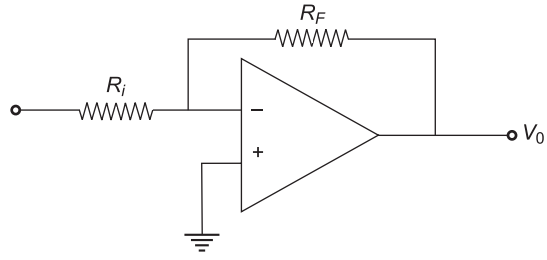


Fig. 6.42

7. An op-amp has a common mode gain of 0.01 and a differential mode gain of 10^5. Its common mode rejection ratio would be
(a) 10^3 (b) 10^5
(c) 10^7 (d) 10^{-7}

8. A differential amplifier is used in the input stage of all op-amps to provide the op-amp with a very high
(a) slew rate (b) bandwidth (c) open-loop gain (d) CMRR

9. The input stage of an op-amp is usually a
(a) level shifter (b) *CE* amplifier
(c) Class-B push-pull amplifier (d) differential amplifier

10. In a circuit if the open loop gain is 10^6 and output voltage is 10 V, the differential voltage will be
(a) 10 μV (b) 100 μV (c) 0.1 V (d) 1 μV

11. For CMRR = 65 dB, determine the common mode output voltage in the given below circuit.

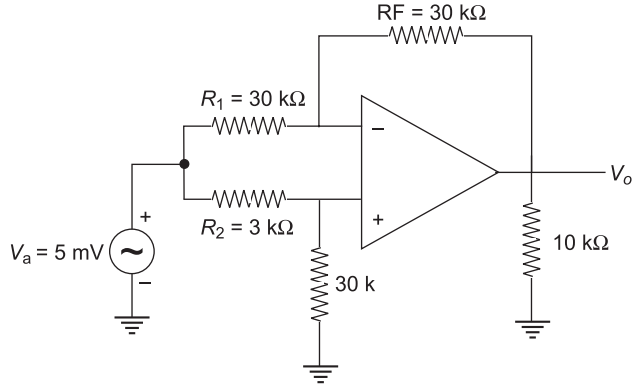


Fig. 6.43

(a) 281.17 mV (b) 28.11 mV (c) 2.81 mV (d) 0.281 mV

ANSWERS

✦ Problems

2. $V_0 = -7.5$ **3.** $V_0 = 11$ V **4.** $A_F = -\dfrac{R_2}{R_1 + (R_1 + R_2)\,\Delta}$ **5.** $\left[(A+1)\dfrac{R_1}{R_2} + 1\right] f_h$

6. 2.5 V **7.** –7 V **10.** $10t \;\; u(t)$ **11.** 1000

✦ Multiple-Choice Questions

1. (d) **2.** (c) **3.** (c) **4.** (a) **5.** (c) **6.** (d) **7.** (c) **8.** (c) **9.** (d) **10.** (a)
11. (a)

CHAPTER 7

Feedback Amplifiers and Oscillators

GOALS AND OBJECTIVES

- Introduction and discussion about types of feedback-voltage series, voltage shunt, current series and current shunt
- To know the importance of feedback in amplifiers
- Description of oscillators—Wein bridge oscillator, tuned oscillator and crystal oscillator

7.1 INTRODUCTION

In a feedback amplifier, voltage or current output is fed back to the input through a modifying network, which determines the magnitude and phase. The feedback centre opposes the input (*negative feedback*) or aids the input (*positive feedback*). The feedback can change certain important characteristics of the amplifier in a desirable manner. The purpose of an amplifier is to amplify

the signal without changing its characteristics except its amplitude. The amplifier that works on the principal of feedback is known as feedback amplifier.

Feedback is a process where a fraction of output (voltage/current) is fed back to the input. Then there become two signals at the input, i.e. the original input signal and fed back signal. Basic feedback is of two types depending upon how and in what phase the input signal and fraction of output are mixed.

When the input signal and the returned signal (fed back signal) are in same phase, the signals are then added up and the resultant output increases, this is called positive or regenerative or direct feedback. Positive feedback causes distortion and instability in amplifiers and hence it is not used for amplifiers; whereas positive feedback amplifier increases the gain and overall power of input signal and hence used in oscillator circuits. The following diagram (Fig. 7.1) shows a positive feedback amplifier.

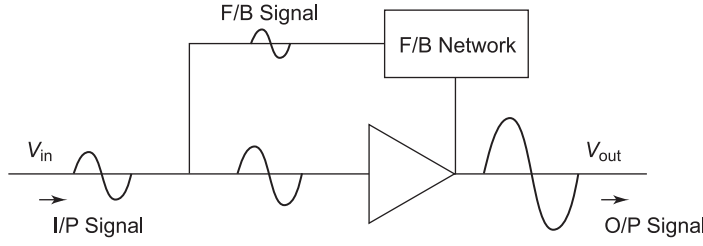


Fig. 7.1 Positive feedback amplifier

In this, voltage gain with positive feedback is given by $A_f = \frac{A}{(1-\beta A)}$, where β is the feedback factor.

When the returned signal has phase opposite to the input signal, the net input signal is a difference of input and feedback signals, and this is called negative or inverse or degenerative feedback. This feedback induces desirable modifications in circuit performance. Though negative feedback reduces the overall gain of the amplifier but it has numerous advantages and hence widely used in amplifier circuits.

The diagram given in Fig. 7.2 shows a negative feedback amplifier.

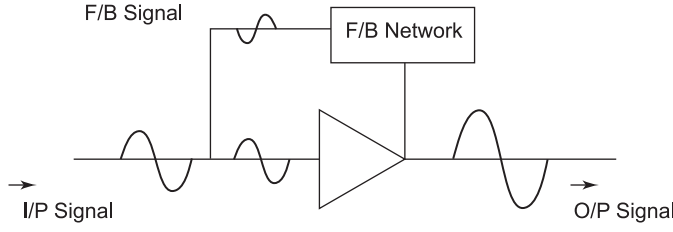


Fig. 7.2 Negative feedback amplifier

In this, voltage gain with negative feedback is given by $A_f = \frac{A}{(1+\beta A)}$, where β is the feedback factor.

7.1.1 Properties of Negative Feedback Amplifier

- ***Desensitize the gain***: It brings stability to amplifier by making gain less sensitive to all kind of variations.
- ***Reduce nonlinear distortion***: The negative feedback makes the output proportional to the input, i.e. reduces non-linear distortion.
- ***Reduce the effect of noise***: It minimizes the contribution by unwanted electric signals . This noise may be generated by circuit components or by extraneous interference.
- ***Control the input and output impedances***: It increases or decreases the input and output impedances. This is done by choosing appropriate feedback topology.
- ***Extend the bandwidth of the amplifier***: By incorporating negative feedback, the bandwidth can be increased.

7.1.2 Advantages of Negative Feedback

In negative feedback amplifier, the gain of the amplifier reduces, however it is still used in almost every amplifier due to its various advantages. Some of its advantages are given below:

- Gains stability
- Significant extension of bandwidth
- Very less distortions
- Decreased output resistance
- Stable operating point
- Reduces noise and other interference in amplifier

All advantages are on cost of reduced gain of amplifier, and hence in negative feedback there is always a trade-off between amplifier gain and other desirable properties.

7.2 TYPES OF FEEDBACK

There are four types of feedback: (1) Voltage series (2) Voltage shunt (3) Current series (4) Current shunt

We will consider *voltage series* feedback whose circuit is drawn in Fig. 7.3. It is also known as series-parallel feedback. Observe that the *feedback is negative*.

From the figure,

$$V_i = V_{\text{in}} - \beta V_o$$

$$V_o = AV_i = AV_{\text{in}} - A\beta V_o$$

Reorganising, we get gain with feedback as

$$\frac{V_o}{V_{\text{in}}} = A_F = \frac{A}{1 + \beta A} \tag{7.1}$$

This is a general result, which applies to all types of feedback circuits.

The amplifier gain reduces by a factor of $(1 + \beta A)$ (7.2)

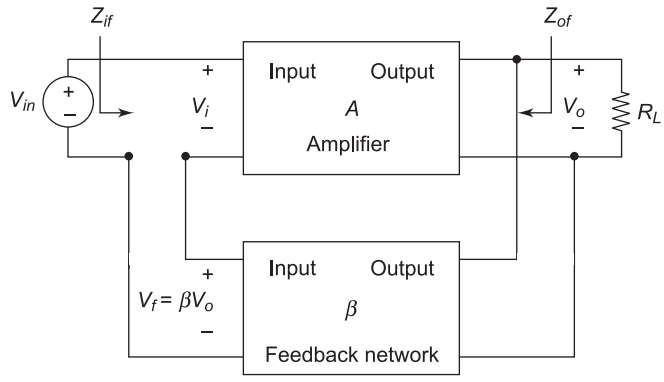


Fig. 7.3 Voltage-series feedback

It can be shown that

Input impedance with feedback, $Z_{if} = Z_i(1 + \beta A)$, increases

Output impedance with feedback, $Z_{of} = Z_o/(1 + \beta A)$, decreases (7.3)

Other types of feedbacks can be analysed on similar lines.

If $\beta A >> 1$ then Eq. (7.1) yields the feedback amplifier gain as

$$A_F = \frac{1}{\beta} \tag{7.4}$$

This means the feedback gain is independent of amplifier gain A. Thus, all the distortions (like amplitude and frequency distortion) do not appear in A_F. This also happens to a noise signal, which gets attenuated by feedback. Any variation in magnitude of A does not appear in A_F, which means A_F has high gain stability. As the cost of these improvements is paid in terms of reduced gain, which can be made up.

Gain and Bandwidth of Feedback Amplifier

As shown above, the negative feedback reduces the amplifier gain. Therefore, as per the general principle, it should increase its bandwidth. In *RC*-coupled amplifiers, the gain reduces at low-frequency and high-frequency ends. So βA_o is no longer much more than unity. As a result, the percent reduction in gain is less at the two frequency ends compared to the mid-band. The reduction in gain and increase in bandwidth of feedback amplifiers are illustrated in Fig. 7.4.

As $f_1 << f_2$ and $f_{1f} << f_{2f}$, therefore,

$\text{BW} \approx f_2$

$\text{BW}_f \approx f_{2f}$

It can be shown that

$$A_o f_2 = A_{of} f_{2f} = \textit{constant product of gain bandwidth} \tag{7.5}$$

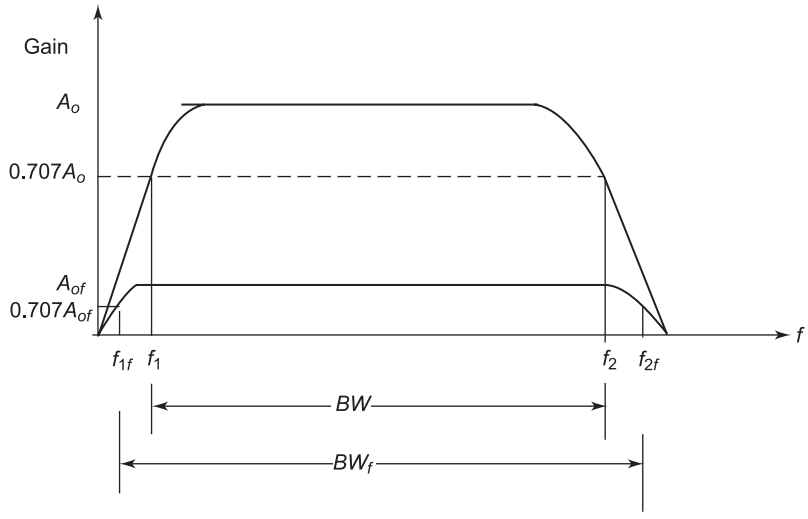


Fig. 7.4 Effect on A_o and bandwidth of negative feedback

EXAMPLE 7.1 An amplifier has a high-frequency response described as

$$A = \frac{A_o}{1+(j\omega/\omega_2)}$$

wherein $A_o = 1000$, $\omega_2 = 10^4$ rad/s.

Find the feedback (negative) factor β, which will raise the upper corner frequency (ω_2) to 10^5 rad/s. What is the corresponding overall gain of the amplifier? Find also the gain–bandwidth product in each case.

Solution

$$A_f = \frac{A_o}{1+\{j\omega/[\omega_2(1+\beta A_o)]\}} = \frac{A(\text{new})}{1+[j\omega/\omega_2(\text{new})]}$$

where,
$$A(\text{new}) = \frac{A_o}{1+\beta A_o}$$

$$\omega_2(\text{new}) = \omega_2(1+\beta A_o)$$

Substituting values, $10^5 = 10^4(1+\beta\times 1000)$

or $\beta = 0.009$

$$A(\text{new}) = \frac{A_o}{1+\beta A_o} = 100$$

Gain-bandwidth products, without and with feedback are

$$\omega_2 A_o = 10^4 \times 10^3 = 10^7$$

$$\omega_2(\text{new})\, A(\text{new}) = 10^5 \times 100 = 10^7$$

Observe that the gain-bandwidth product is maintained constant.

7.3 GAIN STABILITY WITH FEEDBACK

We have seen that the overall gain with negative feedback is

$$A_f = \frac{A}{1+\beta A}$$

Differentiation of the above equation leads to

$$\frac{dA_f}{A_f} = \frac{1}{(1+\beta A)}\left(\frac{dA}{A}\right) \tag{7.6}$$

$$\frac{dA_f}{A_f} \approx \frac{1}{\beta A}\left(\frac{dA}{A}\right), \text{ for } \beta A >> 1 \tag{7.7}$$

This shows that a relative change (dA/A) in the basic amplifier gain is reduced by the factor βA in the relative change (dA_f/A_f) in the overall gain of the feedback amplifier.

7.4 LARGE-SIGNAL AMPLIFIERS

An amplifying system consisting of several stages in cascade has the input stage and the intermediate stage as small-signal. The function of the small-signal stages is to amplify the signal to a large value enough to drive the final stage. The final stage is a large-signal amplifier stage and typically feeds a transducer. In a large-signal amplifier, the output voltage and current swings are so large that we cannot replace the amplifying device while designing a large-signal amplifier, harmonic distortion should not be excessive. The study of large-signal amplifiers deals with calculation of distortion, power output and efficiency.

In single ended amplifiers, i.e., amplifiers using single transistor, appreciable distortion may result due to non-linearity of transfer characteristics. This distortion may be reduced considerably by Push-Pull operation of transistors.

7.5 THE CLASS B AMPLIFIER

To improve the full power efficiency of the previous Class A amplifier by reducing the wasted power in the form of heat, it is possible to design the power amplifier circuit with two transistors in its output stage producing what is commonly termed as a **Class B amplifier** also known as a **push-pull amplifier** configuration.

7.5.1 Push-Pull Amplifier

The Push-Pull amplifier is a power amplifier. It is frequently found in output stages of electronic circuits. It is used whenever high output power at high efficiency and little distortion is required.

The base voltages of the two transistors must be equal in magnitude and in phase opposition with respect to ground. So, the input signal is applied to push-pull amplifier through a transformer. Input transformer with centre tapped secondary winding easily provides such equal and opposite voltages or else we may use any other circuit to generate these signals.

The actual load resistor R_L is connected to the transistors by means of a transformer with centre tapped primary winding. The supply voltage is connected between the emitter terminals and this centre tap.

Push-pull amplifiers use two "complementary" or matching transistors, one being an *NPN*-type and the other being a *PNP*-type with both power transistors receiving the same input signal together that is equal in magnitude, but in opposite phase to each other. This results in one transistor only amplifying one half or 180° of the input waveform cycle while the other transistor amplifies the other half or remaining 180° of the input waveform cycle with the resulting "two halves" being put back together again at the output terminal.

The conduction angle for this type of amplifier circuit is only 180° or 50% of the input signal. This pushing and pulling effect of the alternating half cycles by the transistors gives this type of circuit its amusing "push-pull" name, but more generally known as the **Class B amplifier** shown in Fig. 7.5.

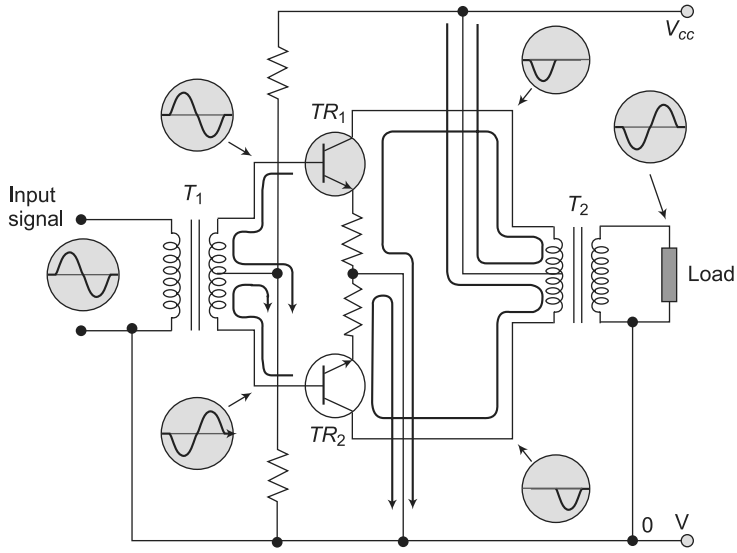


Fig. 7.5 Class B push-pull transformer amplifier circuit

The circuit above shows a standard **Class B amplifier** circuit that uses a balanced centre-tapped input transformer, which splits the incoming waveform signal into two equal halves and which are 180° out of phase with each other. Another centre-tapped transformer on the output is used to recombine the two signals providing the increased power to the load. The transistors used for this type of transformer push-pull amplifier circuit are both *NPN* transistors with their emitter terminals connected together.

Here, the load current is shared between the two power transistor devices as it decreases in one device and increases in the other throughout the signal cycle reducing the output voltage and current to zero. The result is that both halves of the output waveform now swing from zero to twice the quiescent current, thereby reducing dissipation. This has the effect of almost doubling the efficiency of the amplifier to around 70%.

Assuming that no input signal is present, each transistor carries the normal quiescent collector current, the value of which is determined by the base bias which is at the cut-off point. If the transformer is accurately centre tapped then the two collector currents will flow in opposite directions (ideal condition) and there will be no magnetisation of the transformer core, thus minimising the possibility of distortion.

When an input signal is present across the secondary of the driver transformer T_1, the transistor base inputs are in "anti-phase" to each other as shown. Thus, if TR_1 base goes positive, driving the transistor into heavy conduction, its collector current will increase but at the same time the base current of TR_2 will go negative further into cut-off and the collector current of this transistor decreases by an equal amount and vice versa. Hence, negative halves are amplified by one transistor and positive halves by the other transistor, giving this push-pull effect.

Unlike dc condition, these ac currents are additive resulting in the two output half-cycles being combined to reform the sine wave in the output transformer's primary winding that appears across the load.

Class B amplifier operation has zero dc bias as the transistors are biased at the cut-off, so each transistor only conducts when the input signal is greater than the base-emitter voltage. Therefore, at zero input, there is zero output and no power is being consumed. It means the actual Q-point of a Class B amplifier is on the V_{ce} part of the load line as shown in Fig. 7.6.

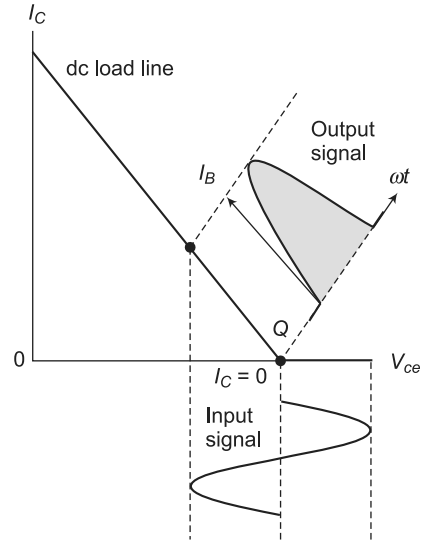


Fig. 7.6 Class B output characteristics curves

The **Class B amplifier** has the big advantage over their Class A amplifier cousins in that no current flows through the transistors when they are in their quiescent state (i.e. with no input signal). Therefore, no power is dissipated in the output transistors or transformer when there is no signal present unlike Class A amplifier stages that require significant base bias, thereby dissipating lots of heat—even with no input signal present. Consequently the overall conversion efficiency (η) of the amplifier is greater than that of the equivalent Class A with efficiencies reaching as high as 70% possible resulting in nearly all modern types of push-pull amplifiers operated in this Class B mode.

7.5.2 Transformerless Class B Push-Pull Amplifier

One of the main disadvantages of the Class B amplifier circuit above is that it uses balanced centre-tapped transformers in its design, making it expensive to construct. However, there

is another type of Class B amplifier called a **complementary-symmetry Class B amplifier** that does not use transformers in its design. Therefore, it is transformerless, using instead complementary or matching pairs of power transistors. As transformers are not needed, this makes the amplifier circuit much smaller for the same amount of output. Also, there are no stray magnetic effects or transformer distortion to effect the quality of the output signal. An example of a "transformerless" Class B amplifier circuit is given below.

The Class B amplifier circuit uses complementary transistors for each half of the waveform.

Although, Class B amplifiers have a much high gain than the Class A types, but one of the main disadvantages of Class B type push-pull amplifiers is that they suffer from an effect commonly known as **crossover distortion**.

Hopefully, we remember from our tutorials about transistors that it takes approximately 0.7 volt (measured from base to emitter) to get a bipolar transistor to start conducting. In a pure Class B amplifier, the output transistors are not "pre-biased" to an "ON" state of operation.

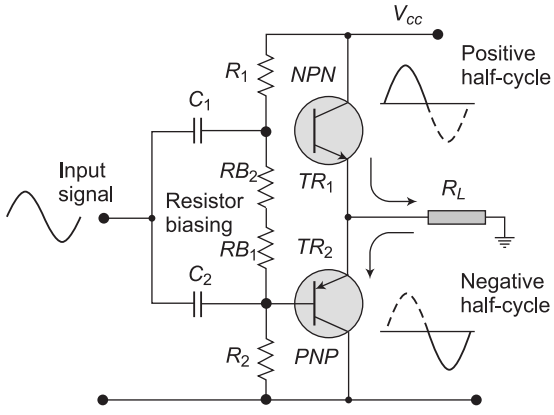


Fig. 7.7 Class B transformerless output stage

This means that the part of the output waveform which falls below this 0.7 volt window will not be reproduced accurately as during the transition between the two transistors (when they are switching over from one transistor to the other), the transistors do not stop or start conducting exactly at the zero crossover point even if they are specially matched pairs. The output transistors for each half of the waveform (positive and negative) will each have a 0.7 volt area in which they are not conducting. The result is that both transistors are turned "OFF" at exactly the same time.

A simple way to eliminate crossover distortion in a Class B amplifier is to add two small voltage sources to the circuit to bias both the transistors at a point slightly above their cut-off point. This would give us what is commonly called a **Class AB amplifier** circuit. However, it is impractical to add additional voltage sources to the amplifier circuit, so *pn*-junctions are used to provide the additional bias in the form of silicon diodes.

7.6 | THE CLASS AB AMPLIFIER

We know that we need the base-emitter voltage to be greater than 0.7 V for a silicon bipolar transistor to start conducting. So if we were to replace the two voltage divider biasing resistors connected to the base terminals of the transistors with two silicon diodes, the biasing voltage applied to the transistors would now be equal to the forward voltage drop of the diode. These two diodes are generally called **biasing diodes** or **compensating diodes** and are chosen to match the characteristics of the matching transistors. The fig. 7.8 below shows diode biasing.

The **Class AB amplifier** circuit is a compromise between the Class A and the Class B configurations. This very small diode-biasing voltage causes both transistors to slightly conduct even when no input signal is present. An input signal waveform will cause the transistors to operate as normal in their active region, thereby eliminating any crossover distortion present in pure Class B amplifier designs.

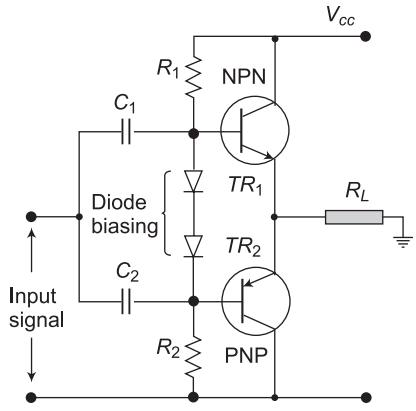


Fig. 7.8 Class AB amplifier

A small collector current will flow when there is no input signal but it is much less than that for the Class A amplifier configuration. It means, the transistor will be "ON" for more than half a cycle of the waveform but much less than a full cycle giving a conduction angle of between 180° to 360° or 50% to 100% of the input signal depending upon the amount of additional biasing used. The amount of diode-biasing voltage present at the base terminal of the transistor can be increased in multiples by adding additional diodes in series.

Class B amplifiers are greatly preferred over Class A designs for high-power applications such as audio power amplifiers and PA systems. Like the Class A amplifier circuit, one way to greatly boost the current gain (A_i) of a Class B push-pull amplifier is to use Darlington transistor pairs instead of single transistors in its output circuitry.

EXAMPLE 7.2 A feedback amplifier comprises two amplifying blocks in tandem; each block having a gain of 100. What should be the gain of the feedback block in order for overall gain to be 100? If the gain of each amplifier block reduces to 50% due to parameter variations, what is the percentage change in the gain of the complete feedback unit?

Solution $A = 100$ (given); forward gain $= A^2 = 10^4$

$$A_f = \frac{10^4}{1 + \beta \times 10^4};\ A_f = 100 \text{ (given)}$$

Therefore, $\beta = 0.0099$

New value of $A = 50$, $A_f(\text{new}) = \dfrac{(50)^2}{(1 + 0.0099 \times 50^2)} = 97.09$

Reduction in overall gain = 100 – 97.09 = 2.91 or 2.91%

❑ ***Observation*** It is seen that a 50% reduction in forward gain causes only 2.91% reduction in gain of feedback amplifier. This means that sensitivity of feedback amplifier gain to change in component (forward) gain is quite low. This is one (important) reason for using feedback technique in amplifiers.

We will not pursue the feedback amplifiers any further but go ahead and discover how feedback can use to set up an oscillatory circuit.

7.7 | OSCILLATORS

✦ Condition for Oscillation

Consider the schematic of the negative feedback amplifier of Fig. 7.9 without any input. As we shall be dealing with sinusoidal quantities voltages, currents would be treated as phasors and gains complex numbers. For the feedback amplifier of Fig. 7.9.

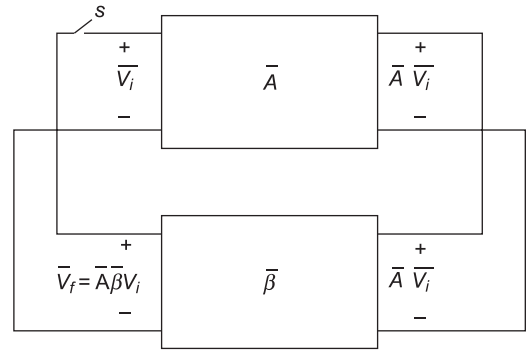


Fig. 7.9 Pertaining to condition for oscillation

Amplifier gain, $\dfrac{\bar{V}_o}{\bar{V}_{\text{in}}} = \bar{A}_F = \dfrac{\bar{A}}{1 + \bar{\beta}\bar{A}}$

If $\bar{\beta}$ is adjusted such that

$$\bar{\beta}\bar{A} = -1 = 1\angle -180° \tag{7.8}$$

the gain tends to become infinity. For $|\bar{\beta}\bar{A}|$ slightly more than unity, the circuit becomes self-oscillatory with no input $V_{\text{in}} = 0$.

The condition

$$\bar{\beta}\bar{A} = -1 \tag{7.9}$$

is known as *Barkhausen condition for oscillation*.

Let us consider the amplifier with no input shown in Fig. 7.9. If $\angle\bar{\beta}\bar{A} = -180°$, the polarity of $\bar{V}_f$ reverses and the feedback becomes positive. As the switch S is closed, a small voltage (may be due to noise) will begin to build up. The oscillation increases in amplitude till steady conditions are reached with non-sinusoidal oscillations due to saturation as shown in Fig. 7.10.

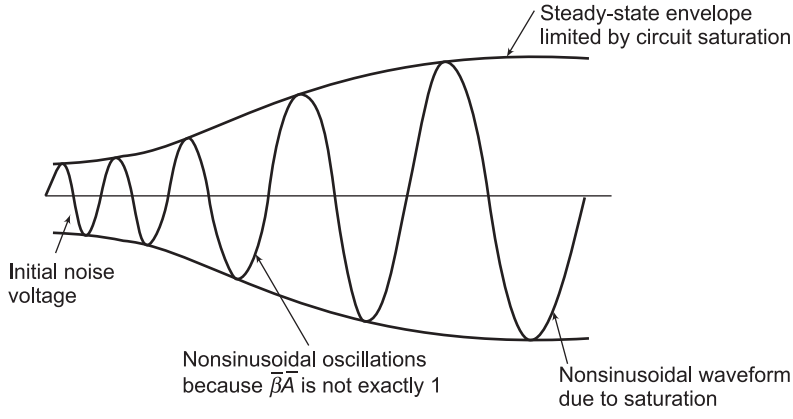


Fig. 7.10 Oscillation build-up with phase reversal of feedback voltage

In practical oscillator circuits, the amplifier gain will be a negative constant at oscillation frequency. For oscillations to occur, $\angle \bar{A}\bar{\beta} = -180°$ and $|\bar{A}\bar{\beta}| > 1$. We will now study various types of oscillators depending upon the circuits to achieve $\angle \bar{A}\bar{\beta} = -180°$. For an amplifier, we shall use op-amp in inverting mode. For BJT based oscillators, the gain can be formed by the method studied in amplifiers.

Introduction to Oscillators

The process of raising the strength of a weak signal without any change in its shape is known as faithful amplification. The dc signal applied to the amplifier is amplified in accordance with the instantaneous value of input ac signal. Nevertheless an oscillator is an energy converter. It receives dc energy and changes it into ac energy of derived frequency. The frequency of oscillations depends upon the constants of the device. The composite electric circuit associated with an active device when used to produce an alternating current is called an oscillator circuit. Oscillators can be both sinusoidal and non-sinusoidal. Non-sinusoidal oscillators are also called Relaxation oscillators which are rich in harmonics.

In this unit, the conditions required for an amplifier to work as an oscillator and the design of different oscillator circuits is discussed. Different oscillator circuits like RC phase shift and Wein bridge oscillators are dealt in an elaborate manner. The conditions that are to be satisfied by these circuits for sustained oscillations are explained. The equations for the output frequencies of different oscillator circuits are derived. The crystal oscillator's construction, working and its advantages over other oscillator types are discussed.

Classification of Oscillators

(1) Based on operating principle
   (a) Negative resistance effect oscillators.
   (b) Feedback oscillators.
(2) Based on waveforms
   (a) Sinusoidal oscillations
   (b) Relaxation oscillations
(3) Based on frequency generation
   (a) AF oscillators
   (b) RF oscillators
   (c) UHF oscillators.
   (d) Microwave oscillators.
(4) According to the circuit employed.
   (a) LC oscillators
   (b) RC oscillators

7.8 PHASE-SHIFT OSCILLATOR

The phase shift is achieved by *RC*-network. Because of loading effect, three *RC*-stages are needed as shown in Fig. 7.11. By writing three nodal equations in phasor form, we can express

$$\bar{\beta} = \frac{\bar{V}_1(j\omega)}{\bar{V}_0(j\omega)}$$

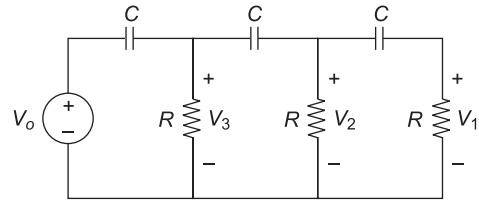


Fig. 7.11 RC phase shifting network

By equating the *j*-part of it, in the denominator, we find the frequency at which β is negative (180° phase shifts). The results we get are,

Frequency of oscillation

$$\omega_o = \frac{1}{RC\sqrt{6}} \tag{7.10}$$

And $$\beta(\omega_o) = -\frac{1}{29}\text{; 180° phase shift} \tag{7.11}$$

For oscillations to occur,

$$\left|\bar{\beta}\right| > \frac{1}{29} \tag{7.12}$$

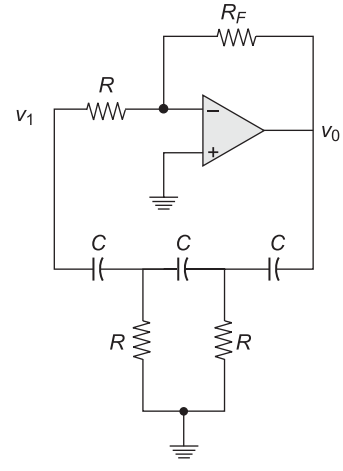


Fig. 7.12 Phase-shift oscillator

The phase-shift oscillator with op-amp is drawn in Fig. 7.12 wherein

Frequency of oscillation,

$$\omega_o = \frac{1}{RC\sqrt{6}} \tag{7.13}$$

$$A = -\frac{R_F}{R}$$

$$A\bar{\beta} = -\frac{1}{29}\left(-\frac{R_F}{R}\right)$$

$$= \frac{R_F}{29R} > 1\text{; by about 5\%}$$

$$R_F > 29R \tag{7.14}$$

EXAMPLE 7.3 For the feedback op-amp circuit of Fig. 7.13, determine the condition for oscillation and the oscillation frequency.

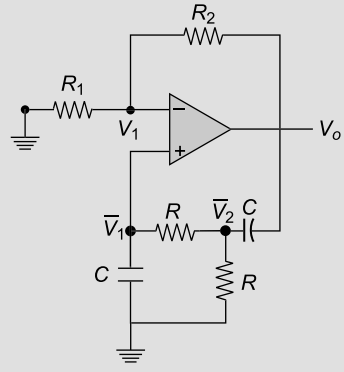


Fig. 7.13

Solution

At the node 2,

$$j\omega C\ (\bar{V}_2 - \bar{V}_o) + \frac{\bar{V}_2}{R} + \frac{\bar{V}_2 - \bar{V}_1}{R} = 0 \tag{i}$$

At the node 1,

$$\frac{\bar{V}_1 - \bar{V}_2}{R} + j\omega C\bar{V}_1 = 0 \tag{ii}$$

or
$$V_2 = (1 + j\omega RC)V_1 \tag{iii}$$

Substituting in Eq. (i) and rearranging, we get

$$\beta = \frac{\bar{V}_1}{\bar{V}_o} = \frac{1}{3 + j(\omega RC - 1/\omega RC)}$$

For β to be real, the j-term should be zero. This happens when

$$\omega_o = \frac{1}{RC} \text{ or } f_o = \frac{1}{2\pi RC} \tag{iv}$$

Then

$$\beta = +\frac{1}{3} \text{ positive feedback}$$

Forward gain

$$A = \frac{\bar{V}_o}{\bar{V}_1} = \left(1 + \frac{R_2}{R_1}\right) \tag{v}$$

For oscillations to occur,

$$A\beta = \frac{1}{3}\left(1+\frac{R_2}{R_1}\right) > 1$$

$$\frac{R_2}{R_1} > 2$$

7.9 WIEN BRIDGE OSCILLATOR

A practical oscillator circuit uses an op-amp and *RC* bridge circuit, with the oscillator frequency set by the *R* and *C* components. Figure 7.14 shows a basic version of a Wien bridge oscillator circuit. Note the basic bridge connection. Two *RC* combination and R_1, R_2 form the bridge. The op-amp output is connected as the bridge input at points *a* and *c*. The bridge circuit output at points *b* and *d* provide negative and positive inputs to the op-amp.

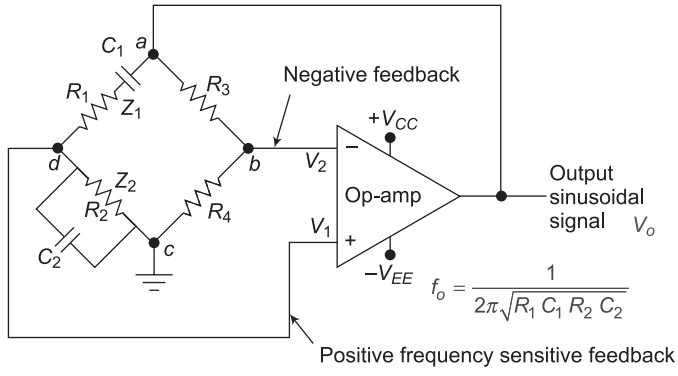


Fig. 7.14 Wein bridge oscillator circuit using an op-amp amplifier

❑ ***Proof*** For simplicity we choose $R_1 = R_2 = R$ and $C_1 = C_2 = C$

Forward gain at inverting terminal

$$A = -\frac{V_o}{V_2} = -\left(1+\frac{R_3}{R_4}\right)$$

Feedback gain at noninverting terminal

$$\beta = \frac{V_o}{V_i} = \frac{Z_2}{Z_1 + Z_2}$$

For oscillations, Barkhousen criterion

$$A\beta = -1$$

or
$$\left(1+\frac{R_3}{R_4}\right)\left(\frac{Z_2}{Z_1+Z_2}\right)=1$$

$$Z_1 = R+\frac{1}{j\omega c},\ Z_2 = \frac{2}{\frac{1}{R}+j\omega c}$$

$$=\frac{1+j\omega Rc}{j\omega c},\quad Z_2 = \frac{R}{1+j\omega cR}$$

Substituting values in Eq. (1) and setling

$$\omega = \omega_0 = \frac{1}{Rc}$$

we get the condition

$$\frac{1}{3}\left(1+\frac{R_3}{R_4}\right)=1$$

or
$$R_3 = 2R_4 \tag{7.15}$$

The frequency of oscillators in them

$$f_o = \frac{1}{2\pi\sqrt{RC}} \tag{7.16}$$

otherwise

$$f_o = \frac{1}{2\pi\sqrt{R_1C_2R_2C_2}} \tag{7.17}$$

$$Z_1 = R+\frac{1}{j\omega c}=\frac{1+j\omega Rc}{j\omega c}=\frac{1+j1}{j\omega c}$$

$$Z_2 = R\,\|\,\frac{1}{j\omega c}=\frac{j\omega Rc}{R+\frac{1}{j\omega c}}=\frac{j\omega Rc\cdot j\omega c}{1+j\omega Rc}=\frac{-\omega Rc\ \omega c}{1+j1}=\frac{-\omega c}{1+j1}$$

$$\frac{Z_2}{Z_1+Z_2}=\frac{\frac{\omega c}{1+j1}}{\frac{1+j1}{j\omega c}-\frac{j\omega c}{1+j1}}=\frac{j\omega^2c^2}{(1+j1)^2+\omega^2c^2}=\frac{j\omega^2c^2}{j2+\omega^2c^2}=\frac{j\omega^2c^2}{\omega^2c^2+j2}=\frac{R_2}{R_1}$$

Thus a ratio of R_1 to R_2 greater than 2 will provide sufficient loop gain for the circuit to oscillate at the frequency of Eq. (7.15).

EXAMPLE 7.4 Calculate the resonant frequency of the Wien bridge oscillator of Fig. 7.15.

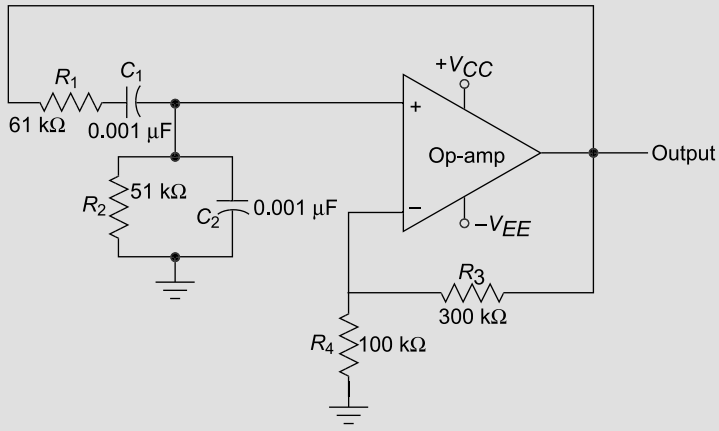


Fig. 7.15 Wein bridge oscillator circuit

Solution

$$f_o = \frac{1}{2\pi RC} = \frac{1}{2\pi(61\times10^3)(0.001\times10^{-6})} = 260.9 \text{ Hz}$$

EXAMPLE 7.5 Design the *RC* elements of a Wien bridge oscillator as in Fig. 7.15 for operation at f_o = 10 kHz.

Solution Using equal value sof *R* and *C*, we can select $R = 130$ kΩ and calculate the required value of C using Eq. (7.16):

$$C = \frac{1}{2\pi f_o R} = \frac{1}{6.28(10\times10^3)(150\times10^3)} = \frac{10^{-9}}{9.42} = 238.5$$

We can use $R_3 = 300$ kΩ and $R_4 = 100$ kΩ to provide a ratio R_3/R_4 greater than 2 for oscillation to take place.

7.10 | TUNED OSCILLATOR CIRCUIT

Tuned-Input, Tuned-Output Oscillator Circuits

A variety of circuits can be built using that shown in Fig. 7.14 by providing tuning in both the input and output sections of the circuit. Analysis of the circuit reveals that the following types of oscillators are obtained when the reactance elements are as designated:

Oscillator Type	Reactance Element		
	X_1	X_2	X_3
Colpitts oscillator	C	C	L
Hartley oscillator	L	L	C
Tuned input, tuned output	LC	LC	—

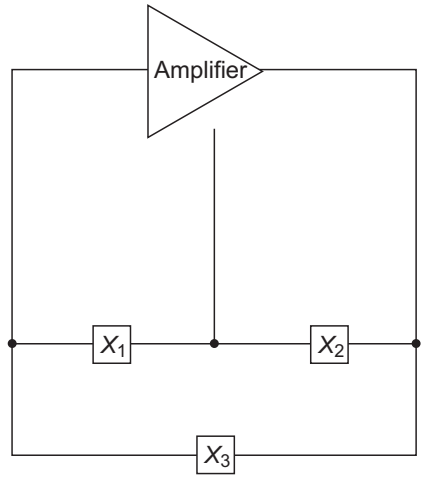


Fig. 7.16 Basic configuration of resonant circuit oscillator

7.11 COLPITTS OSCILLATOR

✦ FET Colpitts Oscillator

A practical version of an FET Colpitts oscillator is shown in Fig. 7.17. The oscillator frequency is found to be

$$f_o = \frac{1}{2\pi\sqrt{LC_{eq}}} \tag{7.18}$$

where

$$C_{eq} = \frac{C_1 C_2}{C_1 + C_2} \tag{7.19}$$

✦ Transistor Colpitts Oscillator

A transistor Colopitts oscillator circuit can be made as shown in Fig. 7.18 whose frequency of oscillation is given by Eq. (7.17).

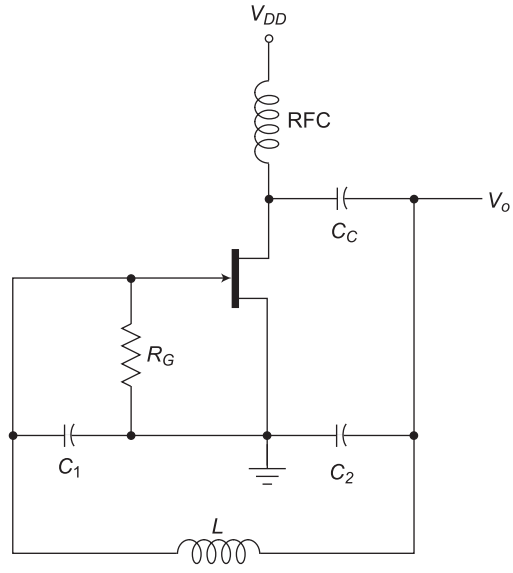


Fig. 7.17 FET Colpitts oscillator

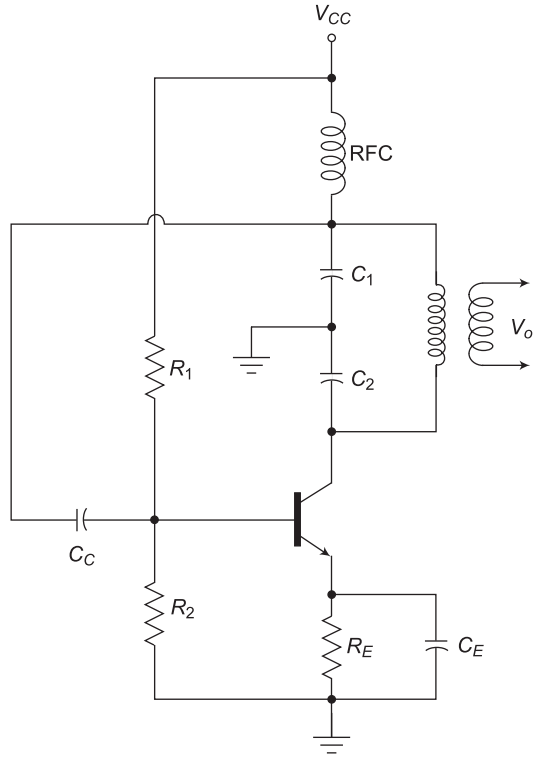


Fig. 7.18 Transistor Colpitts oscillator

✦ IC Colpitts Oscillator

An op-amp Colpitts oscillator circuit is shown in Fig. 7.19. Again, the op-amp provides the basic amplification needed, and the oscillator frequency is set by an *LC* feedback network of a Colpitts configuration. The oscillator frequency is given by Eq. (7.17).

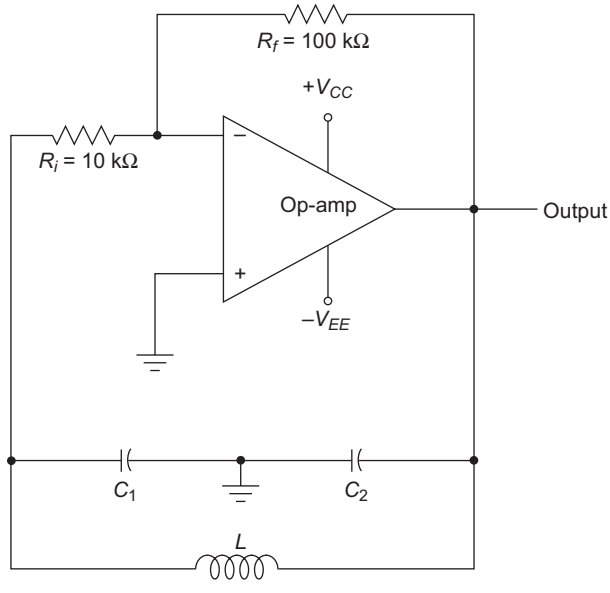


Fig. 7.19 Op-amp Colpitts oscillator

7.12 HARTLEY OSCILLATOR

If the elements in the basic resonant circuit of Fig. 7.16 are X_1 and X_2 (inductors) and X_3 (capacitor), the circuit is a Hartley oscillator.

✦ FET Hartley Oscillator

An FET Hartley oscillator circuit is shown in Fig. 7.20. The circuit is drawn so that the feedback network conforms to the form shown in the basic resonant circuit (Fig. 7.16). Note, however, the inductors L_1 and L_2 have a mutual coupling M, which must be taken into account in determining the equivalent inductance for the resonant tank circuit. The circuit frequency of oscillation is then given approximately by

$$f_o = \frac{1}{2\pi\sqrt{L_{eq}C}} \tag{7.20}$$

with

$$L_{eq} = L_1 + L_2 + 2M \tag{7.21}$$

✦ Transistor Hartley Oscillator

Figure 7.21 shows a transistor Hartley oscillator circuit. The circuit operates at a frequency given by Eq. (7.20).

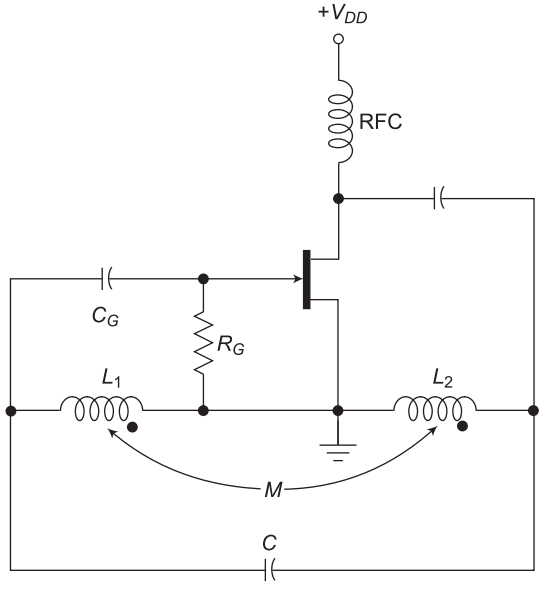


Fig. 7.20 FET Hartley oscillator

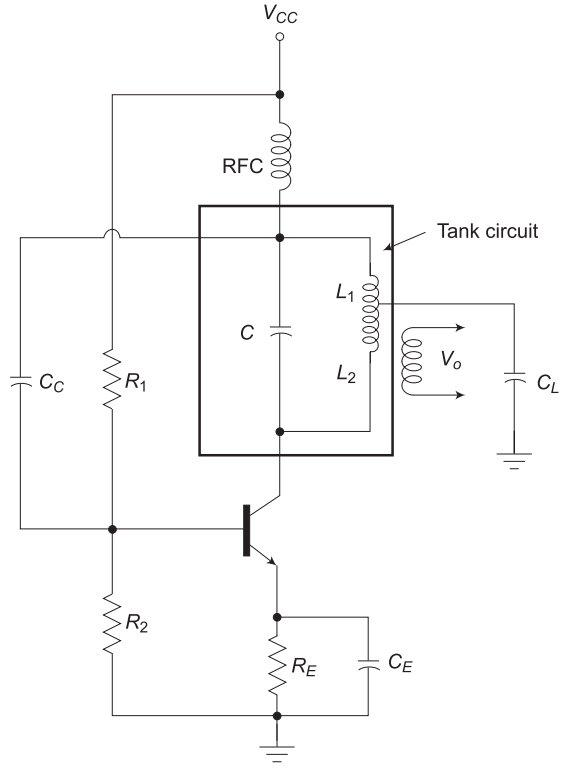


Fig. 7.21 Transistor Hartley oscillator circuit

7.13 | CRYSTAL OSCILLATOR

A crystal oscillator is basically a tuned-circuit oscillator using a piezoelectric crystal as a resonant tank circuit. The crystal (usually quartz) has a greater stability in holding constant at whatever frequency the crystal is originally cut to operate. Crystal oscillators are used whenever great stability is required, such as in communication transmitters and receivers.

Characteristics of a Quartz Crystal

A quartz crystal (one of a number of crystal types) exhibits the property that when mechanical stress is applied across one set of its faces, a difference of potential develops across the opposite faces. This property of a crystal is called the *piezoelectric effect*. Similarly, a voltage applied across one set of faces of the crystal causes mechanical distortion in the crystal shape.

When alternating voltage is applied to a crystal, mechanical vibrations are set up—these vibrations having a natural resonant frequency dependent on the crystal. Although the crystal has electromechanical resonance, we can represent the crystal action by an equivalent electrical resonant circuit as shown in Fig. 7.22. The inductor L and capacitor C represent electrical equivalents of crystal mass and compliance, respectively, whereas resistance R is an electrical equivalent of the crystal structure's internal friction. The shunt capacitance C_M represents the capacitance due to mechanical mounting of the crystal. Because the crystal losses, represented by R, are small, the equivalent crystal Q (quality factor) is high—typically 20,000. Values of Q up to almost 10^6 can be achieved by using crystals.

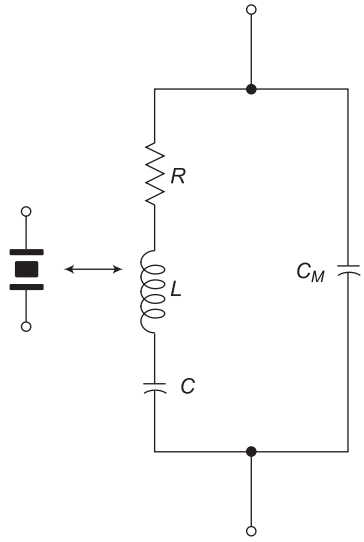


Fig. 7.22 Electrical equivalent circuit of a crystal ($C_M \gg C'$)

The crystal can have two resonant frequencies. One resonant condition occurs when the reactances of the series RLC are equal (and opposite). For this condition, the *series-resonant* impedance is very low (equal to R). The other resonant condition occurs at a higher frequency when the reactance of the series-resonant leg equals the reactance of capacitor C_M. This is a parallel resonance or antiresonance condition of the crystal. At this frequency, the crystal offers a very high impedance to the external circuit. The impedance versus frequency of the crystal is shown in Fig. 7.23.

Series mode,

$$f_1 = \frac{1}{2\pi\sqrt{LC}} \qquad C = C' + C_M \tag{7.22}$$

Parallel mode,

$$f_2 = \frac{1}{2\pi\sqrt{LC_{eq}}} \tag{7.23}$$

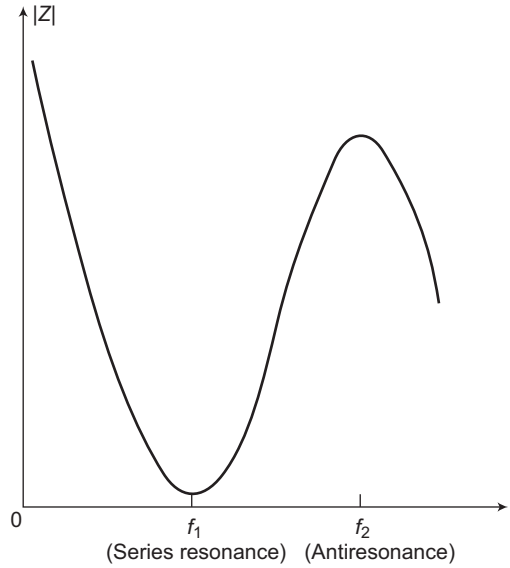


Fig. 7.23 Crystal impedance versus frequency ($f_2 > f_1$)

$$C_{eq} = C' \| C_M = \left(\frac{1}{C'} + \frac{1}{C_M} \right) \approx C_M$$

Series-Resonant Circuits

To excite a crystal for operation in the series-resonant mode, it may be connected as a series element in a feedback path. At the series-resonant frequency of the crystal, its impedance is smallest and the amount of (positive) feedback is largest. A typical transistor circuit is shown in Fig. 7.24. The RFC coil provides for dc bias while decoupling any ac signal on the power lines from affecting the output signal. The voltage feedback from collector to base is a maximum when the crystal impedance is minimum (in series-resonant mode). At operating frequency C_C and C_E are short circuit.

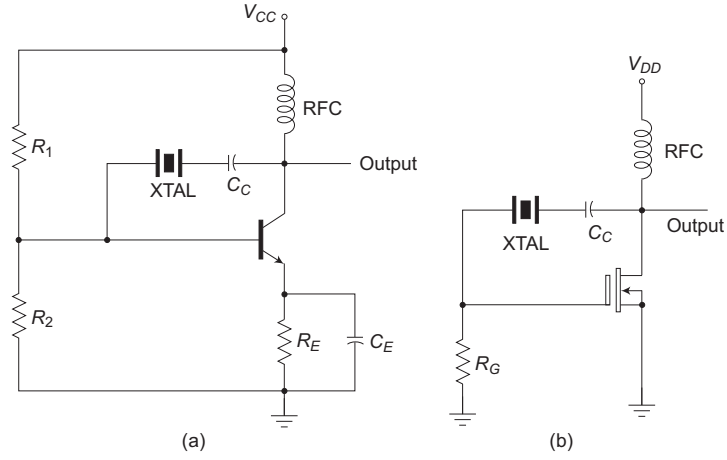


Fig. 7.24 Crystal-controlled oscillator using a crystal (XTAL) in series-feedback path: (a) BJT circuit; (b) FET circuit

The frequency of oscillation of this circuit is determined by the series resonant frequency of the crystal. The circuit frequency stability is set by the crystal frequency stability, which is good.

Parallel-Resonant Circuits

Since the parallel-resonant impedance of a crystal is a maximum value, it is connected in shunt. At the parallel-resonant operating frequency, a crystal appears as an inductive reactance of largest value. Figure 7.25 shows a crystal connected as the inductor element in a modified Colpitts circuit. The basic de bias circuit should be evident. Maximum voltage is developed across the crystal at its parallel-resonance frequency. The voltage is coupled to the emitter by a capacitor voltage divider—capacitor C_1 and C_2.

A *Miller crystal-controlled oscillator* circuit is shown in Fig. 7.26. A tuned *LC* circuit in the drain section is adjusted near the crystal parallel-resonant frequency. The maximum gate-source signal occurs at the crystal antiresonant frequency, controlling the circuit operating frequency.

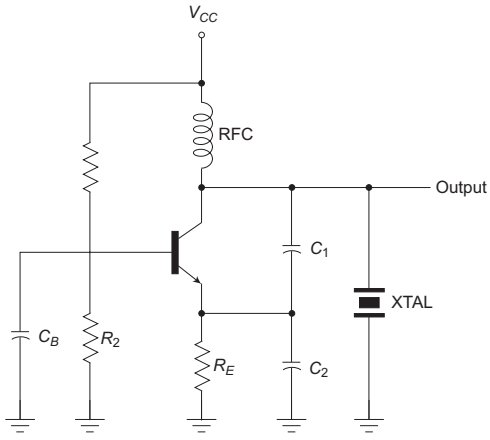


Fig. 7.25 Crystal-controlled oscillator operating in parallel-resonant mode

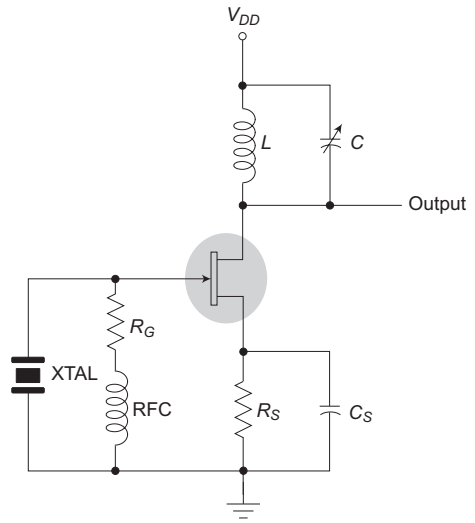


Fig. 7.26 Miller crystal-controlled oscillator

Crystal Oscillator

An op-amp can be used in a crystal oscillator as shown in Fig. 7.27. The crystal is connected in the series-resonant path and operates at the crystal series-resonant frequency. The present circuit has a high gain, so that an output square-wave signal results as shown in the figure. A pair of Zener diodes is shown at the output to provide output amplitude at exactly the Zener voltage (V_Z).

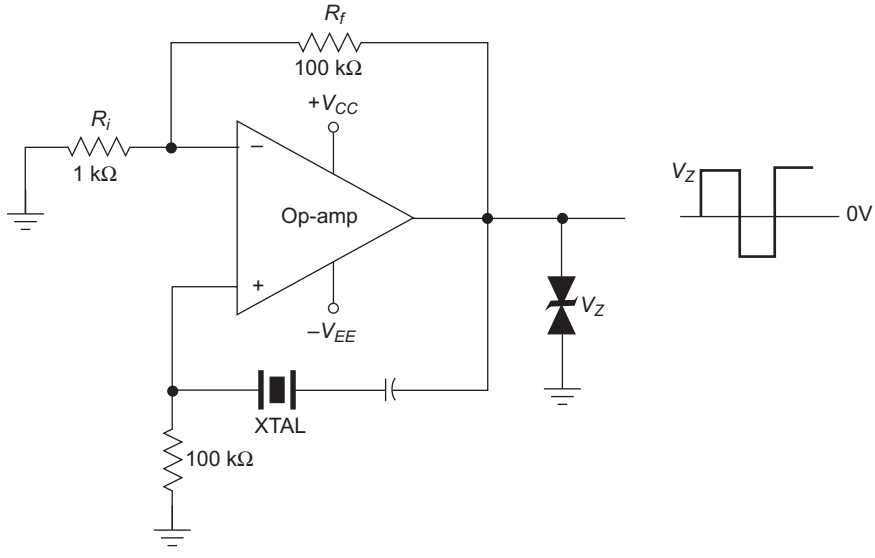


Fig. 7.27 Crystal oscillator using an op-amp

EXAMPLE 7.6 Find the capacitor C and h_{fe} for the transistor to provide a resonating frequency of 10 kHz of a transistorised phase shift oscillator. Assume R_1 = 25 kΩ, R_2 = 57 kΩ, R_C = 20 kΩ, R = 7.1 kΩ and h_{fe} = 1.8 kΩ.

Solution

$R_i' = R_1 \parallel R_2 \parallel h_{fe} = 25\text{k}\Omega \parallel 57\ \text{k}\Omega \parallel 1 - 8\ \text{k}\Omega.$

$$\frac{1}{R_i^1} = \frac{1}{25} + \frac{1}{57} + \frac{1}{1-8}$$

$$\therefore \quad R_i' = 1.631\ \text{k}\Omega$$

Now, $R_i' + R_3 = R$

$$R_3 = R - R_i' = 7.1 - 1.631 = 5.47\ \text{k}\Omega$$

$$K = \frac{RC}{R} = \frac{20}{7.1} = 2.816$$

Now, $$f = \frac{1}{2\pi\ RC\sqrt{6+4K}}$$

$$\therefore \quad 10 \times 10^3 = \frac{1}{2\pi \times 7.1 \times 10^3 \times C \times \sqrt{6+4\times 2.816}}$$

$$C = 5.39.4519$$

$$h_{fe} \geq 4K + 23 + \frac{29}{K}$$

$$h_{fe} \geq 4 \times 2.816 + 23 + \frac{29}{2.816}$$

$$h_{fe} \geq 44.562$$

EXAMPLE 7.7 The frequency sensitive arms of the Wein bridge oscillator uses C_1 = C_2 = 0.001 μF and R_1 = 10 kΩ white R_2 is kept variable. The frequency is to be varied from 10 kHz to 50 kHz, by varying R_2. Find the minimum and maximum values of R_2.

Solution The frequency of the oscillator is given by

$$f = \frac{1}{2\pi\sqrt{R_1 R_2 C_1 C_2}}$$

For f = 10 kHz

$$10 \times 10^3 = \frac{1}{2\pi(\sqrt{10\times 10^3 \times R_2 \times (0.001 \times 10^{-6})^2}}$$

$$R_2 = 25.33\ \text{K}\Omega$$

For $f = 50$ kHz

$$50 \times 10^3 = \frac{1}{\sqrt{10 \times 10^3 \times R_2 \times (0.001 \times 10^{-6})^2}}$$

$$R_2 = 10.13\ \text{k}\Omega$$

so minimum value of R_2 is 1.013 kΩ while the maximum value of R_2 is 25.33 kΩ.

EXAMPLE 7.8 In a transistorised Hartley oscillator, the two inductances are 2 mH and 20 μH while the frequency is to be changed from 950 kHz to 2050 kHz. Calculate the range over which the capacitor is to be varied.

Solution The frequency is given by,

$$f = \frac{1}{2\pi\sqrt{C(L_{eq})}}$$

where $L_{eq} = L_1 + L_2 = 2 \times 10^{-3} + 20 \times 10^{-6} = 0.00202$

For $f = f_{\max} = 2050$ kHz

$$2050 \times 10^3 = \frac{1}{2\pi\sqrt{C \times 0.00202}}$$

$$950 \times 10^3 = \frac{1}{2\pi\sqrt{C \times 0.00202}}$$

$$C = 13.89\ \text{PF}$$

EXAMPLE 7.9 Find the frequency of oscillations of transistorised Colpitts oscillator having tank circuit parameters or C_1 = 150 PF, C_2 = 1.5 nF, and L = 50 μH.

Solution The equivalent capacitance is given by,

$$C_{eq} = \frac{C_1 C_2}{C_1 + C_2} = \frac{150 \times 10^{-12} \times 1.5 \times 10^{-9}}{150 \times 10^{-12} + 1.5 \times 10^{-9}}$$

$$C_{eq} = 136.363 \times 10^{-12}\ \text{F}$$

$$f = \frac{1}{2\pi\sqrt{LC_{eq}}} = \frac{1}{2\pi\sqrt{50 \times 10^{-6} \times 136.363 \times 10^{-12}}} = 1.927\ \text{MHz}$$

EXAMPLE 7.10 A crystal L = 0.4 H, C = 0.085 pF and C_m = 1 PF with R = 5 kΩ. Find

(a) Series resonant frequency
(b) Parallel resonant frequency
(c) By what percent does the parallel resonant frequency exceed the series resonant frequency?
(d) Find the Q factor of the crystal.

Solution

(a) $$f_S = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{0.4\times0.085\times10^{-12}}} = 0.856 \text{ MHz}$$

(b) $$C_{eq} = \frac{CC_M}{C+C_M} = \frac{0.085\times1}{0.085+1} = 0.078 \text{ PF}$$

$$\therefore \quad f_p = \frac{1}{2\pi\sqrt{LC_{eq}}} = \frac{1}{2\pi\sqrt{0.4\times0.078\times10^{-12}}} = 0.899 \text{ MHz}$$

(c) $$\% \text{ increase} = \frac{0.899-0.856}{0.856}\times100 = 5.023\%$$

(d) $$Q = \frac{\omega_S L}{R} = \frac{2\pi f_S L}{R} = \frac{2\pi\times0.856\times10^6\times0.4}{5\times10^3} = 430.272$$

EXAMPLE 7.11 For a particular phase-shift oscillator, the following specifications are given: C = 0.1 μF, R = 3.9 kΩ, and $\left|\frac{R_F}{R_1}\right| = 29$. Determine the frequency of oscillation.

Solution The frequency of oscillation is determined as per the formula,

$$f = \frac{1}{2\pi\sqrt{6}RC} = \frac{1}{2\pi\sqrt{6}\times0.1\times10^{-6}\times3.9\times10^3}$$

Hence, frequency of oscillation = 1.66.7.

EXAMPLE 7.12 A certain Wein-bridge oscillator uses R = 9.4 kΩ, C = 0.01 μF and $RF = 2R_1$. What is the frequency of oscillation.

Solution The frequency of oscillation is calculated as per the formula,

$$f_o = \frac{1}{2\pi RC} = \frac{0.5}{3.14}\times\frac{1}{9.4\times10^3}\times\frac{1}{10^{-2}\times10^{-6}} = 169.3 \text{ kHz}$$

EXAMPLE 7.13 Design a phase-shift oscillator so that f_0 = 1 kHz.

Solution Let $C = 0.1$

Then, from the equation $f_o = \dfrac{1}{2\pi\sqrt{6}RC} = \dfrac{0.065}{RC}$

$$R = \frac{0.065}{(1000)(10^{-7})} = 650\ \Omega = 6.5\ \text{k}\Omega$$

To prevent the loading of the amplifier because of RC networks, it is necessary that $R_1 \geq 10R$

Therefore, let $R_1 = 10R = 65\ \text{k}\Omega$

Then, from the equation $R_F = 29R_1$.

$$R_F = 29(65) = 1885\ \Omega = 1.89\ \text{k}\Omega \text{ (approximately)}$$

(Use R_F = 1 MΩ Potentiometer)

When choosing an op-amp, type 741 can be used at lower frequencies (<1 kHz); however, at higher frequencies, an op-amp such as the LM 318 or LF 351 is recommended because of its increased slow rate.

EXAMPLE 7.14 Design the Wein-bridge oscillator so that f_o = 1 kHz.

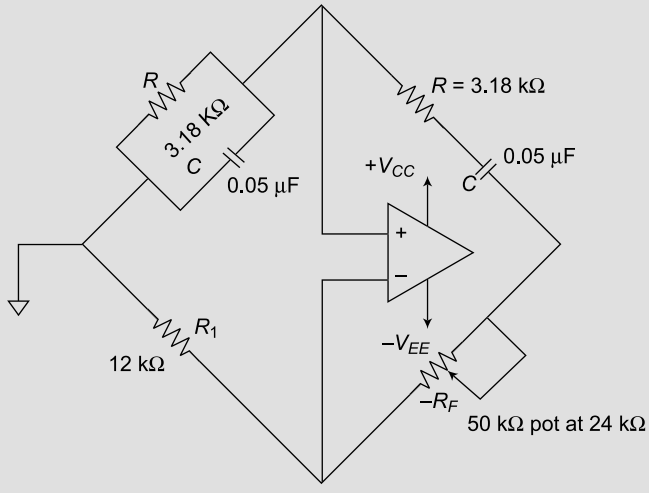


Fig. 7.28

Solution Let $C = 0.05\ \mu F$

Therefore, from the equation $f_o = \dfrac{1}{2\pi RC}$

$$f_o = \frac{0.59}{5\times10^{-8}\times R}$$

$$\therefore \quad R = \frac{0.159}{5\times10^{-8}\times10^{3}} = 0.0318\times10^{5} = 3.18\ k\Omega$$

Now, let $R_1 = 12\ k\Omega$.

Then, from the equation $RF = 2R_1$.

$$R_F = 2 \times 12\ K\Omega = 24\ K\Omega$$

EXAMPLE 7.15 Calculate the Resonant frequency of the Wien-bridge oscillator of Fig. 7.29.

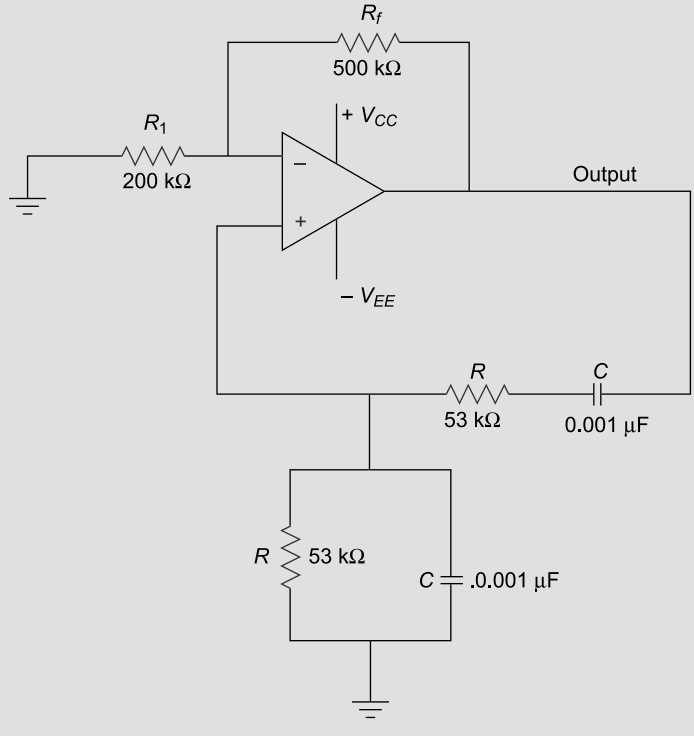


Fig. 7.29 Wien bridge oscillator

Solution Resonant frequency of Wien bridge is given by

$$f_o = \frac{1}{2\pi RC}$$

$$= \frac{1}{2\pi \times (53\times10^3 \times 0.001\times10^{-6})}$$

$$= 3004.44 \text{ Hz.}$$

Take R_f = 500 kΩ and R_1 = 200 kΩ whose ration R_f/R_1 is greater than 2, which provide oscillation.

EXAMPLE 7.16 Design the *RC* elements of a Wien bridge oscillator as shown in Fig. 7.27 for operation at f_0 = 5 kHz.

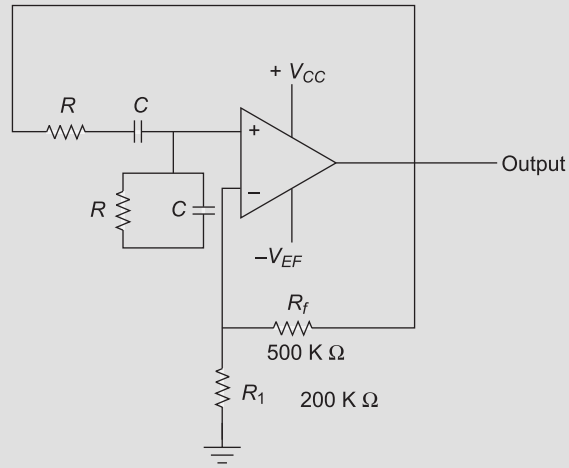


Fig. 7.30 Wien bridge Oscillator

Solution Assume R = 100 kΩ

from the Resonant frequency of Wien bridge oscillator

$$f_o = \frac{1}{2\pi RC}$$

$$C = \frac{1}{2\pi f_0 R}$$

$$= \frac{1}{6.28\times(5\times10^3)\times(100\times10^3)}$$

$$= \frac{10^{-8}}{31.4} = 318 \text{ pF}$$

SUMMARY

➢ Various types of feedback amplifiers and oscillators have been introduced. Gain and bandwidth analysis has been presented.

EXERCISES

➥ Review Questions

1. State the Barkhhausen criterion, that is, the conditions necessary for sinusoidal oscillations to be sustained.
2. In feedback amplifiers, the gain-bandwidth product remains constant. Comment on this statement.
3. Sketch the phase-shift oscillator using (a) an op-amp, and (b) a JFET.
4. By what factor does the feedback reduce the gain of an amplifier?
5. How does negative feedback stabilise the gain of an amplifier?
6. How does negative feedback affect the bandwidth of the amplifier?
7. What is positive feedback and why is it avoided in amplifier circuits?
8. Why is negative feedback used in high gain amplifiers?
9. Explain the concept of feedback in amplifiers. Also, explain why we need to use proper feedback in electronic circuits.

➥ Problems

1. A FET oscillator uses the phase-shifting network as shown in Fig. 7.31.
   (a) What is the frequency of oscillation?
   (b) What is the minimum gain necessary for oscillation?

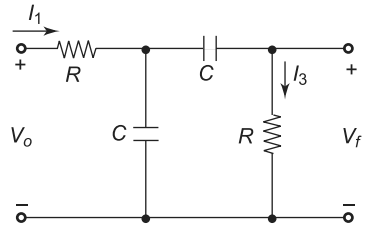


Fig. 7.31

2. An FET phase-shift oscillator has g_m = 6000 μs, r_d = 36 kΩ, feedback resistance R = 6 kΩ. Calculate C for oscillation frequency of 3 kHz.
3. A BJT phase-shift oscillator has the following data: R = 6 kΩ, C = 1600 pF, R_C = 20 kΩ. Calculate the oscillation frequency. What should be the current gain of the BJT?
4. For the lag-lead network of Fig. 7.32, determine $\frac{\bar{V}_o}{\bar{V}_i}$.

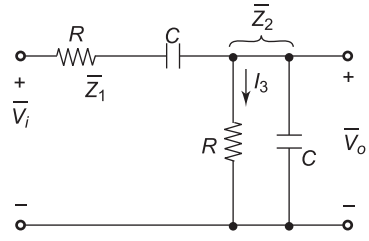


Fig. 7.32

5. The following diagram shows a Wein bridge oscillator using an amplifier with a non-zero output resistance R_{out}. The op-amp is ideal. Derive expressions for the frequency of oscillation and the conditions that R_1/R_2 must satisfy for sustained oscillations.

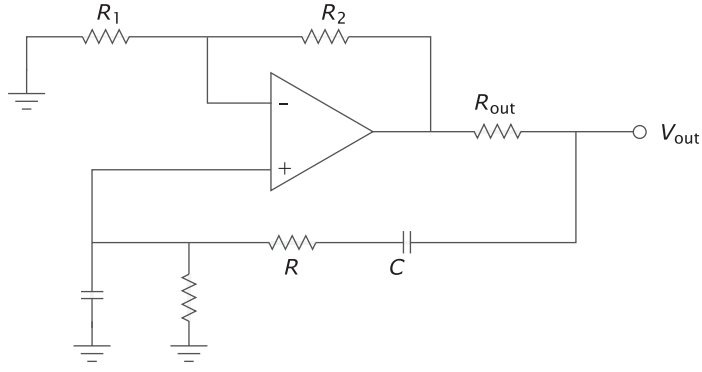


Fig. 7.33

6. The following circuit is used as the phase shifting network for a two-stage FET oscillator. Find the circuits beta, $\beta(s) = \frac{V_f}{V_0}$. Determine the frequency of oscillation and the gain required from the amplifier.

7. Design a 680 kHz Wein-Bridge oscillator. Use an ideal op-amp as your active element.

8. For the oscillator shown below, derive an expression for the frequency of oscillation in terms of R and C. What minimum value of $\frac{R_2}{R_1}$ is required for oscillations to be maintained?

9. For the Colpitts oscillator shown below, find values for L_2, C_3 and R_D appropriate to produce sustained oscillations at 100 kHz if C_1 = 10 nf – Use g_m = 1 mA/V.

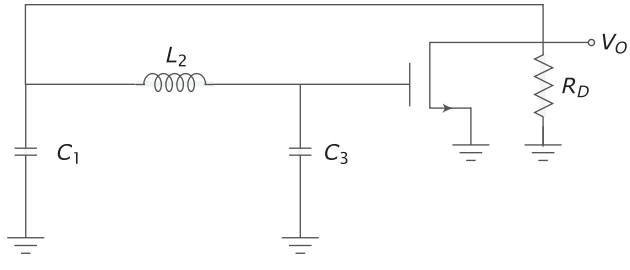


Fig. 7.34

10. Calculate the value of $C_1 = C_2$ for the Wein bridge oscillator to operate at a frequency of 20 kHz. Assume $R_1 = R_2 = 50\ k\Omega$ and $R_3 = 3R_4 = 600\ \Omega$.

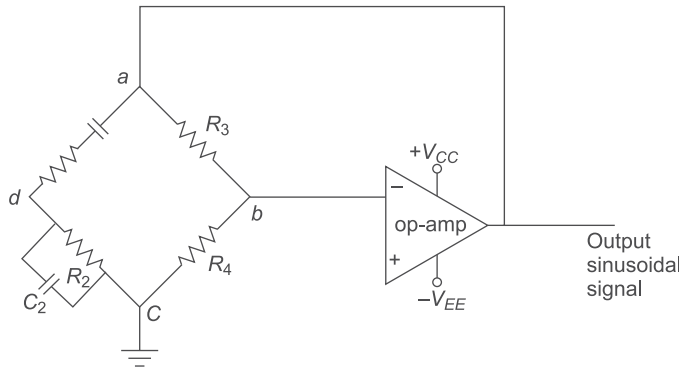


Fig. 7.35

➥ Multiple-Choice Questions

1. In a current-series feedback amplifier, the output resistance
(a) decreases (b) increases (c) remains same (d) none of these

2. The only drawback of using negative feedback in amplifier is that it involves
(a) gain stability (b) frequency stability
(c) temperature sensitivity (d) gain sacrifice

3. An amplifier with no feedback has a gain-bandwidth product of 4 MHz. Its closed-loop gain is 40. The new bandwidth is
(a) 20 kHz (b) 160 MHz (c) 10 MHz (d) 100 kHz

4. Negative feedback in an amplifier
(a) lowers its lower 3 dB frequency (b) raises its upper 3 dB frequency
(c) increases its bandwidth (d) all of these

5. For sustaining oscillations in an oscillator
(a) feedback factor should be unity (b) phase shift should be unity
(c) feedback should be negative (d) both (a) and (b)

6. A Hartley oscillator uses
(a) inductive feedback(b) opacitive feedback (c) resistive feedback (d) none of these

7. A Colpitts oscillator uses
(a) tapped coil (b) inductive feedback (c) tapped capacitance (d) no tuned *LC* circuit

8. The primary advantage of a crystal oscillator is that
(a) it operates on a very low dc supply voltage(b) its frequency of oscillation remains constant
(c) it gives a high output voltage (d) it can oscillate at any frequency

9. In *RC* phase shift oscillator circuits,
(a) feedback factor is less then unity (b) there is no need for feedback
(c) pure sine-wave output is possible (d) transistor parameters determine oscillation frequency

10. The Barkhausen criterion for sustained oscillation is given by
(a) $AB = 1$ (b) $|AB| \geq 1$ (c) $|AB| \leq 1$ (d) $AB = 180°$

11. The value of negative feedback fraction is always
(a) less than 1 (b) more than 1 (c) equal to 1 (d) none of these

12. Negative feedback is employed in .....
(a) oscillators (b) rectifiers (c) amplifiers (d) none of these

13. Which among the following parameters increases due to positive feedback? (i) Input voltage (ii) Output Voltage (iii) Noise (iv) Voltage Gain
(a) i and ii (b) iii only (c) ii and iv (d) all of these

ANSWERS

✦ Problems

1. (a) $1/2\pi RC$, $A(\text{min}) = 3$

2. 433 μF

3. 5.4 kHz, more than 45

4. $\dfrac{j\omega RC}{1 - \omega^2 RC \rightarrow 3j\omega RC}$

✦ Multiple-Choice Questions

1. b **2.** d **3.** d **4.** d **5.** d **6.** a **7.** c **8.** b **9.** c **10.** a
11. a **12.** c **13.** d

CHAPTER 8

Regulated Power Supplies

GOALS AND OBJECTIVES

- Block diagram and explanation of regulated power supplies
- Types of regulators—series voltage, shunt voltage and monolithic linear regulators

8.1 INTRODUCTION

In Chapter 2, we have studied diode full-wave rectifiers with output filter to reduce the ripple. In spite of that, the voltage regulation is unacceptable as dc power supply for electronic circuits, in particular electronic gates, draw peak current. Therefore, special voltage regulators are needed at the output of the power supply. There are two types of regulators–discrete transistor type and IC type.

8.2 SERIES VOLTAGE REGULATOR

The regulator is placed in between (series) the supply and load. The regulator block diagram is drawn in Fig. 8.1.

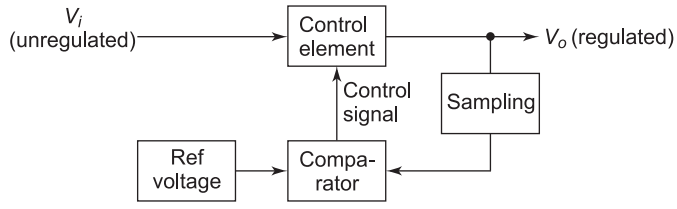


Fig. 8.1 Regulator block diagram

Series Regulator Circuits

✦ Single Transistor

The circuit is drawn in Fig. 8.2. As V_i varies, V_o remains constant as by design

$$V_i \text{ (min)} > V_z$$

$$V_o = V_z - V_{BE} \tag{8.1}$$

As V_o changes, V_{BE} changes slightly, but there is a large change in $I_E = I_L$ restricting the change in V_o.

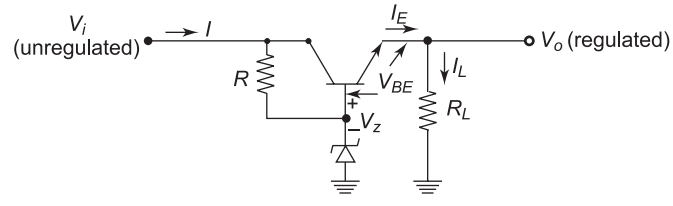


Fig. 8.2 Linear regulator

The operation is in linear range and so it is called *linear regulator*. The efficiency of this regulator is high, 50–70%, and it is used for 10 W load.

$I_E = I_1$ is called *pass-current*.

✦ Protection

The circuitry is suitably modified with additional components for limiting the current, thereby providing short-circuit protection. One of these techniques is 'foldback' current limiting.

8.3 | SHUNT VOLTAGE REGULATOR

The schematic block diagram of the shunt regulator is drawn in Fig. 8.3.

If V_o is high, the control action is to draw shunt current I_{sh} from the input current. $I_L = I_i - I_{sh}$, reduces causing V_o to maintain the desired value. If V_o is low, I_{sh} is added to I_i, and the reverse action takes place.

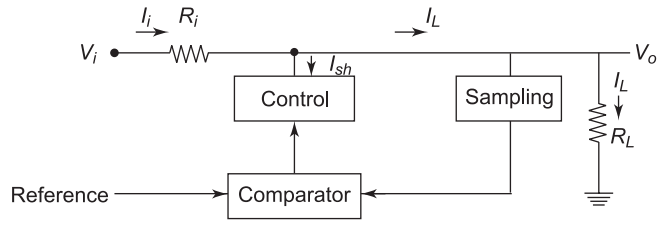


Fig. 8.3 Shunt regulator, block diagram

8.3.1 Shunt Regulator Circuit (Fig. 8.4)

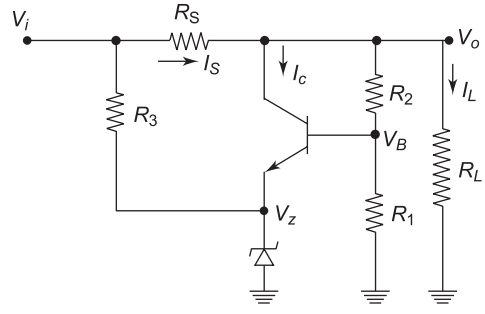


Fig. 8.4 Shunt regulator

The potential divider (R_1, R_2) has high resistance and so the current drawn by it can be ignored. The collector current is the shunt current, which controls the load current.

$$I_L = I_S - I_C$$

to regulate V_o

$$V_B = \frac{R_1}{R_1 + R_2} V_o \tag{8.2}$$

Also, $$V_B = V_z + V_{BE}$$

Thus, $$V_o = \left(1 + \frac{R_2}{R_1}\right)(V_z + V_{BE}) \tag{8.3}$$

which is higher than $V_z - V_o$, and is well regulated except for small variations in V_{BE}.

8.3.2 OP-AMP Shunt Regulator

This circuit is drawn in Fig. 8.5. Because of the feedback gain of op-amp, there is better control over I_C and therefore over V_o.

$$V_o = \left(\frac{R_1}{R_1 + R_2}\right) V_z$$

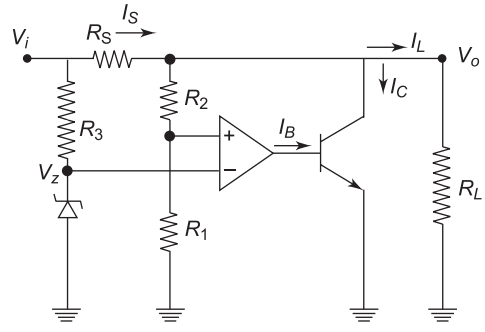


Fig. 8.5 Op-amp shunt regulator

or
$$V_o = \left(1 + \frac{R_2}{R_1}\right) V_z \tag{8.4}$$

V_o is higher than V_z.

It is to be observed that V_o in Eq. (8.4) compared to Eq. (8.3) is independent of V_{BE} and so unaffected by its variations.

✦ Short-Circuit Protection

The shunt regulators are protected against short circuit as such. If output is shorted in Figs. 8.4 and 8.5, no component will be affected. The input current increases to

$$I_s(\text{sc}) = \frac{V_i}{R_s}$$

❑ ***Remark*** Series regulators have better efficiency than shunt regulators, so these are preferred.

EXAMPLE 8.1 Calculate the approximate output voltage of Fig. 8.6. What is the power dissipation of the pass transistor (Q_2)?

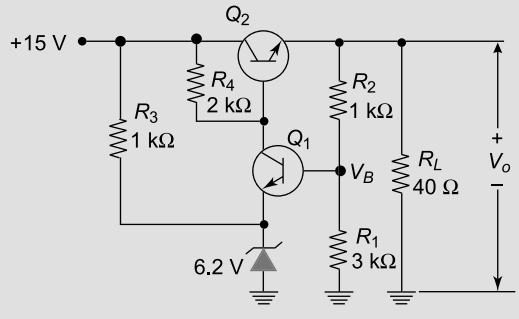


Fig. 8.6

Solution

$$V_B = 6.2 + 0.7 = 6.9 \text{ V}$$

$$V_o = 6.9\left(\frac{1+3}{3}\right) = 9.2 \text{ V}$$

Pass-transistor dissipation

$$P_D = V_{CE}\, I_C$$

$$I_C \approx \frac{V_o}{R_L} = \frac{9.2}{40} = 230 \text{ mA}$$

$$V_{CE} = 15.0 - 9.2 = 5.8 \text{ V}$$

$$P_D = 5.8 \times 230 = 1334 \text{ mW or } 1.334 \text{ W}$$

Reader: What is the approximation in V_o as calculated?

EXAMPLE 8.2 Calculate the approximate efficiency of the regulator of Fig. 8.6.

Solution

$$P_{\text{out}} = 9.2 \times 230 = 2.12 \text{ W}$$

Current drawn by R_3

$$I(R_3) = \frac{15 - 6.2}{1} = 8.8 \text{ mA}$$

$$I_{\text{in}} = I_c + I(R_3) = 230 + 8.8 = 239 \text{ mA}$$

$$P_{\text{in}} = 15 \times 239 = 3.585 \text{ W}$$

$$\eta = \frac{2.12}{3.585} \times 100 = 59.1\%$$

EXAMPLE 8.3 Explain the operation of the two-transistor regulator of Fig. 8.7. What is the expression for V_o?

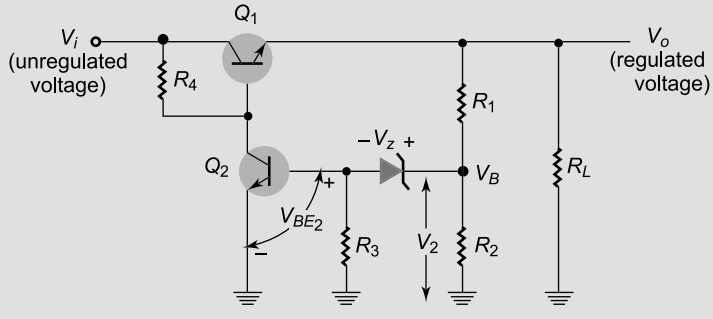


Fig. 8.7

Solution If V_o goes up, V_B increases.

$$V_B = V_z + V_{BE2},\ V_z = \text{constant}$$

As BE_2 increases, the collector current of Q_2 increases and the base current of Q_1 reduces, decreasing the pass current. So V_o returns to regulated value.

$$V_o = V_B\left(\frac{R_1 + R_2}{R_2}\right) = \left(1 + \frac{R_1}{R_2}\right)(V_z + V_{BE_2})$$

Changes in BE_2 are of negligible order.

8.4 MONOLITHIC LINEAR REGULATORS

There is a wide variety of linear **IC voltage regulators** with pin counts from 3 to 14. All are series regulators because they are more efficient than the shunt regulator. Some IC regulators are used in special applications in which external resistors can set the current, limiting the output voltage, and so on. By far, the most widely used IC regulators are those with only three pins: one for the unregulated input voltage, one for the regulated output voltage, and one for ground.

Available in plastic or metal packages, the three-terminal regulators have become extremely popular because they are inexpensive and easy to use. Aside from two optional bypass capacitors, three-terminal IC voltage regulators require no external components.

8.4.1 Basic Types of IC Regulators

Most IC voltage regulators have one of these types of output voltage: fixed positive, fixed negative, or adjustable. IC regulators with fixed positive or negative outputs are factory-trimmed to get different fixed voltages with magnitudes from about 5 to 24 V. IC regulators with an adjustable output can vary the regulated output voltage from less than 2 to more than 40 V.

IC regulators are also classified as standard, low power and low drop-out. *Standard IC regulators* are designed for straightforward and non-critical applications. With heat sinks, a standard IC regulator can have a load current of more than 1 A.

If load currents up to 100 mA are adequate, *low-power IC regulators* are available in TO-92 packages, the same size used for small-signal transistors like the 2N3904. Since these regulators do not require heat sinking, they are convenient and easy to use.

The **drop-out voltage** of an IC regulator is defined as the minimum headroom voltage needed for regulation. For instance, standard IC regulators have a drop-out voltage of 2 to 3 V. This means that the input voltage has to be at least 2 to 3 V, which is greater than the regulated output voltage, for the chip to regulate to specifications. Applications when 2 to 3 V of headroom is not available, *low drop-out IC regulators* can be used. These regulators have typical drop-out voltages of 0.15 V for a load current of 100 mA and 0.7 V for a load current of 1 A.

8.4.2 On-Card Regulation versus Single-Point Regulation

With *single-point regulation*, we need to build a power supply with a large voltage regulator and then distribute the regulated voltage to all the different *cards* (printed-circuit boards) in the system. This creates problems. To begin with, the single regulator has to provide a large load current equal to the sum of all the card currents. Second, noise or other **ElectroMagnetic Interference (EMI)** can be induced on the connecting wires between the regulated power supply and the cards.

Because IC regulators are inexpensive, electronic systems that have many cards often use *on-card regulation*. This means that each card has its own three-terminal regulator to supply the voltage used by the components on that card. By using on-card regulation, we can deliver an unregulated voltage from a power supply to each card and have a local IC regulator take care of regulating the voltage for its card. This eliminates the problems of the large load current and noise pick-up associated with single-point regulation.

8.4.3 Load and Line Regulation Redefined

Up to now, we have used the original definitions for load and line regulation. Manufacturers of fixed IC regulators prefer to specify the change in load voltage for a range of load and line conditions. Here are definitions for load and line regulation used on the data sheets of fixed regulators:

Load regulation = ΔV_{out} for a range of load current

Line regulation = ΔV_{out} for a range of input voltage

For instance, the LM7815 is an IC regulator that produces a fixed positive output voltage of 15 V. The data sheet lists the typical load and line regulation as follows:

Load regulation = 12 mV for I_L = 5 mA to 1.5 A

Line regulation = 4 mV for V_{in} = 17.5 V to 30 V

The load regulation will depend on the conditions of measurement. The foregoing load regulation is for T_J = 25°C and V_{in} = 23 V. Similarly, the foregoing line regulation is for T_J = 25°C and I_L = 500 mA. In each case, the junction temperature of the device is 25°C.

8.4.4 The LM7800 Series

The LM78XX series (where XX = 05, 06, 08, 10, 12, 15, 18, or 24) is typical of the three-terminal voltage regulators. The 7805 produces an output of +5 V, the 7806 produces +6 V, the 7808 produces +8 V, and so on, up to 7824, which produces an output of +24 V.

Figure 8.8 shows the functional block diagram for the 78XX series. A built-in reference voltage V_{ref} drives the non-inverting input of an amplifier. The voltage regulation is similar to our earlier discussion. A voltage divider consisting of R'_1 and R'_2 samples the output voltage and returns a feedback voltage to the inverting input of a high-gain amplifier. The output voltage is given by

$$V_{out} = \frac{R'_1 + R'_2}{R'_1} V_{ref}$$

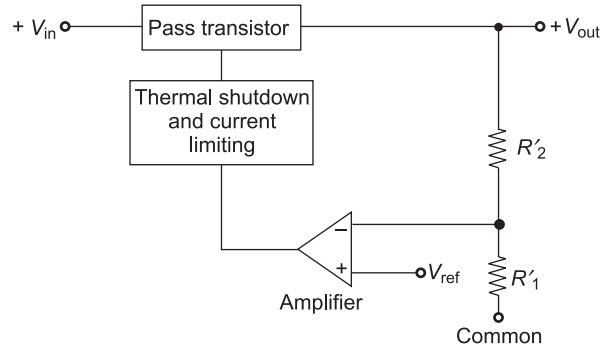


Fig. 8.8 Functional block diagram of three-terminal IC regulator

In this equation, the reference voltage is equivalent to the zener voltage in our earlier discussions. The primes attached to R'_1 and R'_2 indicate that these resistors are inside the IC itself, rather than being external resistors. These resistors are factory-trimmed to get the different output voltages (5 to 15 V) in the 78XX series. The tolerance of the output voltage is ±4 percent.

The LM78XX includes a pass transistor that can handle 1 A of load current, provided that adequate heat sinking is used. Also included are thermal shutdown and current limiting. **Thermal shutdown** means that the chip will shut itself off when the internal temperature becomes too high, around 175°C. This is a precaution against excessive power dissipation, which depends on the ambient temperature, type of heat sinking, and other variables. Because of thermal shutdown and current limiting, devices in the 78XX series are almost indestructible.

8.4.5 Fixed Regulator

Figure 8.9(a) shows an LM7805 connected as a fixed voltage regulator. Pin 1 is the input, Pin 2 is the output and Pin 3 is ground. The LM7805 has an output voltage of +5 V and a maximum load current over 1 A. The typical load regulation is 10 mV for a load current between 5 mA and 1.5 A. The typical line regulation is 3 mV for an input voltage of 7 to 25 V. It also has a ripple rejection of 80 dB, which means that it will reduce the input ripple by a factor of 10,000. With an output resistance of approximately 0.01 Ω, the LM7805 is a very stiff voltage source to all loads within its current rating.

When an IC is more than 6 from the filter capacitor of the unregulated power supply, the inductance of the connecting wire may produce oscillations inside the IC. This is why manufacturers recommend using a bypass capacitor C_1 on Pin 1 [Fig. 8.9(b)]. To improve the transient response of the regulated output voltage, a bypass capacitor C_2 is sometimes used on Pin 2. Typical values for either bypass capacitor are from 0.1 to 1 μF. The data sheet of the 78XX series suggests 0.22 μF for the input capacitor and 0.1 μF for the output capacitor.

Any regulator in the 78XX series has a drop-out voltage of 2 to 3 V, depending on the output voltage. This means that the input voltage must be at least 2 to 3 V greater than the output voltage. Otherwise, the chip stops regulating. Also, there is a maximum input voltage because of excessive power dissipation. For instance, the LM7805 will regulate over an input range of

approximately 8 to 20 V. The data sheet for the 78XX series gives the minimum and maximum input voltages for the other preset output voltages.

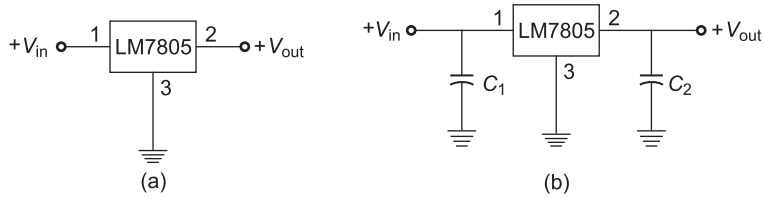


Fig. 8.9 (a) Using a 7805 for voltage regulation (b) Input capacitor prevents oscillations and output capacitor improves frequency response

SUMMARY

➢ The regulated power supplies are explained and different types of regulators are discussed.

EXERCISES

➥ Review Questions

1. Explain an IC.
2. Describe the scheme of regulation of a series regulator. Why should it be less than the input voltage?
3. What is a pass transistor?
4. What determines the output voltage in a series regulator? Why should it be less than the input voltage?
5. Describe the scheme of regulation of a shunt regulator.
6. In applications why is a series regulator preferred over shunt regulator?
7. Draw the circuit diagram of a series regulator with a single transistor and also with op-amp.
8. Why is current-limiting protection essential for series regulator?
9. Draw the circuit of a series regulator with current limiting. Describe its operation.
10. Draw the circuit of a series regulator with feedback current limiting. Describe its operation. Why is it superior to ordinary current limiting?
11. For IC regulators:
    (a) What is dropout voltage?
    (b) Define line and load regulation.
    (c) What is meant by on-card regulation?
    (d) What is an adjustable regulator and its range of adjustment? What are its three terminals?
    (e) Explain the operation of a three-terminal regulator of Fig. 8.8.

➥ Problems

1. For the series regulator of Fig. 8.10, calculate V_o and zener diode current. What will be these values if V_i reduces to 12 V?
2. Identify the kind of regulatoring in Fig. 8.11. Calculate V_o and I_z.

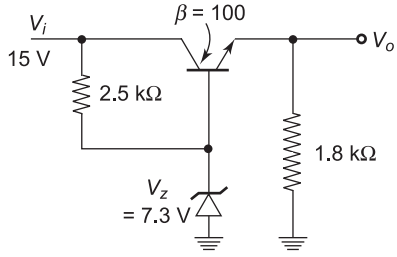


Fig. 8.10

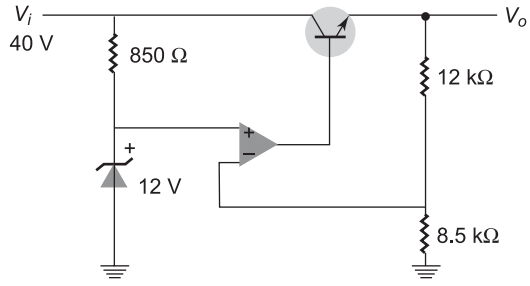


Fig. 8.11

3. For the regulator circuit of Fig. 8.12, calculate the approximate value of V_{out}. What is the power dissipation of the pass transistor?

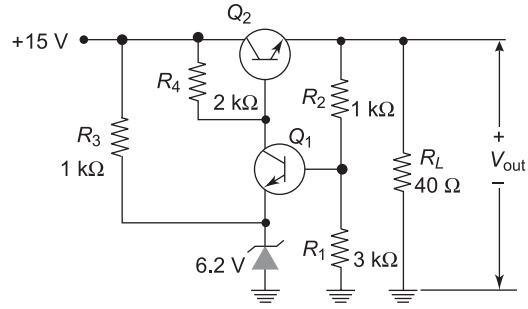


Fig. 8.12

Hint: I_c is approximately same as load current.

4. Calculate the efficiency of the regulator of Fig. 8.12.

➥ Multiple-Choice Questions

1. A voltage regulator circuit is that which
- **(a)** smoothens the ac variation in dc output voltage
- **(b)** converts ac voltage to dc voltage
- **(c)** maintains a constant dc output voltage in spite of the fluctuations in ac voltage or load current
- **(d)** none of the above

2. Which of the following voltage regulators is preferred for providing large values of load current?
(a) Transistor series regulator (b) Zener diode shunt
(c) Transistor shunt regulator (d) None of these

3. An ideal voltage regulator has voltage regulation of
(a) 50 (b) 1 (c) 100 (d) 0

4. The main job of a voltage regulator is to provide a nearly …….. output voltage.
(a) fluctuating (b) smooth (c) constant (d) sinusoidal

5. The output voltage of a step-down-type switching voltage regulator depends on
(a) input voltage (b) transistor on time (c) duty cycle (d) all of these

6. In a feedback series voltage regulator circuit, the output voltage is regulated by controlling the ….
(a) reference voltage
(b) magnitude of input voltage

(c) gain of the feedback transistor
(d) voltage drop across the series pass transistor

7. A switching voltage regulator can be of the following type:
(a) Inverting (b) Step down (c) Step up (d) All of these

8. The power efficiency of a switching voltage regulator is much higher than that of a linear regulator because of
(a) cut-off (b) saturation (c) high duty cycle (d) it is like a switch

9. Which of the following is not an essential element of dc power supply?
(a) Filter (b) Voltage amplifier (c) Rectifier (d) Voltage regulator

10. As compared to voltage regulators made up of discrete components, IC regulators have the inherent advantage(s) of
(a) Self-protection against over temperature (b) remote control
(c) current limiting (d) all of these

ANSWERS

✦ Problems

1. $V_o = 8$ V, $I_2 = 3.08$ mA, V_o no change, $I_z = 1.88$ mA
2. Series regulator, $V_o = 28.94$, $I_z = 0.033$ mA
3. $V_{out} = 9.2$ V, $P_D = 1.33$ W
4. 61.4%

✦ Multiple-Choice Questions

1. (c) **2.** (a) **3.** (d) **4.** (c) **5.** (d) **6.** (d) **7.** (d) **8.** (d) **9.** (b) **10.** (d)

CHAPTER 9

Digital Logic

GOALS AND OBJECTIVES

- Introduction to analog/discrete (digital) signal and frequency spectrum
- Switching and logic levels of digital waveforms
- Discussion of characterisation of digital ICs—speed operation, power description, current and voltage parameters, noise immunity and fan-out
- Description and functional operation of logic gates

9.1 INTRODUCTION

In an analog (continuous) form of information (signal), the signal may acquire any value in a range of the independent variable (say time). In a discrete (or digital) form, the signal can have any value but it would remain constant over periods of time (called *sampling periods*), and also at sampling instant. An audio signal (which is analog) can be recorded on a phonograph in the form of a continuous groove of depth varying in accordance with signal strength (at constant speed, which converts time to distance variable). Alternatively, discrete values of audio signal could be recorded on a laser disc in a pattern of flat areas of holes, which reflect or do not reflect light. Of course, discretisation must be over short-enough intervals so as to preserve the amplitude and frequency information of the original analog signal.

Figure 9.1 shows the process of discretising (or sampling) an analog signal, which could also be reversed.

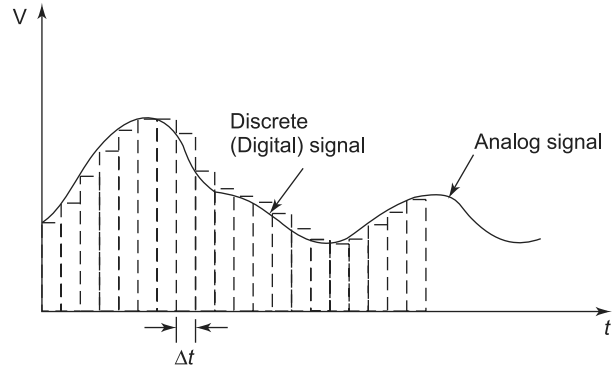


Fig. 9.1 Analog/discrete (digital) signal

Usually, nonphysical signals (generally called information) are inherently discrete in the form of numbers pertaining to intervals of time, e.g. monthly cash flow of a company. The information could also be a set of numbers pertaining to the value of a variable/parameter recorded (at usually uniform intervals) during say, an environmental study. Discrete signals, physical/nonphysical, are conventionally and speedily processed in digital form and so are also called *digital signals*. Digital signal could be processed independently disassociated from the original intervals called *off-line* processing, or may have to be processed, for example, in a digital control system within the constraint imposed by the time interval, called *on-line* processing.

Digital signals or numbers, in general, are processed by means of digital systems using the concept of binary numbers and Boolean algebra. A digital computer is a general-purpose digital item. The numbers are coded in the form of binary (ON/OFF) electrical pulses and processed by means of logic gates and memory cells. The logic gates and memory cells exploit the controlled switch behaviour of electronic devices. *Binary signals* are used extensively in communication, control and instrumentation systems as well in computers. Binary signals have the great advantage of being far less susceptible to disturbance (noise) compared to analog signals.

Even in large-scale digital systems, only a few different operations need to be performed, although these have to be repeated many times at a very fast speed. Logic, arithmetic and memory in conjunction with input, output devices/circuits are the five constituents of a digital system. In this chapter, we shall begin by studying the basic logic gates and their electronic realisation. Appropriate concepts of binary numbers and *Boolean algebra* will be introduced simultaneously. The building blocks of four major technologies—Transistor-Transistor Logic (TTL), Emitter-Coupled Logic (ECL), NMOS and CMOS. We will restrict ourselves to logic gates and simple combinational circuits.

9.1.1 Frequency Spectrum of the Signal

To fabricate electromagnetic circuits, information regarding the frequency content of signals involved is essential. As any signal having high frequency components may, under certain conditions, result in intersignal interference. In order to obtain a signal in the frequency domain, Fourier transforms are implied. This frequency domain representation is called signal spectrum. The advantages of using frequency domain are the easiness of synthesis and analyses of signal behaviour in a particular environment, especially in the presence of multiple signals. Fig. 9.2(a) shows three sine waves of different amplitudes and frequencies. Their frequency domain representation as obtained using Fourier transform is shown in Fig. 9.2(b) given below.

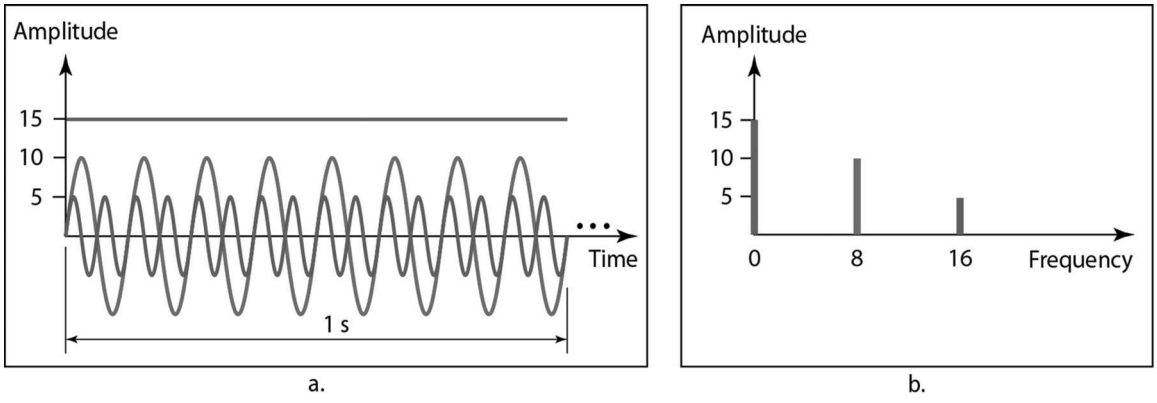


Fig. 9.2 (a) Time-domain representation of three sine waves with frequencies 0, 8, and 16, (b) Frequency domain representation of the same three signals

It has applications in image processing, signal processing, and data communications, where we need to send a composite signal, i.e. a signal made of many simple sine waves.

9.2 SWITCHING AND LOGIC LEVELS

Consider the simple circuit of Fig. 9.3 with switch *S*. The switch can be in two distinct states—OFF/ON. With *S* in OFF state, the output voltage is V_o = 5 V (HIGH) and with *S* in ON state, V_o = 0 V (LOW). In digital circuits, variables which can acquire only two values are indicated as '0' and '1' invoking a branch of algebra known as *Boolean algebra* and these are known as *Boolean* or *binary variables*. We can identify OFF as '0' and ON as '1' at the input and LOW as '0' and HIGH as '1' at the output. We immediately see that input '0' results in output '1' and vice versa, i.e. the circuit (also called switching circuit) of Fig. 9.3 accomplishes state inversion. Such circuits are known as *logic gates*. This particular circuit is an inverter or a NOT gate. It is symbolically represented in Fig. 9.4(a) with the table of input-output relationship called *truth table* as in Fig. 9.4(b), and input-output waveform relationship as in Fig. 9.4(c).

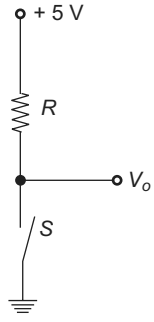
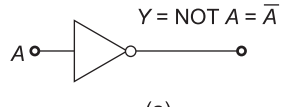


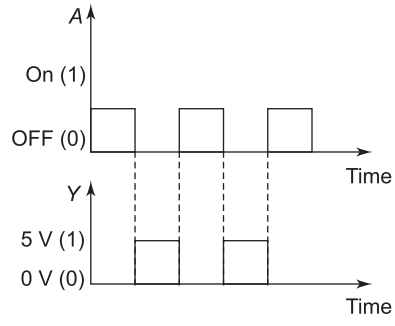
Fig. 9.3 Switching circuit



(a)

A	Y
1	0
0	1

(b) Truth Table



(c) Input/Output waveforms

Fig. 9.4 Inverter or NOT gate

In Boolean algebra, output is expressed as

$$Y = \text{NOT } A = \overline{A} \text{ or complement of } A \tag{9.1}$$

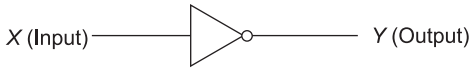
A is known as a Boolean variable.

It can easily be recognised by a truth table that $\overline{\overline{A}} = A$.

In electronic switching circuits, 0 and 1 states are represented by voltage levels which are indeed a range of values with a clear margin between the high end of LOW and low end of HIGH. For example, if 0 is normally represented as 0 V and 1 as 5 V, their voltage ranges may be 0 to 0.3 V and 4.7 to 5 V with a margin of 4.4 V(4.7 – 0.3 = 4.4 V).

9.2.1 Digital Logic Inverter

Inverter is a device performing the basic logical function called inversion, i.e. when the logic 0 is applied as an input, the output will be logic 1, and vice-versa. Figure 9.5 given below shows the logic inverter.



(a)

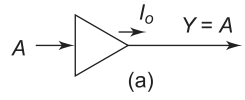
Input	Output
A	NOT A
0	1
1	0

(b)

Fig. 9.5 (a) Inverter Symbol (b) Inverter truth table

✦ Buffer

When a larger current I_o has to the fed to a load than what can be provided by a gate, a buffer is connected between the gate and load. Its symbol and truth table are shown in Fig. 9.6.



(a)

A	Y
1	1
0	0

(b)

Fig. 9.6 (a) Buffer symbol (b) Buffer truth table

9.3 DIGITAL WAVEFORM

The digital waveforms logic levels shown in Fig. 9.7 are ideal where 5 V = high = 1 and 0 V = low = 0. Also the change over from H(1) to L(0) and vice versa is in zero time.

Actual Waveform (Voltage Level)

In high output, the digital circuit acts as a source which feeds the load. Therefore,

$$V_o(H) < 5 \text{ V (say)}$$

It can have minimum value

$$V_{oH\,\min} = 3.5 \text{ V (say)}$$

This is to be recognised that $H(1)$

$$V_o = +5 \text{ V to } 3.5 \text{ V}$$

In low state, the digital circuit acts as a sink receiving current from the source. So,

$$V_o(L) > 0$$

Its maximum value to be recognised as zero is

$$V_o = 0 \text{ to } V_{oL\max} = 0.2 \text{ V (say)}$$

The forbidden region is then

$$(V_{oH\min} - V_{oL\max}) = 3.5 - 0.2 = 3.3 \text{ V}$$

where the signal cannot be recognised as $H(1)$ or $L(0)$.

The above conclusions are illustrated in Fig. 9.7.

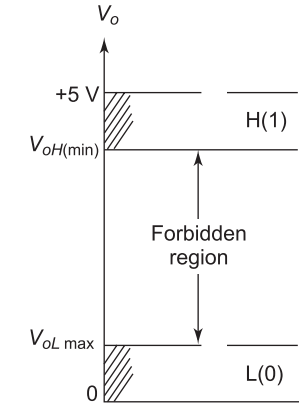


Fig. 9.7 Logic level profile

9.4 CHARACTERISATION OF DIGITAL ICs

The various characteristics of digital ICs that can be used to compare their performances are

1. Speed of operation
2. Power dissipation
3. Current and voltage parameter
4. Noise immunity
5. Fan-out

✦ Speed of Operation

A pulse signal in getting through an inverter is not an exact inverted replica. Further, it gets delayed in time called propagation delay which includes the switching of the transistor. This delay is once again illustrated by somewhat idealised input/output waveforms of an inverter. Typical gate delays are 1–20 nanoseconds. This delay limits the frequency of the input signal, and the cycle time must be larger than the propagation delay.

✦ Power Dissipation

The output voltage and current waveforms of a logic gate are shown in Fig. 9.8. As neither $V(0)$ nor I_{OFF} are zero, there is a power consumption under static condition (0 or 1). Also there is a power consumption during the switching intervals $T_1 < t < T_2$ and $T_3 \le t \le T_4$ which is the dynamic power. The dynamic power rises to a peak value and then falls off.

Both static and dynamic dissipation contribute to the total power consumption of a gate. Sometimes static dissipation is a dominant factor. In VLSI with CMOS fabrication technology, the dynamic power dissipation is the dominant factor as static dissipation is low.

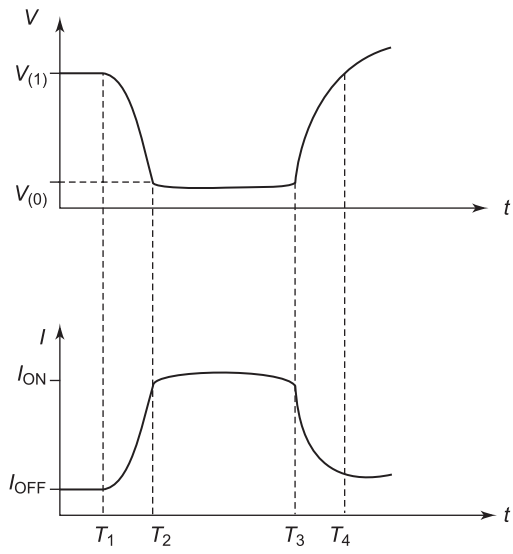


Fig. 9.8 Voltage and current waveforms

✦ Noise Immunity

This is essentially a measure of how much stray noise voltage the device can handle without giving any error at the output level.

✦ Fan-out

It denotes the number of gates that a gate in HIGH output state can feed without drop in its output voltage more than the allowable noise margin $(NM)_H$.

✦ Fan-in

It is the number of gates that can be allowed to sink their current into feeder gate in low state.

9.5 | LOGIC GATES

Logic gate is an electronic circuit, which accepts a binary input and produces a binary output, namely 0 and 1. The inverter (NOT) logic gate in general accepts one or more inputs and produces one output. Apart from the NOT gate considered in Section 9.2, it has one input and one output, but a logic gate in general accepts one or more inputs and produces one output. There are six other types of logic gates.

Input to a gate will be designated by binary variables A, B, C, etc. and the output will be indicated by the binary variable Y. As stated earlier, binary variables can take on values 0 and 1 which are electronically represented by LOW and HIGH voltage levels. In terms of Boolean algebra, the function of a logic gate will be represented by a binary expression. Boolean algebra will be dealt with in detail in Chapter 11.

9.5.1 AND Gate

Consider the diode circuit of Fig. 9.9. Assuming the diodes to be ideal, it is immediately obvious that the output will be HIGH only if both the inputs are HIGH and so both diodes are OFF. Such a logic operation is called AND and is represented by the Boolean expression

$$Y = A \text{ AND } B = A \bullet B \tag{9.2}$$

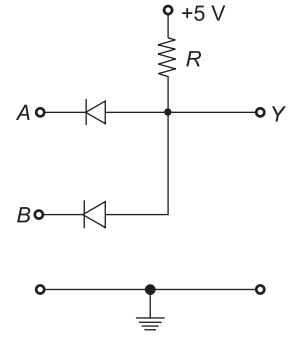
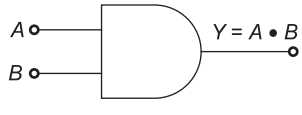


Fig. 9.9 An AND gate

Here dot indicates ANDing. The logic symbol for an AND gate is drawn in Fig. 9.10(a) with its truth table is given in Fig. 9.10(b).

Additional information that can be drawn from the truth table are the Boolean relations

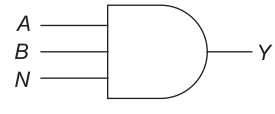
$A \bullet 0 = 0;$ $B = 0$
$A \bullet 1 = A;$ $B = 1$
$A \bullet A = A;$ $B = A$



(a) Symbol

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(b) Truth table



(c) AND gate with N inputs

Fig. 9.10 AND gate

In general, an AND gate can accept more than one input, which is symbolically represented in Fig. 9.10(c) and whose Boolean expression is

$$Y = A \cdot B \cdot C \cdot \cdots \cdot N \tag{9.3}$$

Note: In an AND gate, output is 1 only when all inputs are 1(HIGH).

9.5.2 OR Gate

Consider the diode circuit of Fig. 9.11. The output will be HIGH, if A or B (or both) are HIGH. Such a logic operation is called OR and is expressed by the Boolean expression

$$Y = A \text{ OR } B = A + B$$

The symbolic representations of OR and its truth table are given in Figs. 9.12(a) and (b). It is evident from the truth table that

$$A + 0 = A;\ B = 0 \tag{9.4}$$
$$A + 1 = 1;\ B = 1$$
$$A + A = A;\ B = A$$

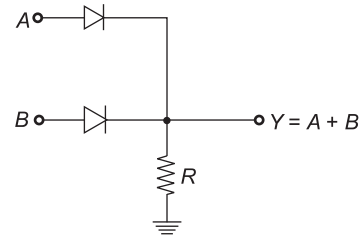
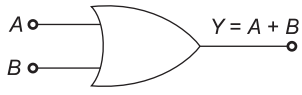


Fig. 9.11 An OR gate

An *N*-input OR gate is represented in Fig. 9.12(c) whose Boolean expression is

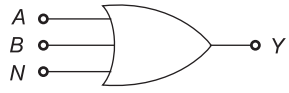
$$Y = A + B + C + \cdots + N \tag{9.5}$$



(a) Symbol

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

(b) Truth table



(c) OR gate with *N* inputs

Fig. 9.12 OR gate

Note: Output of OR gate is 0 only when both (all) inputs are 0 (low).

✦ Bubbled (Small Circle) Notation

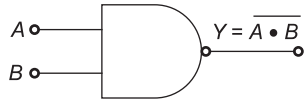
A bubble at the output and/or input of a gate means *inversion*. This notation has already been used in NOT gate (inverter).

9.5.3 NAND Gate

It is an AND gate followed by a NOT gate represented by the symbol of Fig. 9.13(a), wherein the small circle (bubble) indicates NOT operation. The NAND truth table is shown in Fig. 9.13(b). Its Boolean expression is

Y = NOT (A AND B) = $\overline{A \cdot B}$; complement of AND (9.6)

A NAND gate could have more than two inputs.



(a) Symbol

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

(b) Truth table

Fig. 9.13 NAND gate

Note: The output is 0 only when both inputs are 1 (High).

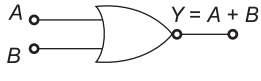
9.5.4 NOR Gate

Logically, a NOR is expressed as

Y = NOT (A OR B) = $\overline{A + B}$; complement of OR (9.7)

The symbolic representation of NOR gate and its truth table are shown in Fig. 9.14(a) and (b). There could be more than two inputs.

Note: Output is 1 only when both inputs are 0 (Low).



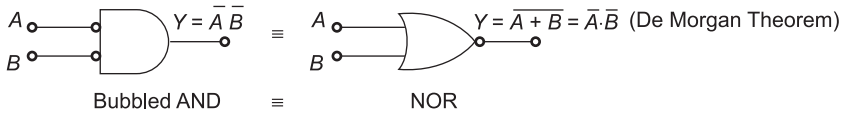
(a) Symbol

A	B	Y
0	0	1
1	0	0
0	1	0
1	1	0

(b) Truth table

Fig. 9.14 NOR gate

9.5.5 Bubbled AND Gate



Truth table proof

A	B	$\overline{A}$	$\overline{B}$	$\overline{A}\cdot\overline{B}$	$A + B$	$\overline{A + B}$
0	0	1	1	1	0	1
0	1	1	0	0	1	0
1	0	0	1	0	1	0
1	1	0	0	0	1	0

Similarly

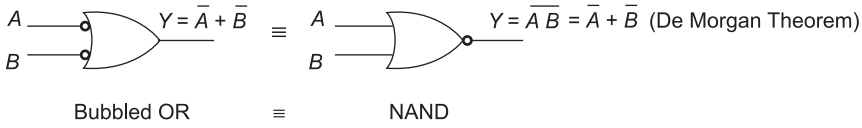


Fig. 9.15

The equalities shown there follow from the application of *De Morgan theorems:*

1. $$Y = \overline{A + B} = \overline{A}\,\overline{B} \tag{9.8}$$

2. $$Y = \overline{AB} = \overline{A} + \overline{B} \tag{9.9}$$

These theorems are proved in truth table of Fig. 9.15.

✦ Duality Theorem

From a Boolean Identity another identity (Dual) can be obtained by the following step:

- Change AND to OR and OR to AND
  $\bullet \rightarrow +$ $\quad + \rightarrow \bullet$
- Change 1 to 0 and 0 to 1 (complement)
  $\bullet \rightarrow +$ $\quad + \rightarrow \bullet$

Example

1. $A + 1 = 1$
   New identity dual
   $A \cdot 0 = 0$
2. Also
   $A \cdot 1 = A$
   Then $A + 0 = A$
3. $A(B + C) = AB + AC$
   Dual $A + BC = (A + B) \cdot (A + C)$
4. Further change each NAND to NOR and NOR to NAND complement input and output.

9.5.6 Universality of NAND/NOR Gates

NAND and NOR gates are called universal logic gates, because any logic operation can be realised by using these gates. This is illustrated by some examples.

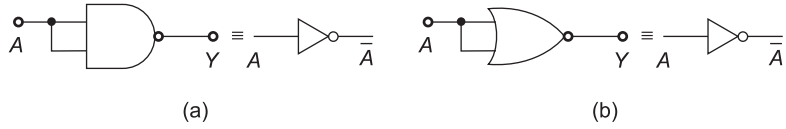
Consider the NAND gate of Fig. 9.16(a) with input A at both ports. Then

$$Y = \overline{A \cdot A} = \overline{A} = \text{NOT } A$$

Similarly for the NOR gate of Fig. 9.16(b),

$$Y = \overline{A + A} = \overline{A} = \text{NOT } A$$

The reader should verify these by writing the truth tables.



(a) (b)

Fig. 9.16 NOT gate realisation by NAND/NOR

9.5.7 OR Realisation by NAND/NOR

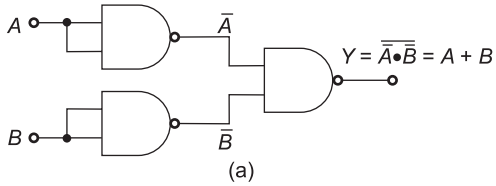
Consider the logic circuit of Fig. 9.17(a) composed of NANDs. Here,

$$Y = \overline{\overline{A} \cdot \overline{B}}$$

Its truth table is given in Fig. 9.17(b). It immediately follows that[1]

$$Y = \overline{\overline{A} \cdot \overline{B}} = A + B = A \text{ OR } B$$

Thus, the logic circuit of Fig. 9.17(a) is an OR realisation by NAND gates.



(a)

A	$\overline{A}$	B	$\overline{B}$	$\overline{A} \cdot \overline{B}$	$\overline{\overline{A} \cdot \overline{B}}$	$A+B$
0	1	0	1	1	0	0
0	1	1	0	0	1	1
1	0	0	1	0	1	1
1	0	1	0	0	1	1

(b)

Fig. 9.17 OR realisation by NANDs

OR realisation by NOR gates is given in Fig. 9.18 whose operation is obvious.

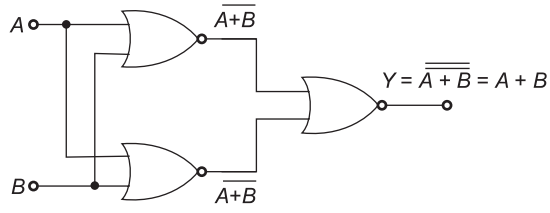


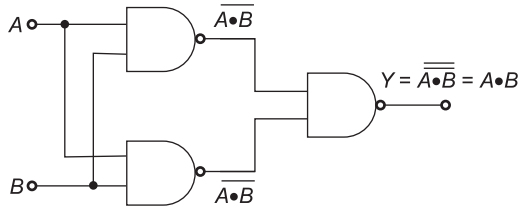
Fig. 9.18 NOR realisation of OR

1. This indeed is application of De Morgan's theorems.

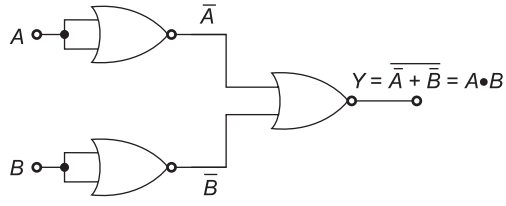
EXAMPLE 9.1 Consider the logic circuits of NAND/NORs of Figs 9.19(a) and (b) respectively. Find the output in each case.

Solution NAND circuit of Fig. 9.19(a) and NOR circuit of Fig. 9.19(b) are two realisations of AND. While the operation is obvious for the NAND circuit of Fig. 9.19(a), for the NOR circuit of Fig. 9.19(b), the AND operation is proved in the truth table of Fig. 9.19(c).

Observe that circuits of NAND realisation of OR [Fig. 9.19(a)] and NOR realisation of AND [Fig. 9.19(b)] have the same structure; vice versa also applies. See Figs 9.18 and 9.19(a). The truth table is given in Fig. 9.19(c).



(a) NAND realisation of AND



(b) NOR realisation of AND

A	$\overline{A}$	B	$\overline{B}$	$\overline{A} + \overline{B}$	$\overline{\overline{A} + \overline{B}}$	$A \cdot B$
0	1	0	1	1	0	0
0	1	1	0	1	0	0
1	0	0	1	1	0	0
1	0	1	0	0	1	1

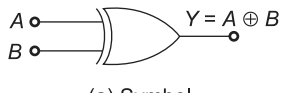
(c) Truth table of the circuit

Fig. 9.19

9.5.8 Exclusive OR (EXOR) Gate

The output of this gate is high only when either of the inputs is high but not when both inputs are high, i.e. either, but not both. The gate symbol and truth table are shown in Figs 9.20(a) and (b) respectively. In terms of Boolean algebra,

$$Y = A \text{ XOR } B = A \oplus B \tag{9.10}$$



(a) Symbol

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

(b) Truth table

Fig. 9.20 EXOR gate

NAND implementation of Exclusive OR is shown in Fig. 9.21.

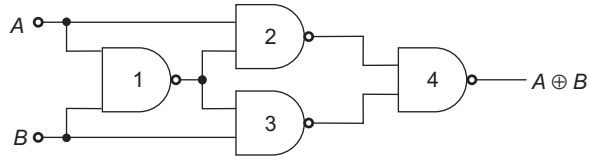


Fig. 9.21 NAND implementation of EXOR

Output gate 1 $\overline{AB}$

Output gate 2 $\overline{\overline{AB}A}$

Output gate 3 $\overline{\overline{AB}B}$

Output gate 4 $Y = \left[\overline{\overline{\overline{AB}A} \cdot \overline{\overline{AB}B}}\right]$

Applying De Morgan's theorem repeatedly,

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

or $$Y = \overline{AB} \cdot A + \overline{AB} \cdot B = \overline{A}B(A + B)$$

or $$Y = (\overline{A} + \overline{B})(A + B)$$

or $$Y = \overline{A}B + A\overline{B}$$

Its truth table is given below.

A	B	$\overline{A}$	$\overline{B}$	$\overline{A}B$	$A\overline{B}$	Y
0	0	1	1	0	0	0
0	1	1	0	1	0	1
1	0	0	1	0	1	1
1	1	0	0	0	0	0

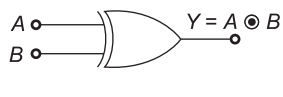
Thus,

$$Y = A \oplus B, \text{ exclusive } OR$$

Note: For multiple inputs, EXOR output will be 1 if the number of 1 inputs is odd.

9.5.9 Exclusive NOR Gate

It is the gate complementary to EXOR gate. Its symbol and truth table are shown in Figs 9.22(a) and (b) respectively. The output (Y) is 1 only when both inputs are same (0, 0) or (1, 1)



(a) Symbol

A	B	Y
0	0	1
0	1	0
0	1	0
1	1	1

(b) Truth table

Fig. 9.22 Exclusive NOR gate

9.5.10 AND-OR-Invert (AOI) Circuit

And-OR-Invert circuit is drawn in Fig. 9.23. Its output is

$$Y = \overline{AB + CD}$$

Each of the AND gates could have more than two inputs.

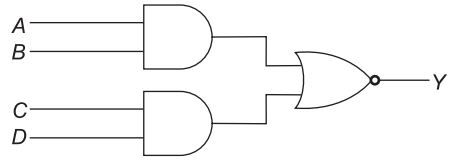


Fig. 9.23 AOI circuit

9.6 REPRESENTATION OF LOGIC FUNCTION

The logic variables and functions are often represented using binary representation, i.e. bit value 1 or 0. In digital designing of combinational circuits, one needs to have minimum Boolean expressions so as to reduce the circuit complexity and ease of design. The basic Boolean expressions are derived from the truth tables using following representations:

1. Sum of Product
2. Product of Sum

9.6.1 Sum of Product (SOP)

The Boolean expression represented by performing OR operation on different AND operations is called Sum of Product (SOP) expression. In simple words, the logic sum of two or more product terms, e.g. $A\overline{B} + \overline{A}\,B\overline{C} + AC$, is called SOP representation.

In order to illustrate the same, let us consider the truth table of XOR gate given as follows:

A	*B*	*Y* (Output)
0	0	0
0	1	1
1	0	1
1	1	0

The truth table may be represented by Boolean expression:

$$Y = \overline{A}\,B + A\overline{B} \tag{9.11}$$

By substituting all the values of A and B in this truth table, you will always get the output F conforming to the specific truth table. The product of all the variables in each row of the truth table is termed as minterm. So the above table has four minterms: $\overline{A}\,\overline{B}$, $\overline{A}\,B$, $A\overline{B}$, and AB, which are equal to the number of rows corresponding to m_0, m_1, m_2 and m_3. The minterms implied to output 1 are known as implicants. It is the sum of the implicants which provides the equation for the function of SOP for Eq. (9.11).

A truth table can also be specified as:

$$F(A, B) = \sum m\,(1, 2)$$

So if we have n variables in the input, it implies 2^n minterms as the output.

9.6.2 Product of Sum (POS)

The Boolean expression represented by performing AND operation on different OR operations is called Product of Sum (POS) expression. In simple words, the logic product of two or more summing terms, e.g. $(A + \overline{C})\cdot(B + C)\cdot(\overline{A} + B + C)$ is called POS representation.

The Boolean expression for the truth table stated above (of XOR operation) may also be written using POS as:

$$Y = (A + B)\,(\overline{A} + \overline{B})\,Q \tag{9.12}$$

The product of input variables in each row of the truth table is called minterm, and the sum of the complements of input variables in each row is called maxterm. In the given example, $(A + B)$, $(A + \overline{B})$, $(\overline{A} + B)$, and $(\overline{A} + \overline{B})$, are the maxterms. The number of maxterms in the truth table is equal to the number of the rows in the table. The function can be specified by taking the product of those maxterms which have the output 0. It must be noted that each maxterm can be obtained by complement of minterms and vice-versa.

Digital functions can be specified in the form of implied maxterms as follows:

$$F(A, B) = \prod M(0,3)$$

Similarly, n variables forming n OR terms with each variable being primed or unprimed, provide 2^n combinations called maxterms or standard sums, and vice-versa.

✦ Inhibit (Enable) Operation

Consideration a two-input AND gate with an additional *Strobe* (S) terminal through an inverter represented by a bubble.

The output

$$Y = AB\overline{S}$$

If $A = 1$, $B = 1$, output $Y = 1$ only if $S = 0$, the gate is enabled. However, if $S = 1$.

$$Y = AB\overline{S} = 0$$

The gate is inhibited.

The above applies to AND gate with any number of inputs.

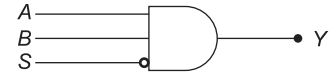


Fig. 9.24 AND gate with Strobe

✦ Two-input Gate

Input B through bubble (inverted);

$$Y = A\overline{B}$$

If $B = 1$, gate is inhibited and so

$$Y = 0$$

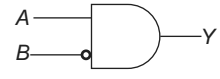


Fig. 9.25 Two-input gate

EXAMPLE 9.2 Realise the following gates using universal NAND gate combinations:

(a) NOT gate (b) AND gate (c) OR gate (d) NOR gate (e) XOR gate

Solution

(a) NOT gate using NAND

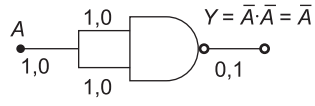


Fig. 9.26

(b) AND gate using NAND

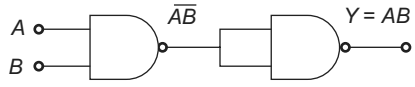


Fig. 9.27

(c) OR gate using NAND

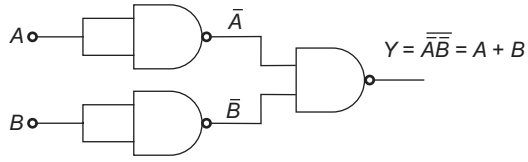


Fig. 9.28

(d) NOR gate using NANDs

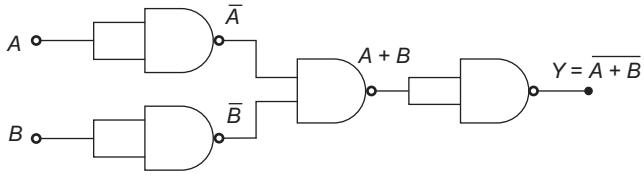


Fig. 9.29

(e) XOR gate using NAND

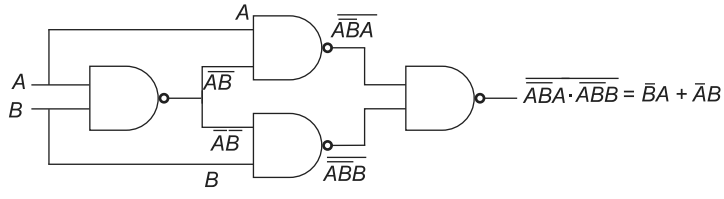


Fig. 9.30

EXAMPLE 9.3 Realise following gates using universal NOR gate combinations:

(a) NOT gate (b) OR gate (c) AND gate (d) NAND gate (e) XOR gate

Solution

(a) NOT gate using NOR

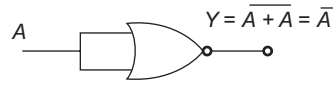


Fig. 9.31

(b) OR gate using NOR

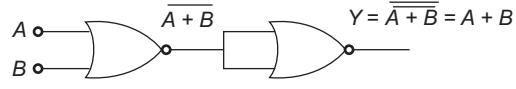


Fig. 9.32

(c) AND gate using NOR

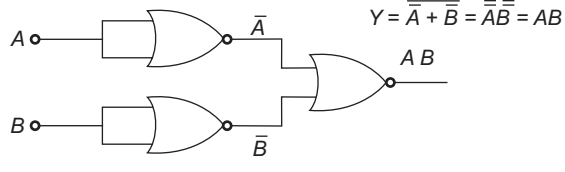


Fig. 9.33

(d) NAND gate using NOR

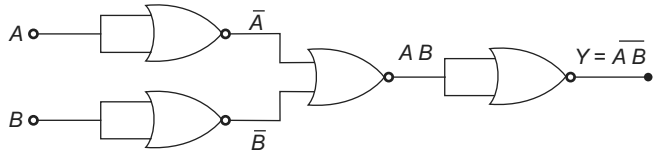


Fig. 9.34

(e) XOR gate using NOR

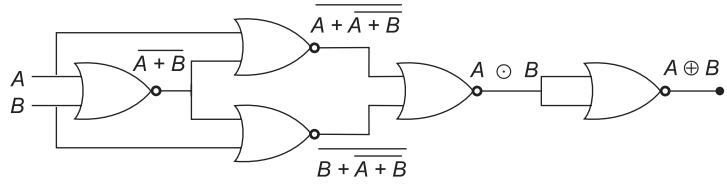


Fig. 9.35

EXAMPLE 9.4 Enlist the Basic Boolean laws.

Solution

1. Commutative law
   (a) $A + B = B + A$
   (b) $AB = BA$
2. Associative law
   (a) $(A + B) + C = A + (B + C)$
   (b) $(AB)C = A\,(BC)$
3. Distributive law
   (a) $A(B + C) = AB + AC$
   (b) $A + (BC) = (A + B)\,(A + C)$
4. Identity law
   (a) $A + A = A$
   (b) $AA = A$
5. Redundance law
   (a) $A + AB = A$
   (b) $A(A + B) = A$
6. Inverse law
   (a) $A + \overline{A} = 1$
   (b) $A \cdot \overline{A} = 0$

EXAMPLE 9.5 In brief, explain the characteristics of logic gates and mention the number for TTL.

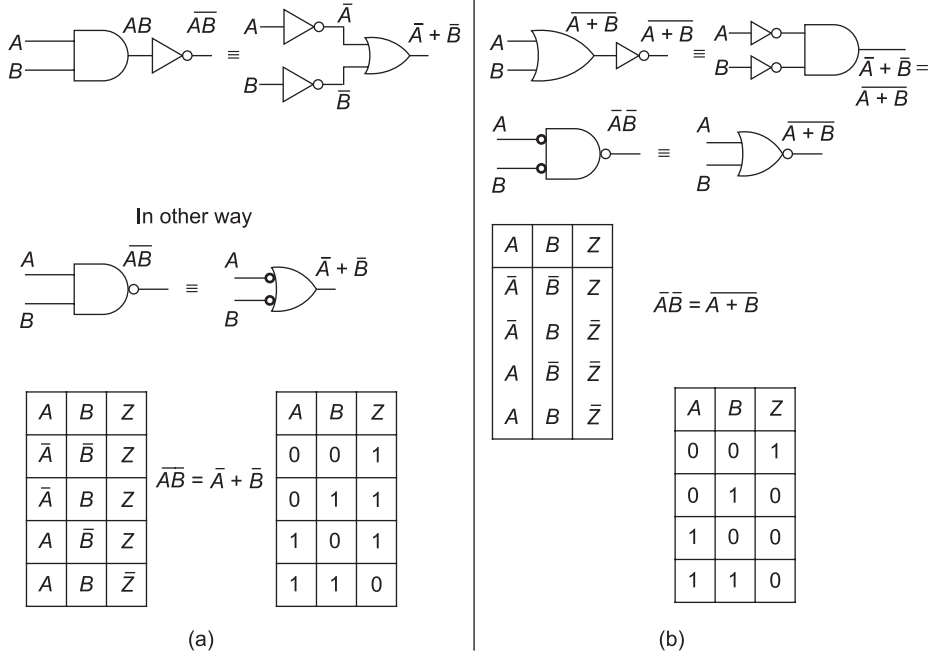
Solution

Fan-out: The measure of the maximum number of logic gates that can be driven by a single logic gate without affecting the specified operation characteristics of the driving gate is called fan-out. A standard fan-out for TTL is ten (10).

Fan-in: The measure of the maximum number of inputs that can be connected to a logic gate without affecting the specified operational characteristics of the driven logic gate is called fan-in. A standard fan-in for TTL is ten(10).

EXAMPLE 9.6 Prove De Morgan's theorem.

Solution Using basic logic gates, we prove (a) $\overline{AB} = \bar{A} + \bar{B}$ (b) $\overline{A + B} = \bar{A}\,\bar{B}$



A	B	Z
$\bar{A}$	$\bar{B}$	Z
$\bar{A}$	B	Z
A	$\bar{B}$	Z
A	B	$\bar{Z}$

$\overline{AB} = \bar{A} + \bar{B}$

A	B	Z
0	0	1
0	1	1
1	0	1
1	1	0

(a)

A	B	Z
$\bar{A}$	$\bar{B}$	Z
$\bar{A}$	B	$\bar{Z}$
A	$\bar{B}$	$\bar{Z}$
A	B	$\bar{Z}$

$\bar{A}\,\bar{B} = \overline{A + B}$

A	B	Z
0	0	1
0	1	0
1	0	0
1	1	0

(b)

Fig. 9.36

EXAMPLE 9.7 Simplify the following Boolean expressions.

(a) $A + AC$

$A + AC = A\,(1 + C)$

$= A\,1$

$= A$

(b) $A + \bar{A}B$

$A + \bar{A}B = (A + \bar{A})(A + B)$

$= A + AB + \bar{A}B = A + B(A + \bar{A})$

$= (A + B)$

(c) $\boldsymbol{A + \overline{A}\,B + AB\overline{C}}$

$$\begin{aligned} A + \overline{A}\,B + AB\overline{C} &= A(1 + B\overline{C}) + \overline{A}\,B \\ &= A1 + \overline{A}\,B \;= A \cdot 1 + \overline{A} \cdot B \\ &= A + \overline{A}\,B \\ &= (A + \overline{A})\,(A + B) \\ &= 1(A + B) \\ &= (A + B) \end{aligned}$$

(d) $\boldsymbol{A + \overline{A}\,B + ABC + A\overline{C}}$

$$\begin{aligned} A + \overline{A}\,B + ABC + A\overline{C} &= (A + \overline{A}) \cdot (A + B) + ABC + A\overline{C} \\ &= 1 \cdot (A + B) + A(BC + \overline{C}) \\ &= A + B + A(B + \overline{C})\,(C + \overline{C}) \\ &= A + B + AB + A\overline{C} = A(1 + B + \overline{C}) + B = A + B \end{aligned}$$

(e) $\boldsymbol{\overline{A}C + \overline{AC}}$

$$\begin{aligned} \overline{A}C + \overline{AC} &= \overline{A}C + \overline{A} + \overline{C} \\ &= \overline{A}(C + 1) + \overline{C} \\ &= \overline{A} + \overline{C} \end{aligned}$$

(f) $\boldsymbol{(B + \overline{C})(\overline{B} + C) + \overline{\overline{A} + B + \overline{C}}}$

$$\begin{aligned} (B + \overline{C})(\overline{B} + C) + \overline{\overline{A} + B + \overline{C}} &= B\overline{B} + \overline{B}\,\overline{C} + C\overline{C} + BC + A\overline{B}C \\ &= \overline{B}\,\overline{C} + BC + A\overline{B}C \\ &= \overline{B}\,\overline{C} + (B + A\overline{B})C \\ &= \overline{B}\,\overline{C} + (B + \overline{B})\,(B + A)C \\ &= \overline{B}\,\overline{C} + (B + A)C \\ &= \overline{B}\,\overline{C} + BC + AC \end{aligned}$$

(g) $\boldsymbol{\overline{\overline{A + B} + \overline{C}}}$

$$\begin{aligned} \overline{\overline{A + B} + \overline{C}} &= (A + B)C \\ &= AC + BC \end{aligned}$$

(h) $\boldsymbol{\overline{AB + \overline{A\overline{B}} + A}}$

$$\overline{AB + \overline{A\overline{B}} + A} = \overline{A\overline{B}} \cdot AB \cdot \overline{A}$$

$$= \overline{AB} \cdot 0$$
$$= 0$$

(i) $\boldsymbol{AB + \bar{A} + \overline{AB}}$

$$AB + \bar{A} + \overline{AB} = AB + \bar{A} + \bar{A} + \bar{B}$$
$$= AB + \bar{A} + \bar{B}$$
$$= AB + \overline{AB}$$
$$= 1$$

(j) $\boldsymbol{\overline{(\bar{A} + C)(B + \bar{D})}}$

$$\overline{(\bar{A} + C)(B + \bar{D})} = \overline{(\bar{A} + C)} + \overline{(B + \bar{D})} = (\bar{\bar{A}} \cdot \bar{C}) + (\bar{B} \cdot \bar{\bar{D}}) = (A \cdot \bar{C}) + \bar{B} \cdot D$$
$$= A\bar{C} + \bar{B}D$$

EXAMPLE 9.8 Prove the following equalities.

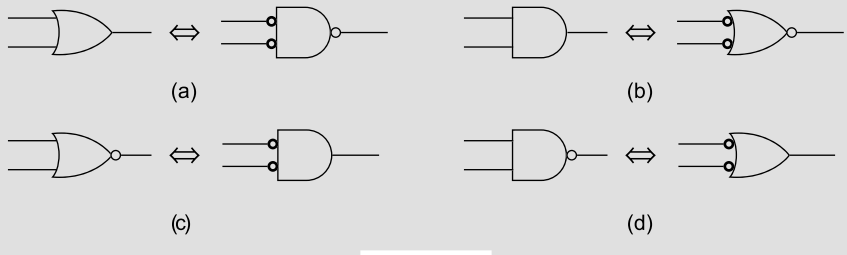


Fig. 9.37

Solution Let inputs be A, B and output be Y. We will use De Morgan's theorems.

Left = Right

(a) $Y = A + B = \overline{\bar{A} \cdot \bar{B}} = A + B$

(b) $Y = AB = \overline{\bar{A} + \bar{B}} = AB$

(c) $Y = \overline{A + B} = \bar{A}\bar{B}$

(d) $Y = \overline{AB} = \bar{A} + \bar{B}$

EXAMPLE 9.9 For the NAND-NAND circuit of Fig. 9.38, write the truth table given if the change from each step of the truth table needs 1 μs time; draw the timing diagram.

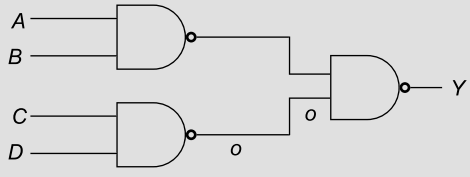


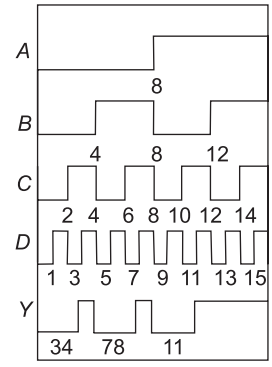
Fig. 9.38

Solution In a NAND, output is zero only if both inputs are 1. The truth table is given in Fig. 9.39(a) and timing diagram in Fig. 9.39(b). It is to be noted that the inputs change in synchronism at each circuit time.

$$Y = \overline{\overline{AB}\cdot\overline{CD}} = AB + CD$$

A	B	C	D	Y
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	1
1	0	0	0	0
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1

(a) Truth table



(b) Timing diagram

Fig. 9.39

EXAMPLE 9.10 Convert the following circuit in a Boolean circuit.

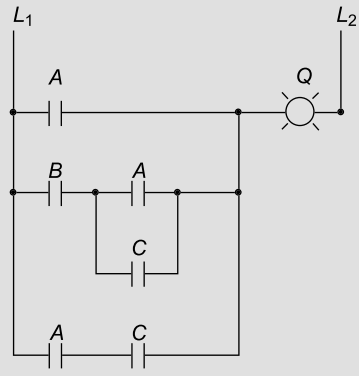


Fig. 9.40

Solution The resistive circuit for the same is shown below in Fig. 9.41:

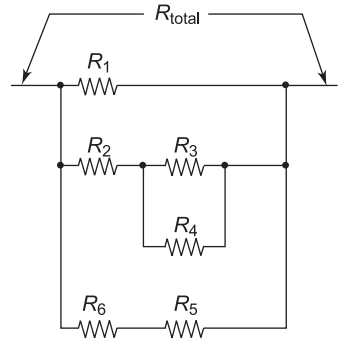


Fig. 9.41

The total resistance of the circuit is, $R = R_1 \parallel [(R_3 \parallel R_4 - R_2)] \parallel (R_5 - R_6)]$. In the above circuit, the parallel contacts represent the Boolean addition and series contacts represent the Boolean multiplication. The Boolean expression for the circuit is shown below.

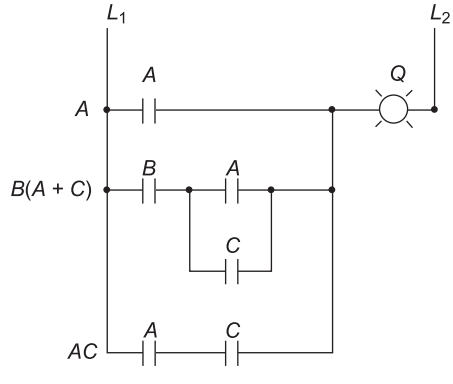


Fig. 9.42

The minimization of Boolean expression is as follows:

$A + B(A + C) + AC$

↓ Distributed terms

$A + AB + BC + AC$

↓ Applying rule $A + AB = A$ to 1st and 2nd terms

$A + BC + AC$

↓ Applying rule $A + AB = A$ to 1st and 3rd terms

$A + BC$

Therefore, $Q = A + BC$

EXAMPLE 9.11 Calculate SOP and POS representation of Boolean function $F(A, B, C, D) = S(1, 3, 4, 5, 6, 12, 13, 14)$.

Solution

$F(A, B, C, D) = S(1, 3, 4, 5, 6, 12, 13, 14)$

AB \ CD	C′D′	C′D	CD	CD′
A′B′	0	1	1	0
A′B	1	1	0	1
AB	1	1	0	1
$\overline{AB}$	0	0	0	0

(i) Sum of Products (grouping of 1's)

Make a group of two by grouping $m1$ and $m3 = A'B'D$. Make a group of four by grouping $m4$, $m5$, $m12$, and $m13 = BC'$. Make a group of four by grouping $m4$, $m12$, $m6$, and $m14 = BD'$. Read the terms from each group and sum them $F(A,B,C,D) = A'B'D + BC' + BD'$.

(ii) Product of Sums

Step 1 Find F' by grouping 0's, expressing 0's (not 1'). Make a group of four by grouping $m0$, $m2$, $m8$, and $m10 = B'D'$. Make a group of four by grouping $m8$, $m9$, $m11$, and $m10 = AB'$. Make a group of two by grouping $m7$ and $m15 = BCD$. Read the terms from each group and sum them $F'(A, B, C, D) = B'D' + AB' + BCD$.

Step 2 **Applying De Morgan's theorem**

$(F')' = (B'D' + AB' + BCD)'$ $F(A, B, C, D)$

$= (B + D)(A' + B)(B' + C' + D')$

SUMMARY

- Analog and digital signals have been introduced. Digital ICs are discussed.
- Resumption and functional operation of logic gates is presented.

EXERCISES

Review Questions

1. How can you convert octal numbers to binary and vice versa?
2. How can you convert hexadecimal to binary and vice versa?
3. Write the basic rules for addition and subtraction of binary numbers.
4. Take any two 8-bit binary numbers. Illustrate how to add and subtract the two.
5. What are signed numbers? What is their range? How do they compare with unsigned numbers?
6. What is 2's complement of a negatives number? How is it found?
7. How is 2's complement in subtraction? If the answer is negative, how can you find its magnitude?
8. State De Morgan's laws.
9. What is the ASCII Code?
10. What is BCD?
11. What is a BCO to decimal converter? What gates are used at the output end?
12. What is the input and output of a half adder?
13. Draw the circuit for a full adder using two half adders.
14. In a full adder, the sum is obtained as the output of an XOR gate. Show how.
15. What is the input and output of a full adder?
16. Distinguish between analog and digital signals.
17. Draw an ideal digital waveform. Superimpose on it the actual wave. What is its rise and fall time?
18. Sketch the voltage level of a digital signal, which determines high and low.
19. When is the output of an AND gate 1?
20. When is the output of an OR gate 0?
21. Draw the symbol of an XOR gate and write its truth table.
22. Show that a bubbled AND gate is equivalent to NOR gate.
23. What is the meaning of universality of NAND and NOR gates?
24. How can you realise XOR gate by NAND gates?
25. Draw the symbol and write the truth table of the exclusive NOR gate.
26. Define canonical forms.

Problems

1. Draw the logic circuit whose Boolean equation is

   $Y = \overline{A + B} + \overline{C}$

2. Draw the logic circuit described by

   $Y = (\overline{ABC})\overline{D}$

3. Draw the logic circuit given by the Boolean equation

   $Y = \overline{A}BC + A\overline{B}C + \overline{A}B\overline{C}$

You may use triple-input ANDs.

4. Write the truth table for Fig. 9.43

5. In Fig. 9.44, determine Y if (i) both switches are open, and (ii) both switches are closed.

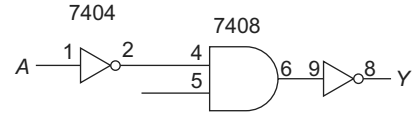


Fig. 9.43

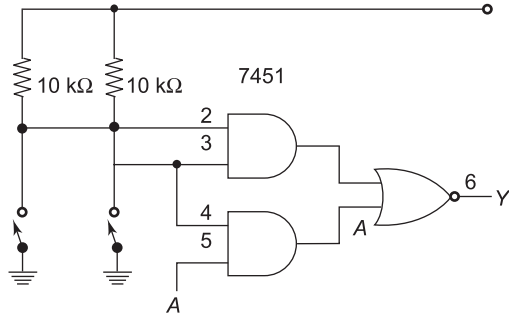
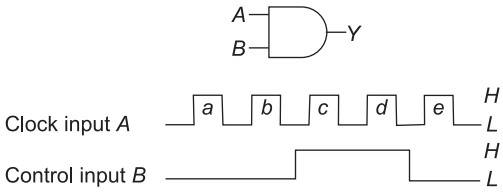
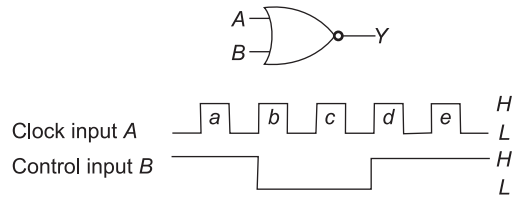


Fig. 9.44

6. For Figs 9.45(a) and (b), sketch the output (Y) waveforms.



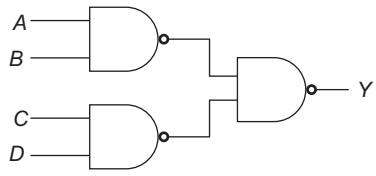
(a) Pulse train problem.



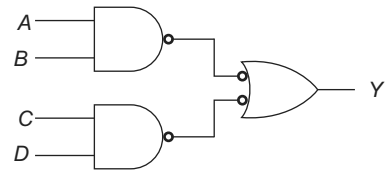
(b) Pulse train problem.

Fig. 9.45

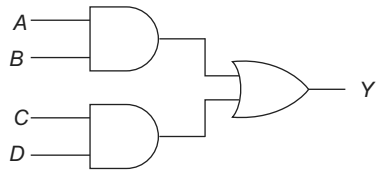
7. Prove that the figures below are logically equivalent:



(a)



(b)



(c)

Fig. 9.46

8. Assume that the switch in Fig. 9.47 is controlled by a voltage, such that $V_i(0)$ makes the switch open and $V(1)$ closes it. Let the switch have an ON resistance of 20 Ω. Find the value of R that ensures $V_o(0) > 4$ V and $V_o(1) < 0.5$ V.
9. The waveforms shown in Fig. 9.48 are applied to the logic circuit shown therein. Plot the output waveform.

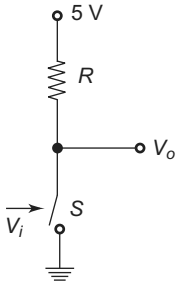


Fig. 9.47

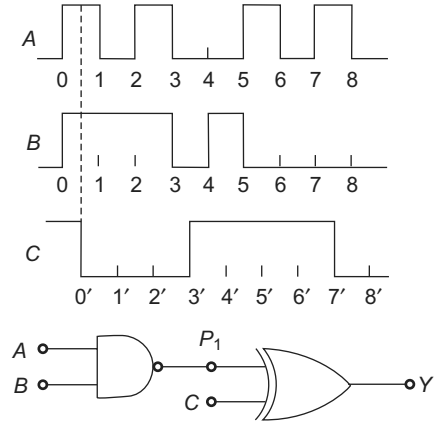


Fig. 9.48

10. Sketch an EXOR gate. Connect its input terminals to High(1). Show that it gets on as an inverter.
11. What is the output of the circuit Q?

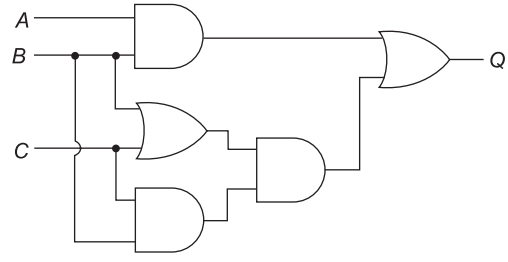


Fig. 9.49

12. Using the truth table given below, define SOP and POS.

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

13. The standard SOP form of the expression $A'B + A'C$ is …….

➥ Multiple-Choice Questions

1. The universal gates are
(a) NAND gate (b) AND gate (c) NOR gate (d) EX-OR gate

2. The output of a two-input EX-OR gate is
(a) $\overline{a}b + a\overline{b}$ (b) $\overline{ab + bc}$ (c) $\overline{a}\,\overline{b} + ab$ (d) none of these

3. How many NAND gates will require to implement an AND gate?
(a) 2 (b) 3 (c) 4 (d) 5

4. The equation $\overline{a \cdot b}$ represents
(a) X-OR gate (b) X-NOR gate (c) NAND gate (d) NOR gate

5. NAND and NOR are called universal gates because
(a) they have least propagation delay
(b) they are available every where
(c) all the other gates can be implemented using these gates
(d) they are easy to process

6. A three-variable input can have how many compilations of input?
(a) 8 (b) 9 (c) 3 (d) None of these

7. According to commutative law,
(a) $A + B = B + A$ (b) $AB = BA$ (c) $\overline{A + B} = \overline{A} \cdot \overline{B}$ (d) both (a) and (b)

8. The fan-in par standard TTL is
(a) 100 (b) 10 (c) 1 (d) 12

9. $A + \overline{A}B + ABC + \overline{A}C$ can be reduced as
(a) $A + B + C$ (b) $\overline{A} + B + C$ (c) $A + B + \overline{C}$ (d) none of these

10. $AB + \overline{A} + A\overline{B}$ is
(a) 1 (b) 0 (c) A (d) $\overline{A}$

11. Which property of delta function indicates the equality between the area under product of function with shifted impulse and the value of function located at unit impulse instant?
(a) Replication (b) Sampling (c) Scaling (d) Product

12. Which mathematical notation specifies the condition of periodicity for a continuous time signal?
(a) $x(t) = x(t + T_0)$ (b) $x(n) = x(n + N)$ (c) $x(t) = e^{-\alpha t}$

13. _____ signals can have an infinite number of values in a range.
(a) Analog (b) Digital (c) (a) or (b) (d) None of these

14. _____ describes the position of the waveform relative to time 0.
(a) Frequency and phase (b) Phase
(c) Frequency and amplitude (d) None of these

15. For the SOP expression $AC + BC + C$, the number of 1s in the output are
(a) 1 (b) 2 (c) 4 (d) 3

16. If $A' + B + C' + D = 0$, the values of A, B, C, and D will be
(a) $A = 1, B = 0, C = 0, D = 0$ (b) $A = 1, B = 0, C = 1, D = 0$
(c) $A = 0, B = 1, C = 0, D = 0$ (d) $A = 1, B = 0, C = 1, D = 1$

➥ Fill in the Blanks

1. In an integrated circuit, all active and passive components are ……..on a single chip of …….
2. In thin film ICs, ……. components are added externally.

➥ True/False

1. In the current scenario, only silicon is used to fabricate ICs.
2. An OP-AMP represents a linear IC.

ANSWERS

✦ Problems

4. (i) $\bar{A}$ (ii) 1

11. $AB + BC\,(B + C)$

12. SOP: $S = (A'.B) + (A.B')$; $C = A.B$
POS: $S = (A + B)\,.\,(A' + B')$; $C = (A + B)\,.\,(A + B')\,.\,(A' + B)$

13. $A'B'C + A'BC' + A'BC$

✦ Multiple-Choice Questions

1. (a) and (c) **2.** (a) **3.** (b) **4.** (a) **5.** (c) **6.** (a) **7.** (d) **8.** (b) **9.** (a)
10. (a) **11.** (b) **12.** (a) **13.** (a) **14.** (b) **15.** (d) **16.** (b)

✦ Fill in the Blanks

1. fabricated, ??? **2.** active

✦ True/False

1. True **2.** True

CHAPTER 10

Digital Electronics

GOALS AND OBJECTIVES

- ❑ Introduction to number systems—decimal number, binary number, hexadecimal number and complement of binary number system
- ❑ Introduction and explanation of binary number system and number conversion
- ❑ Conversion of octal number system—octal to binary and binary to octal conversion
- ❑ Conversion of hexadecimal number system—hexadecimal to binary and binary to hexadecimal
- ❑ Simplification of complex Boolean expressions
- ❑ Analyze the relation of Boolean expressions to truth tables and logic diagrams
- ❑ Explanation of Boolean algebra theorems and their operations
- ❑ Classification of digital circuits—logic gates, combinational circuits, multiplexers and decoders
- ❑ Introduction of sequential circuits and general model of sequential circuits

10.1 INTRODUCTION

The term *digital* refers to any process that is accomplished using discrete units. Each of these could be used as a unit or group of units to express a whole number. In contrast with digital numbers, analog numbers are represented as directly measurable quantities such as volts, speed and distance. It means analog numbers represent real things. Both the analog and digital methods are used for computing. The abacus was the first computing device; the word "*abacus*" means making *marks in the dust*. From the beginning, the Greeks and Romans developed a system of counting involving beads on a marble table. The beads were grouped for units, tens, hundreds, and so on. Since then, we have come a long way. Today computers have been developed which can perform millions of operations per second.

10.2 NUMBER SYSTEMS

There are four number systems of arithmetic that are used in the digital systems:

1. Decimal
2. Binary
3. Hexadecimal
4. Octal

10.2.1 Decimal Number System

The decimal number system has ten symbols, and any number of any magnitude can be expressed by using this system of positional weighting. For example, 6841 can be broken down as

$$6841 = 6000 + 800 + 40 + 1$$
$$= 6 \times 10^3 + 8 \times 10^2 + 4 \times 10^1 + 1 \times 10^0$$

The principle of positional weighting can be extended to any number system. A number can be represented by the equation

$$N = d_n * r^n + d_{n-1} * r^{n-1} + \ldots d_2 * r^2 + d_1 * r^1 + d_0 * r^0 \quad (10.1)$$

where N is the value of the entire number,
d_n is the value of the n^{th} digit from the decimal point, and
r is the *radix* or *base*.

Number of digitals $(n + 1)$; $(0, 1, 2, \ldots, n)$

10.2.2 Binary Number System

In this, each position in a number can take only one of two values: 0 or 1. These positions are called *bits*, a contraction of the words *binary digits*. The primary advantage of using this binary counting system as opposed to the 10-discrete line method in the decimal number system is that it minimises the number of lines required to two. Binary numbers are used

extensively throughout all digital systems because of the very nature of electronics. A "1" can be represented by a saturated transistor, a light turned on, a relay energised or a magnet magnetised in a particular direction. A "0" can be represented as a cut-off transistor, a light turned off, a de-energised relay, or the magnet magnetised in the opposite direction. We will study the binary number system in detail for the above reasons. Table 10.1 gives the binary representations of decimal numbers from 0 to 15.

Table 10.1 *Binary representations of decimal numbers from 0 to 15*

Binary	Decimal
0000	0
0001	1
0010	2
0011	3
0100	4
1101	5
0110	6
0111	7
1000	8
1001	9
1010	10
1011	11
1100	12
1101	13
1110	14
1111	15

From the table it is clear that using 4 bits we can represent decimal numbers from 0 to 15. Like the decimal system, the binary is also positionally weighted. Each position represents a particular value of 2^n.

$$1_2 = 0 \times 2^0 = 1$$

$$1010_2 = 1 * 2^3 + 0 * 2^2 + 1 * 2^1 + 0 * 2^0 = 8 + 0 + 2 + 0 = 10$$

The base is indicated by the subscript. Because of this property of positional weighting, the procedure for converting a binary number to decimal is very similar to that of breaking a decimal number into its weighted values. In Eq. (10.1), *d*'s will all be 0's or 1's, the *r*'s will all be 2, and the radix and the *n*'s will be various powers of 2, depending on the position of the digit with reference to the binary point.

EXAMPLE 10.1 Convert 1101101_2 to decimal.

Solution

$$N = d_6 * r^6 + d_5 * r^5 + d_4 * r^4 + d_3 * r^3 + d_2 * r^2 \_ d_1 * r^1 + d_0 * r^0$$
$$= 1 * 2^6 + 1 * 2^5 + 0 * 2^4 + 1 * 2^3 + 1 * 2^2 + 0 * 2^1 + 1 * 2^0$$
$$= 64 + 32 + 0 + 8 + 4 + 0 + 1$$
$$= 109_{10}$$

EXAMPLE 10.2 Convert 101010101 to decimal; the radix may not be always indicated.

Solution

$$N = d_8 \times r^8 + d_7 \times r^7 + d_6 \times r^6 + d_5 \times r^5 + d_4 \times r^4 + d_3 \times r^3 + d_2 \times r^2 + d_1 \times r^1 + d_0 \times r^0$$
$$= 1x2^8 + 0x2^7 + 1x2^6 + 0x2^5 + 1x2^4 + 0x2^3 + 1 \times 2^2 + 0x2^1 + 1x2^0$$
$$= 256 + 0 + 64 + 0 + 16 + 0 + 4 + 0 + 1$$
$$= 341_{10}$$

The above example provides the way for converting binary numbers to decimal numbers. However, in digital circuits, it is necessary to convert from decimal to binary also.

✦ Converting Decimal to Binary

This is the reverse process of the previous section. The first method requires the table of powers of 2. Starting with the decimal number to be evaluated, obtain the largest power of 2 from the table without exceeding the original number. Subtract the number selected from table from the original number and repeat this process for the remainder until it becomes zero. Finally, add the binary numbers obtained from Table 10.2. The result will give the corresponding binary number.

Table 10.2 *Decimal-to-binary conversion*

	Powers of 2	
2^n	n	2^{-n}
1	0	1.0
2	1	0.5
4	2	0.25
8	3	0.125
16	4	0.0625
32	5	0.03125

(*Contd.*)

64	6	0.015625
128	7	0.0078125
256	8	0.00390625
512	9	0.001953125
1024	10	0.0009765625
2048	11	0.00048828125
4096	12	0.000244140625
8192	13	0.0001220703125
16348	14	0.00006103515625
32768	15	0.000030517578125
65536	16	0.0000152587890625

EXAMPLE 10.3 Convert 69_{10} to binary.

Solution From the above table, 64 is the largest number without exceeding 69.
$64 = 1000000_2$; $69 - 64 = 5$; from the table, 4 is the largest number without exceeding 5.

$4 = 100_2$; $5 - 4 = 1$, one is the largest number in the table

$1 = 1_2$; Adding the binary numbers

$$\begin{array}{r} 1000000 \\ 100 \\ 1 \\ \hline \end{array}$$

Therefore, $69_{10} = 10000101_2$

The second method is successive division by 2 and the remainders recorded. In this, the last remainder is the *most significant* bit (*MSB*).

EXAMPLE 10.4 Convert 69 to binary.

Solution

$$\begin{array}{l} 2)\underline{69} \\ 2)\underline{34-1} \\ 2)\underline{17-0} \\ 2)\ \underline{8-1} \\ 2)\ \underline{4-0} \\ 2)\ \underline{2-0} \\ \quad 1-0 \end{array} \uparrow$$

Read the remainders from bottom to top.

$69_{10} = 1000101_2$. This is same as the result of Example 10.3.

✦ Binary Addition and Subtraction

Addition is accomplished in a similar manner to that in decimal addition.

EXAMPLE 10.5 Add 101111_2 and 10111_2.

Solution

Number 1 1 0 1 1 1 1

Number 2 1 0 1 1 1

$1\ 0\ 0\ 0\ 1\ 1\ 0_2 \rightarrow 0$ + carry 1

During the addition, First column 1 + 1 = 10

Second column $1 + 1 + 1 = (0 + 1) + \text{carry}\ 1 \rightarrow 1\quad 1$

↑ carry

EXAMPLE 10.6 Add 1111 and 1111.

Solution

Number 1:	1 1 1 1
Number 2:	1 1 1 1
Carry:	1 1 1
	$1\ 1\ 1\ 1\ 0_2$

Binary subtraction is performed in a manner similar to that in decimal subtraction. Because there are only two digits in binary, its subtraction often requires more borrowing operation than decimal numbers.

EXAMPLE 10.7 Subtract 1110 from 10000.

Solution

10000

1110

00010_2

We find that in the second column 1 cannot be subtracted from 0. So a 1 must be borrowed the from third column but it is a 0. In this example, 1 is available at the fifth column. So borrow this 1, leaving behind a 0. Then 1 is(1 + 1) in the fourth column. We borrow 1 leaving behind 1 in the fourth column. Finally, successive borrowing makes (1 + 1) in the second column from which we subtract 1, yielding 1 as answer in the second column. At this stage, we have the answers for zeroth and first column. The third, fourth and fifth columns are

Column	5	4	3
	0	1	1
		1	1
	0	0	0

Thus, the complete answer is

$$00010_2 = 2_{10}$$

EXAMPLE 10.8 Subtract 10101 from 101010.

Solution

$$\begin{array}{r} 101010 \\ 10101 \\ \hline \\ 10101_2 \\ \hline \end{array}$$

✦ Binary Multiplication

EXAMPLE 10.9 Multiply 1111_2 by 101_2.

Solution

$$\begin{array}{r} 1111 * 101 \\ \hline 1111 \\ 0000 \\ 1111 \\ \hline 1001011 \\ \hline \end{array}$$

The above method is a paper method and cannot be used by a computer. Multiplying a 4-bit number by another 4-bit number yields an eight-bit number results.

✦ Binary Division

Like multiplication, division can be accomplished by two methods: Paper and computer method. The paper method is long-division procedures.

EXAMPLE 10.10 Divide 110110_2 by 110_2.

Solution

```
      1001
110)110110
    110
     0110
      110
        0
```

✦ Computer Division Method

The computer method uses successive subtraction. Assume the above example with 8-bit representation; dividend is to be divided by the divisor in 4-bit representation. That is, dividend is 00110110, and divisor is 0110. The quotient will be formed in the right half of the *MQ* register and the remainder in the left half. The dividend is first placed in the *MQ* register and divisor in the *D* register. The divisor is then subtracted from the dividend. The result is considered positive if the most significant bit (the far left bit or the ninth bit) is a 0, and negative if it is a 1. If the result is positive then the error has occurred for the quotient would be greater than four bits.

MQ	0011 0110	
Subtract *D*	0110	
MQ	1101 0110	Result is negative; the division is valid
Add *D*	0110	
MQ	0011 0110	
Shift *MQ* left	0110 1100	
Subtract *D*	0110	
	0000 1100	Result is positive
Add 1 to quotient	1	
MQ	0000 1101	
Shift *MQ* left	0001 1010	
Subtract *D*	0110	
	1011 1010	Result is negative
Add *D*	0110	
MQ	0001 1010	Put a zero quotient
Shift *MQ* left	0011 0100	
Subtract *D*	0110	
	1101 0100	Result is negative

Add *D*	0110	
	0011 0100	Put a zero quotient
Shift *MQ* left	0110 1000	
Subtract	0110	
	0000 1000	Result is positive
	0000 1000	
Add I to quotient	1	
	0000 1001	Final answer
	Remainder Quotient	

If the result is negative then the quotient will be four or less bits that is sufficiently small to be contained in a 4-bit register. The *MQ* register is next shifted left one bit and the divisor subtracted from it. If the result is positive, 1 is added to the *Least Significant Bit* (*LSB*) of the *MQ* register, where the quotient is accumulated. If it is negative, the divisor is added to *MQ* and the *MQ* shifted left one bit. This effectively puts a 0 in this bit of the quotient. The process is continued until the *MQ* register has been shifted left four bits the number of bits in the D register. The remainder is then in the left half of the *MQ* register and the quotient in the right half.

10.2.3 Hexadecimal Number System

The hexadecimal number system was born out of the need to express binary numbers concisely and is by far the most commonly used number system in computer. The hexadecimal number is formed from a binary number (word) by grouping bits in groups of four bits each, starting at the binary point. This is a logical way of grouping since computer words come in 8 bits, 12 bits, 16 bits, 32 bits, and so on. In a group of 4 bits, the decimal number 0–15 can be represented by a unique symbol; 0, 1, ..., 9 and 10–15 are represented by symbols A, B, C, D, E and F.

✦ Converting Binary to Hexadecimal

Binary numbers can be easily converted to hexadecimal by grouping in groups of four starting at the binary point.

EXAMPLE 10.11 Convert 1010111011110101_2 to hexadecimal.

Solution

1010	1110	1111	0101	Group in four = from LSB
A	E	F	5	Convert each number

Table 10.3 *Conversion table*

Decimal	Hexadecimal	Binary
0	0	0000
1	1	0001
2	2	0010
3	3	0011
4	4	0100
5	5	0101
6	6	0110
7	7	0111
8	8	1000
9	9	1001
10	A	1010
11	B	1011
12	C	1100
13	D	1101
14	E	1110
15	F	1111
16	10	10000
32	20	100000
64	40	1000000
100	64	01100100
255	FF	1111 1111
511	1FF	1 1111 1111
1024	400	100 0000 0000

✦ Converting Hexadecimal to Binary

It can be converted to binary by converting each digit.

EXAMPLE 10.12 Convert $FA876_{16}$.

Solution

F	A	8	7	6
1111	1010	1000	0111	0110

Thus, the solution is 1111101010000111010110_2.

✦ Converting Hexadecimal to Decimal

Hexadecimal number is also a positionally weighted system, where the radix is 16.

$$N = d_4 * 16^4 + d_3 * 16^3 + d_2 * 16^2 + d_1 * 16^1 + d_0 * 16^0$$

EXAMPLE 10.13 Convert $FA27D_{16}$.

Solution

$$F * 16^4 + A * 16^3 + 2 * 16^2 + 7 * 16^1 + D * 16^0$$
$$= 15 * 65536 + 10 * 4096 + 2 * 356 + 7 * 16 + 13$$
$$= 10,24,637$$

✦ Converting Decimal to Hexadecimal

Any decimal number can be converted to hex by successively dividing by 16. The remainders can then be converted to hex, and read up from the bottom to obtain the hexadecimal results.

EXAMPLE 10.14 Convert 57345.

Solution

16) 57345 ↑
16) 3584 – 1
16) 224 – 0
14 – 0

14 00 1 ⟶ Convert decimal to hexadecimal notation.
↓
E 0 0 1_{16}

The result is $E001_{16}$.

10.2.4 Octal Numbers

This was used extensively by early minicomputers. The octal system is formed by grouping bits in groups of 3, starting at the binary point.

✦ Binary to Octal Conversion

EXAMPLE 10.15 Convert 111110101101_2 to octal.

Solution

Split the number into group of three	111	110	101	101
	7	6	5	5

The result is 7655_8

EXAMPLE 10.16 Convert 6754_8 to binary.

Solution

6	7	5	4
110	111	101	100

The result is 110111101100_2.

EXAMPLE 10.17 Convert 867_{10} to octal number. It is simply a successive division by 8.

Solution

$$
\begin{array}{l}
8)\,867 \\
8)\,108 - 3 \\
8)\,13 - 4 \\
\quad 1 - 5
\end{array}
\uparrow
$$

The result is $(1543)_8$

10.2.5 Complement of Binary Numbers

The computer arithmetic process is done not only with positive numbers; it is also done with negative numbers. In such a condition, signed and unsigned numbers are dealt by the processors. The 1's and 2's complement operation is useful for this type of arithmetic process.

1's complement of a binary number is just an inversion of individual bits. For example, the 1's complement of 1010101 is found by inversion of each bit, i.e. 0 to 1 and 1 to zero conversion.

1's complement of $(1010101)_2$
is $(0101010)_2$

The 2's complement of a positive binary number whose integral part has p digits is also found by subtracting the number from 2^p. For instance, 2's complement of 11010 is

$$2^5 - 11010 = 100000 - 11010 = 00110$$

2's complement of 1011.001 is

$$2^4 - 1011.001 = 10000 - 1011.001$$
$$= 0100.110$$

The 2's complement of 0 is defined to be 0.

The 2's complement of a given binary number is also found by adding 1 to the least significant bit of the 1's complement of that binary number. The 2's complement of the number $(10101011)_2$ is found by two steps. First, convert the given number into its 1's complement by inversion.

1010101 → 0101010 is the 1's complement of the number. Then add one to the least significant bit (LSB). So

0101010
1

0101011_2 – 2's complement of the number is $(1010101)_2$. 2's complement is used to represent negative numbers.

The above procedure also applies for a decimal binary number. For example,

To find 2's complement of	0.1110100
1's complement	0.0001011
	+1
2's complement	0.0001100
To find 2's complement of	1011.001
1's complement	0100.110
	+1
2's complement	0100.111

✦ 2's Complement of Floating Point Numbers

Example: What would be the 2's complement of 010111.1100?
Ans: 101000.0100

The exponent is represented using excess-N and the significant is represented using signed binary, which allows the significant to be conveniently split into a sign bit and an unsigned magnitude that can be processed independently.

To create 2's complement of 010111.1100, proceed as:

Step 1:Start on the least significant bit and located the first 1 marked red 01011.1100.
Step 2:Then flip every bit after that first one example: 1 change to 0 and vice versa.
Step 3:2's complement is 010111.1100→101000.0100.

✦ Difference of Two Positive Binary Numbers ($m - n$)

1. To m add 2's complement of n.
2. If the sum has a carry at the left end, delete it. The result is $(m - n)$.
3. If the sum has no carry at the left end, take 2's complement of the sum. Attach a negative sign. The result is $(m - n)$.

EXAMPLE 10.18 Given m = 11010110, n = 01000101, determine (a) $(m - n)$ and (b) $(n - m)$.

Solution

(a) 2's complement of n

	01000101
1's complement	10111010
	+1
2's complement	10111011

	Add in	11010110
	Delete carry →1	10010001
	$(m - n)$	10010001
(b)	2's complement of m	
		11010110
	1's complement	00101001
		+1
	2's complement	00101010
	Add n	01000101
	No carry	01101111
	1's complement	10010000
		+1
	2's complement $(n - m)$	–10010001

10.3 BOOLEAN ANALYSIS OF LOGIC CIRCUITS

Logic circuits are the electronic circuits which have a number of logic gates and perform logical operations on the given data. These circuits function in binary manner, i.e. each of input or output given to the circuit is either 0 or 1. In 1854, Boolean (mathematician) found that the logic has a connection with mathematics and performed the mathematical analysis of logic circuits. On the basis of his investigation on law of thought, he constructed a logical algebra. This logical algebra is called **Boolean algebra**. Boolean algebra is extensively used in digital electronics. It is used for the simplification, design and analysis of the digital circuits. Boolean algebra follows a set of rules and laws dealing with the relation between logic and Boolean variables, by which logical operations can be expressed mathematically with single equation. Logical operators (AND, OR, NOT, XOR, NAND, NOR) are further used to simplify and manipulate the equations into new forms.

Boolean algebra differs from regular mathematics algebra in a way that Boolean constants and variables deal with only two discrete values, i.e. 0 and 1 (predefined voltage levels: 1 representing HIGH level and 0 representing LOW level). In digital system, Boolean value 0 relates to any voltage in range from 0–0.8 V and Boolean value of 1 relates to any voltage of 2–5 V.

10.4 BOOLEAN ALGEBRA THEOREMS

A Boolean algebra is a set of binary operations, + and *, and a unary operation, –, and elements 0, 1 such that the following laws hold: commutative and associative laws for addition and multiplication, distributive laws both for multiplication over addition and for addition over multiplication. The applications of digital logic involve functions of the AND, OR, and NOT operations.

In Boolean algebra, a variable A called *bilateral*, can take on two values, 1/0. Three basic binary operations are defined here before proceeding further.`

✦ AND—Multiplication Symbol (.)

$Y = A$ AND $B = A \cdot B$

Truth Table

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

In AND operation, the output is 1 only if both A and B are 1; otherwise it is zero.

✦ OR—Addition Symbol (+)

$Y = A + B$

Truth Table

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

In OR operation, the output is '1' if either A or B or both are 1.

✦ NOT—Complement, Symbol (super –, super', may also be used)

$Y = \overline{A}$

Truth Table

A	Y
0	1
1	0

In a NOT operation, the output is the reverse of input.

❑ *Special Case*

AND Operation

$A \cdot 0 = 0, \quad A \cdot 1 = A$

$A \cdot A = A, \quad A \cdot \overline{A} = 0$

OR Operation

$A + 0 = A, A + 1 = 1, A + A = A, A + \overline{A} = 1$

These results follow from the truth tables by replacing B with 0, 1 or $\overline{A}$

$\overline{\overline{A}} = A$

Boolean identities of common use are presented in Table 10.4.

Table 10.4 *Some identities of Boolean algebra*

Name	AND form	OR form
Identity law	$1 \cdot A = A$	$0 + A = A$
Null law	$0 \cdot A = 0$	$1 + A = 1$
Idempotent law	$A \cdot A = A$	$A + A = A$
Inverse law	$A \cdot \overline{A} = 0$	$A + \overline{A} = 1$
Commutative law	$A \cdot B = B \cdot A$	$A + B = B + A$
Associative law	$(A \cdot B) \cdot C = A \cdot (B \cdot C)$	$(A + B) + C = A + (B + C)$
Distributive law	$A + BC = (A + B) \cdot (A + C)$	$A(B + C) = AB + AC$
Absorption law	$A \cdot (A + B) = A$	$A + AB = A$
De Morgan's law	$\overline{A \cdot B} = \overline{A} + \overline{B}$	$\overline{A + B} = \overline{A}\,\overline{B}$

As the AND operation is unambiguous, we may omit the dot(.). Thus, we can write $A \cdot B \cdot C$ as ABC.

✦ Proofs of Some Results of Table 10.4

❑ *Distributive Law*

$$\begin{aligned}(A + B)(A + C) &= AA + AC + BA + BC \\ &= A + A(B + C) + BC \\ &= A[1 + (B + C)] + BC \\ &= A + BC \quad \text{as} \quad 1 + (B + C) = 1\end{aligned}$$

❑ *Absorption Law*

$$\begin{aligned}A(A + B) &= AA + AB = A + AB \\ A(1 + B) &= A\end{aligned}$$

✦ De Morgan's Theorem

$$\overline{A + B} = \overline{A} \cdot \overline{B} \quad \text{and} \quad \overline{A \cdot B} = \overline{A} + \overline{B}$$

This is the most important logic theorem for digital electronics, which states that any logical binary expression remains unchanged if we

1. Change all variables to their complements,
2. Change all AND operations to ORs,
3. Change all OR operations to ANDs, and
4. Take the complement of the entire expression.

A practical operational way to look at De Morgan's theorem is that the inversion bar of an expression may be broken at any point and the operation at that point is replaced by its opposite (i.e. AND replaced by OR or vice versa).

Table 10.5 *Truth table to prove De Morgan's theorems*

A	B +	$\overline{A}$	$\overline{B}$	$\overline{A \cdot B}$	$\overline{A} + \overline{B}$	$\overline{A + B}$	$\overline{A} \cdot \overline{B}$
0	0	1	1	1	1	1	1
0	1	1	0	1	1	0	0
1	0	0	1	1	1	0	0
1	1	0	0	0	0	0	0

10.5 DIGITAL CIRCUITS

Digital circuits are classified into two major categories: Combinational circuits and Sequential circuits. *Combinational circuits* are the circuits where output depends on the present input only. The *sequential circuit* produces the output on the basis of both present and previous inputs. It shows that sequential circuits have memory. Before studying these two types, first let us study the basic logic gates which will form the digital circuits.

10.5.1 Logic Gates

Logic gates are the building blocks of digital circuits. They are used to create digital circuits and even complex integrated circuits. Complex integrated circuits are complete circuits ready to perform several functions—microprocessors and microcontrollers are the best examples—the inner sub-circuits are formed using several logic gates. Combinations of logic gates form circuits designed with specific tasks in mind. For example, logic gates are combined to form circuits to add binary numbers (adders), set and reset bits of memory (flip-flops), multiplex multiple inputs, etc. The following section will give a brief idea of the different gates.

❑ ***Representation of 0 and 1 in Logic Gates*** In an electronic logic gate, 0 is represented by *low voltage* and 1 by *high voltage*.

✦ NOT Gate

The NOT gate is also known as an *inverter*, because it inverts the input to its opposite. The NOT gate accepts one input and the output is the opposite of the input as shown in Fig. 10.1 and its truth table below it. In other words, a low-voltage input (0) is converted to a high-voltage output (1); high voltage (1) is converted to low voltage (0).

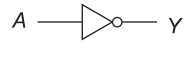


Fig. 10.1 Inverter

Table 10.6 *Truth table of NOT gate*

A (Input)	Y (Output)
0	1
1	0

✦ AND Gate

As the name implies, an AND logic gate performs an "AND" logic operation, which is a logic multiplication. It has at least two inputs. So, if A and B are its inputs, at the output we will find $Y = A \cdot B$. The AND gate symbol is shown in Fig. 10.2 and its truth table in Table 10.7.

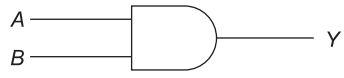


Fig. 10.2 AND logic gate

Table 10.7 *Truth table of AND gate*

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

Another way to express AND logic gate is that its output will only be at "1" when all its inputs are also at "1". Otherwise, its output will be "0".

✦ OR Gate

As its name implies, an OR logic gate performs an "OR" logic operation, which is logic addition. It has at least two inputs. So, if A and B are its inputs, at the output we will find $Y = A + B$. The OR gate symbol is drawn in Fig. 10.3 and its truth table besides it.

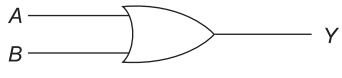


Fig. 10.3 OR logic gate

Table 10.8 *Truth table of OR gate*

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

Another way to express OR logic gate is that its output will only be at "0" when all its inputs are also at "0". Otherwise, its output will be "1".

✦ XOR Gate

XOR stands for *Exclusive-OR*. An XOR gate compares two values and if they are different, its output will be "1". The XOR operation is represented by the symbol ⊗. So $Y = A \otimes B$. The XOR logic gate symbol is given in Fig. 10.4 and its truth table along its side.

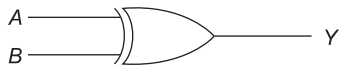


Fig. 10.4 XOR logic gate

Table 10.9 *Truth table of XOR gate*

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

So its output will only be at "0" when all its inputs have the same value. Otherwise, its output will be "1".

✦ NAND Gate

This is an AND gate with the output inverted. The output is high when either of inputs *A* or *B* is high, or if neither is high. In other words, it is high, going low only if both *A* and *B* are high. Its symbol is presented in Fig. 10.5 with the truth table.

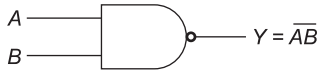


Fig. 10.5 NAND logic gate

Table 10.10 *Truth table of NAND gate*

A	B	Out
0	0	1
0	1	1
1	0	1
1	1	0

✦ NOR Gate

This is an OR gate with the output inverted. The output is high only when neither *A* nor *B* are high. That is, it is normally high but any kind of nonzero input will make it low. Its symbol and truth table are presented in Fig. 10.6 and in Table 10.11, respectively.

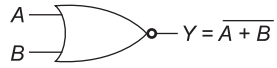


Fig. 10.6 NOR logic gate

Table 10.11 *Truth table of NOR gate*

A	B	Out
0	0	1
0	1	0
1	0	0
1	1	0

✦ XNOR Gate

It is the complement of XOR gate. Its output is 1 when both inputs are equal, i.e. (0, 0) or (1,1). XNOR operation is represented as $Y = A \odot B$. The symbol of XNOR gate and its truth table are given in Fig. 10.7 and Table 10.12.

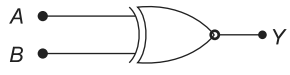


Fig. 10.7 XNOR logic gate

Table 10.12 *Truth table of XNOR gate*

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	1

✦ Non-inverter or Buffer

At a non-inverter, also known as a buffer, the value entered on its input will be found on its output. You may think that this is a crazy logic gate, since it does nothing. That's not true; it has several important applications in digital electronics, as we will explain below. Its symbol and truth table are presented in Fig. 10.8 and Table 10.13.

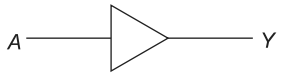


Fig. 10.8 Non-inverter or buffer

Table 10.13 *Truth table*

A	Y
0	0
1	1

A typical application for a buffer is to increase the *fan-out* of a given logic gate. Fan-out is the maximum number of gates a given integrated circuit is capable of being connected to. For example, if a given logic gate has a fan-out of three gates, its output can be only connected directly to three other logic gates. If you need to connect its output to more logic gates, you can use a buffer to increase the number of logic gates you can connect this output to.

10.5.2 Examples of Combinational Circuits

The half adder is an example of a simple functional digital circuit built from two logic gates. The half adder adds two one-bit binary numbers (A and B). The output is the sum of the two bits (S) and the carry (C). Figure 10.9 shows the schematic of half-adder circuit.

Note that the same two inputs are directed to two different gates. The inputs to the XOR gate are also the inputs to the AND gate. The input "wires" to the XOR gate are tied to the input wires of the AND gate.

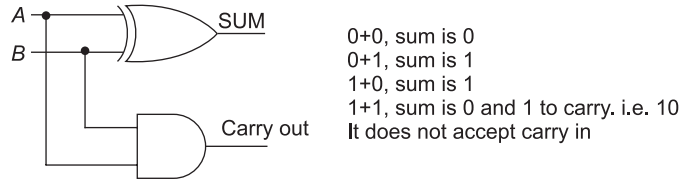


Fig. 10.9 Half adder

Table 10.14 Half-adder truth table

$S = A \oplus B$

$C = AB$

Row	A	B	C	S
0	0	0	0	0
1	0	1	0	1
2	1	0	0	1
3	1	1	1	0

The full-adder circuit adds three one-bit binary numbers (A, B, C_i) and outputs two one-bit binary numbers, a sum (S) and a carry (C_0), If you look at Fig. 10.10 closely, you'll see the full adder is simply two half adders joined by an OR.

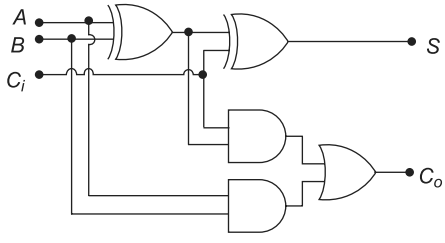


Fig. 10.10 Full adder

Table 10.15 *Full-adder truth table*

Row	A	B	C_{in}	C_{out}	Out
0	0	0	0	0	0
1	0	0	1	0	1
2	0	1	0	0	1
3	0	1	1	1	0
4	1	0	0	0	1
5	1	0	1	1	0
6	1	1	0	1	0
7	1	1	1	1	1

The full adder is usually a component in a cascade of adders, which add 8, 16, 32, etc. binary numbers. The carry input for the full-adder circuit is from the carry output from the circuit "above" itself in the cascade. The carry output from the full adder is fed to another full adder "below" itself in the cascade.

An n-bit adder can then be designed by connecting the carry out and carry in lines of n full adders. Figure 10.11 shows a 4-bit adder. This design is called a *ripple-carry* adder. Similarly, four 4-bit adders can be connected to form a 16-bit adder, etc.

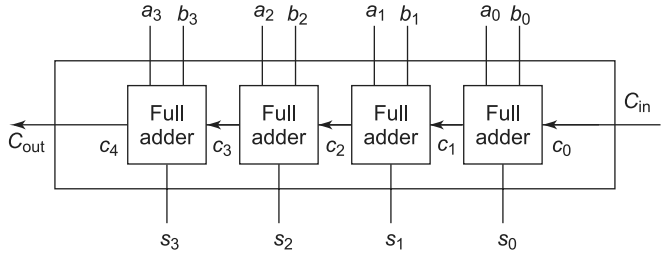


Fig. 10.11 4-bit adder

10.5.3 Multiplexers

The objective of a multiplexer is to select one signal from a group of 2^n inputs, to be an output on a single output line. For example, an 8-to-1 multiplexer (mux) is diagramed as a black box as shown in Fig. 10.12. Lines D_0, …D_7 are the data input lines and F is the output line. Lines A, B, and C are called the select lines. They are interpreted as a three-bit binary number, which is used to choose one of the D lines to be output on the line F.

✦ Implementation

Design of an 8-to-1 multiplexer using traditional minimisation techniques would require minimising a Boolean function of 11 variables. Instead, a mux can be designed with a regular pattern of AND and OR gates, as shown in Fig. 10.13.

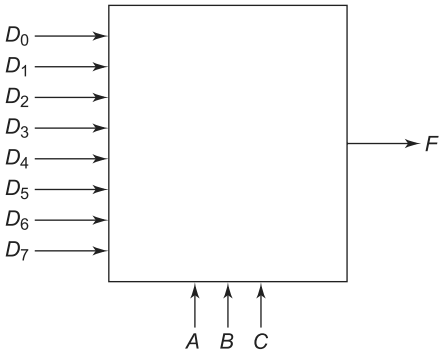


Fig. 10.12 8-to-1 multiplexer

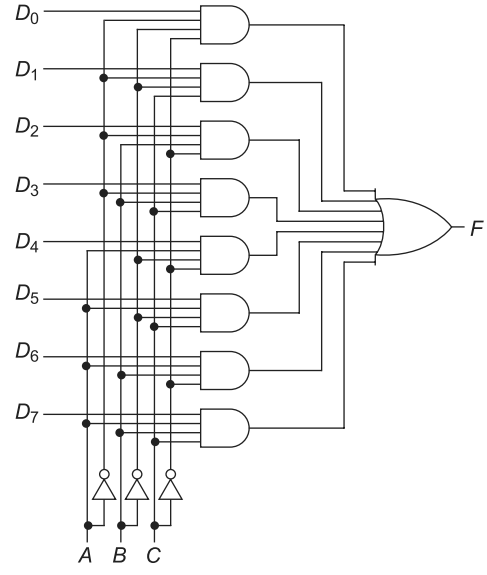


Fig. 10.13 Implementation of an 8-to-1 multiplexer

Any literal $A = 1$, $\bar{A} = 0$

Suppose we want to output $Y = D_5$, then the three inputs to the corresponding AND gate should be 10s. Then

$$\left.\begin{array}{ll} D_5 = 0, & Y = 0 \\ D_5 = 1, & Y = 1 \end{array}\right\} \rightarrow Y = D_5$$

The corresponding select inputs are

$$5 \rightarrow 101 \text{ or } A\bar{B}C$$

And the output is $A\bar{B}C \rightarrow D_5$

Similarly to output, D_6

$$6 \rightarrow 110 \rightarrow AB\bar{C}$$

Output is $AB\bar{C} \rightarrow D_6$

The output Y will be sum (+) of all eight (2^3) outputs.

✦ Application

Choose one of several registers to be used as ALU (Arithmatic Logic Unit) input. A 2^n-to-1 multiplexer can be used to implement an arbitrary Boolean function of n variables, by associating each input line of the multiplexer with a row of the truth table for the function.

10.5.4 Decoders

The objective of the decoder is to decode an n-bit binary number, producing a signal on one of 2^n output lines. Figure 10.14 shows an 3-to-8 decoder.

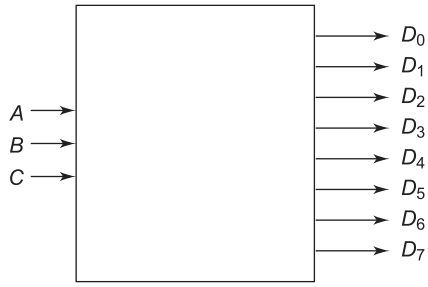


Fig. 10.14 3-to-8 decoder

✦ Implementation

A decoder is often implemented with an additional input called an "enable" line. When the line is enabled, the circuit is a decoder. When it is disabled, all the outputs are 0. The same circuit can be used as a de-multiplexer, which directs a single data input line to one of 2^n output lines, depending on the values of n select lines. Figure 10.15 shows the 3-to-8 decoder implementation.

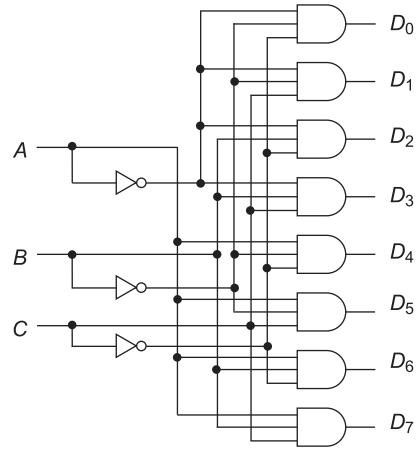


Fig. 10.15 3-to-8 decoder implementation

✦ To Produce Output 1 at D_5

$5 \rightarrow 101$

If input is $A = 1$, $B = 0$, $C = 1$ or $A\bar{B}C$, the output is 1 and D_5. It can be checked that all other outputs are 0.

To produce 1 at D_3 we input $3 \rightarrow 011$ or $\bar{A}BC$

❑ ***Reader*** Find the input to produce 1 at D_7.

❑ ***Application*** Decode memory address for reads and writes to random access memory.

10.6 INTRODUCTION TO SEQUENTIAL CIRCUITS

The circuits we have studied so far (gates, multiplexers, decoders, adders) are all examples of combinational circuits. They have the property that their outputs depend only on the current input values (after some delay). Because of this property, combinational circuits are memoryless. Their outputs cannot depend on past input values which are no longer present. How can our logic design theory be extended to introduce memory elements? The simplest memory circuit is called an *SR* latch, as shown in Fig. 10.16.

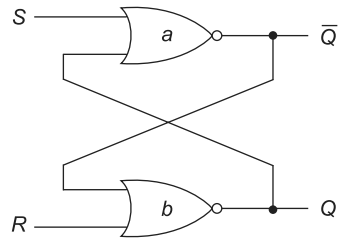


Fig. 10.16 SR flip-flop

✦ Operation

We remind ourselves here that a NOR gate produces output 1 only when both inputs are 0, otherwise the output is 1.

❑ ***Set State*** Output $Q = 1$, $\bar{Q} = 0$

❑ ***Reset State*** Output $Q = 0$, $\bar{Q} = 1$

Let the *SR* FF be in set state. We make $S = 0$ and $R = 0$. The inputs to '*a*' NOR are (0, 1); so $\bar{Q} = 0$, and inputs to '*b*' NOR are (0, 0); so $Q = 1$. Therefore, *SR* F/F remains in set state, i.e. no state change.

It similarly shows that when the *SR* F/F is in reset state and we make $S = 0$, $R = 0$, it remains in reset.

Thus, if both inputs are zero ($S = R = 0$), the output state does not change.

If output is in set state ($Q = 1$, $\bar{Q} = 0$), we input $S = 1$, $R = 0$, F/F remains in set state. Let us input S = 0, $R = 1$. The inputs to '*b*' NOR are (0, 1) causing $Q = 0$. This in turn makes the input to '*a*' NOR as (0, 0) and so $\bar{Q}$ changes to 1. Thus, the state changes from set state to reset state.

Similarly, if the output is in reset state ($Q = 0$, $\bar{Q} = 1$), it remains so if $S = 0$ and $R = 1$. But changes to set state if $S = 1$ and $R = 0$.

Both $S = 1$ and $R = 1$ are not permitted as this would cause a conflict, and outcome would be uncertain.

The above operations of *SR* F/F are summarised in Table 10.16.

Table 10.16 *Truth table for SR flip-flop*

Inputs		Present state	Next state
S	R	Q_n	Q_{n+1}
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	1
1	1	Not allowed	

It is seen that the output (Q and $\bar{Q}$) depends on the S, R inputs and the previous state. Thus, the latch holds one bit of data. Circuits involving latches are *called sequential* circuits. They are included memory elements.

Within a digital computer, a clock is used to synchronise changes in the contents of memory elements. A clock is a signal which oscillates between 0 and 1, as shown in Fig. 10.17.

Fig. 10.17 Clock signal

The clock can be applied to a latch so that change in the latch's value can only occur when the clock is in the "1" (high) state. The result is a clocked *SR* latch shown in Fig. 10.18.

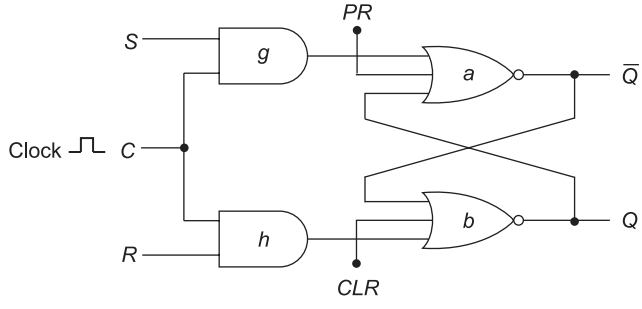


Fig. 10.18 Clocked SR flip-flop

When $C = 0$, the output of both AND gates is zero. So Q cannot change: the latch is non-operational.

When $C = 1$, output of gate g is $1.S = S$ and output of gate h is $1 \cdot R = R$, the latch is now operational.

The term *level-triggered* is applied to this type of latch to indicate that its ability to change value depends on the level (low or high) of the clock signal.

✦ Clear and Preset

In Fig. 10.18, one additional input is provided to each NOR gate; CLR on Q side NOR and PR on $\overline{Q}$ side NOR. When CLR and PR are 0, the circuit behaves as an *SR* flip-flop as in Fig. 10.19. In the absence of a clock pulse ($C = 0$), the outer inputs to NOR gates are zero. Making CLR = 1 and PR = 0, results in $Q = 0$ and $\overline{Q} = 1$, thereby clearing the flip-flop. Instead making PR (preset) = 1 and CLR = 0, results in $Q = 1$ and $\overline{Q} = 0$; thereby presetting the flip-flop. These are asynchronous inputs as their action does not need a clock pulse. The symbol of an *SR* flip-flop is sketched in Fig. 10.19. It may not be necessary to always label CLR and PR.

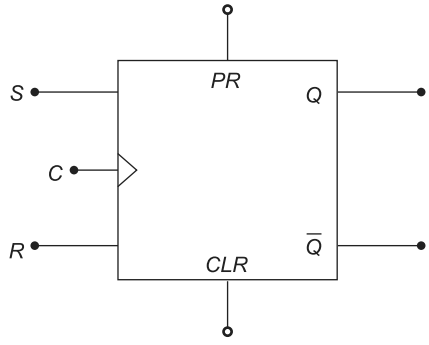


Fig. 10.19 Symbol of SR flip-flop

10.7 | FLIP-FLOPS

For more precise synchronisation, "flip-flops" are used. They are edge-triggered that means they can change with change in the value of the clock signal; that is, on its rising or falling edge. One way to accomplish this is by the use of a small circuit, called a *pulse generator*, shown in Fig. 10.20.

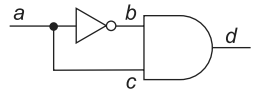


Fig. 10.20 Pulse generator (edge triggering)

When *a* is connected to a clock signal, a short pulse will be generated on *d*, wherever *a* makes a transition from 0 to 1 and 1 to 0.

There are four commonly used types of flip-flops: *D*, *T*, *RS* and *JK*. The behaviour of each of them can be described by a state table. The state table (Table 10.17) shows how the value of the flip-flop changes in response to its inputs.

Table 10.17 *The state table of various flip-flops*

RS (reset-set)		
R	**S**	**Q_{next}**
0	0	Q
0	1	1
1	0	0
1	1	undefined

D (data)	
D	**Q_{next}**
0	0
1	1

JK		
J	**K**	**Q_{next}**
0	0	Q
0	1	0
1	0	1
1	1	$\overline{Q}$

T (toggle)	
T	**Q_{next}**
0	Q
1	$\overline{Q}$ (toggle)

✦ JK Flip-Flop (FE) (Fig. 10.21)

SR, *FF* is converted to *JK*, *FF* by feeding $\overline{Q}$ to the upper *AND* and *Q* to lower *AND*. It also takes illegal inputs values of *SR* ($S = R = 1$).

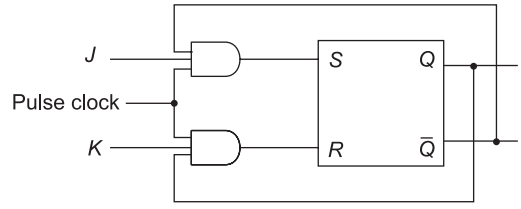


Fig. 10.21 JK Flip Flop

As the basic flip-flop is *SR*, we will consult Table 10.17 for each set *SR* to determine the output of the *JK* flip-flop. The values of *S*, *R* are governed by inputs *J*, *K* and *Q* (and so $\overline{Q}$ to the two AND gates). The outputs for various combinations of inputs, (*J*, *K*) when a pulse is applied, i.e. *PT* (Positive Trigger) or *NT* (Negative Trigger) are given below.

- $J = 0, K = 0, Q = 0 \Rightarrow S = 0, R = 0 \Rightarrow Q_{next}$, no change
  $J = 0, K = 0, Q = 1 \Rightarrow S = 0, R = 0 \Rightarrow Q_{next}$, no change
- $J = 0, K = 1, Q = 0 \Rightarrow S = 0, R = 0 \Rightarrow Q_{next} = 0$, no change
  $J = 0, K = 1, Q = 1 \Rightarrow S = 0, R = 1 \Rightarrow Q_{next} = 0$, State resets
- $J = 1, K = 0, Q = 0 \Rightarrow S = 1, R = 0 \Rightarrow Q_{next} = 1$, State sets
  $J = 1, K = 0, Q = 1 \Rightarrow S = 0, R = 0 \Rightarrow Q_{next} = 1$, no change

The above results confirm the first of the truth table of *JK* flip-flop (Table 10.19).
Consider now (1, 1) inputs which were entries not permitted in an *SR* flip-flop.

- $J = 1, K = 1, Q = 0 \Rightarrow S = 1, R = 0 \Rightarrow Q_{next} = 1 = \bar{Q}$, State sets
  $J = 1, K = 1, Q = 1 \Rightarrow S = 0, R = 1 \Rightarrow Q_{next} = 0 = \bar{Q}$, State resets

The above results confirm the *JK* truth table (Table 10.16).

It is also concluded from above that as long as *JK* flip-flop inputs are $J = K = 1$, the state changes from 0 to 1, or 1 to 0, that is it *toggles*. However, this causes a problem. If any clock pulse is too long, the output state will change more than once and the final state of the flip-flop will be indeterminate. To avoid this problem, a *JK* flip-flop is constructed with two *SR* flip-flops in master-slave connection as in Fig. 10.22.

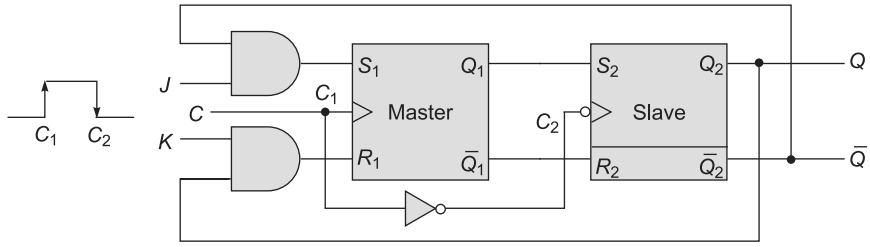


Fig. 10.22 JK master slave flip-flop

✦ Operation

By virtue of master to slave connection,
$S_2 = Q_1, R_2 = \bar{Q}_1$

It means that input to the slave is always 0/1 or 1/0. Therefore, as pulse *C* goes low, $C_2 = 1$ and the slave transfers the output of the master to the system output $Q = Q_2 = Q_1$.

When $C_1 = 1$, the master produces output Q_1 as per the *JK* action so long as $J = K \neq 1$. Consider one case.

$$J = 1, K_1 = 0, Q = 0 \Rightarrow S_1 = 1, R_1 = 0, \xrightarrow{C = C_1 = 1} \cdot Q_1 = 1 = S_2, \xrightarrow{C = 0, C_2 = 1} Q_2 = Q = 1$$

This agrees with the *JK* truth table 10.16. The *JK* action occurs for all input combinations other than $J = K = 1$.

For $J = K = 1$, toggling should happen at output without any indeterminate solution.

$$J = 1, K = 1, Q = 1 \Rightarrow S_1 = 0, R_1 = 1 \xrightarrow{C = C_1 = 1} Q_1 = 0 \Rightarrow S_2 = 0\ , R_2 = 1 \xrightarrow{C = 0, C_2 = 1} Q = 0 = \bar{Q}$$

$$J = 1, K = 1, Q = 0 \Rightarrow S_1 = 1, R_1 = 0 \xrightarrow{C = C_1 = 1} Q_1 = 1 \Rightarrow S_2 = 1\ , R_2 = 0 \xrightarrow{C = 0, C_2 = 1} Q = 1 = \bar{Q}$$

Thus confirms the toggle operation of the *JK* master-slave flip-flop. It is to be noted that the master flip-flop operates when $C = 1$ whereas the slave operation flows when pulse goes low, i.e. $C = 0$. When the slave operation occurs, the master is inoperative and so unwanted toggling cannot occur, which was the case in *JK* flip-flop of Fig. 10.21. Because of this sequential action of slave (after master has become inoperative) and *C* goes low, the *JK* master-slave flip-flop is also referred to as *trailing edge triggered flip-flop*.

A *D* flip-flop action is achieved by a *JK* flip-flop by connecting to *K* through a NOT gate as in Fig. 10.23(a). The input (1/0) is reproduced at the output (1/0).

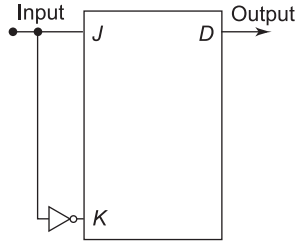


Fig. 10.23(a) D flip-flop

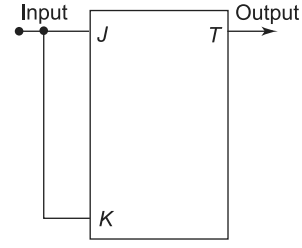


Fig. 10.23(b) T flip-flop

The reader may check from the *JK* truth table.

A *T* flip-flop function is obtained from a *JK* flip-flop by connecting *J* to *K* and input at *J* as in Fig. 10.23(b). The reader may check against *JK* truth table.

10.7.1 Shift Register

For example, consider the following circuit shown in Fig. 10.24 containing two *RS* flip-flops. When a clock pulse occurs, the value of the left flip-flop is copied to the right flip-flop. This circuit is known as a *shift register*.

Flip-flops are used within a CPU (Central Processing Unit) to implement registers. A register is an ordered group of *n* flip-flops. For example, in the Intel architecture, EAX is a 32-bit register. It can hold a 32-bit binary integer.

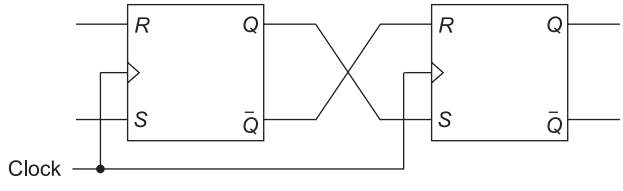


Fig. 10.24 Shift (right) register

The working principle of a shift register has already been introduced in Fig. 10.19 using *SR* flip-flop. In Fig. 10.24, we present a four-bit shift register employing *D* as flip-flops, which *trailing edge triggered* indicated by a small circle at the back of the clock triangle.

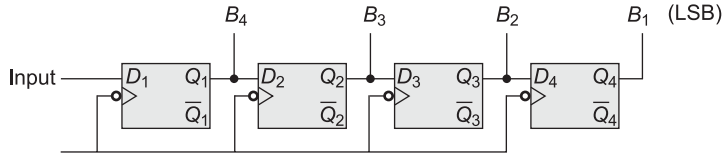


Fig. 10.25 Shift register

Let the input sequence be A_1, A_2, A_3, A_4 and the initial state of the register be $Q_1 = Q_2 = Q_3 = Q_4 = 0$. Before the first pulse arrives, the four data (D) bits are $D_1 = A_1$, $D_2 = Q_1 = 0$, $D_3 = Q_2 = 0$. And $D_4 = Q_3 = 0$. After the first pulse arrives, date (D) in all the flip-flops shifts to Qs and the state of the regular read from left to rights becomes A_1, 000. After the second pulse, the state of the register becomes A_2, A_1 00 and finally at the end of fourth pulse it is A_4, A_3, A_2, A_1. The register state can be read simultaneously at $B_4\ B_3\ B_2\ B_1$. The register is named *serial-in, parallel-out* shift register.

By putting in four more pulses, the output can be read serially at B_1.

10.7.2 Binary Counter

A counter is a sequential circuit that counts the number of input pulses. A counter that counts in terms of binary is called a binary counter. The count output of an n-bit binary counter is 2^n states. Therefore, it can count from 0 to $(2^n - 1)$. The number of states of a counter is referred to as it modulus (m). For an n-bit counter, $m \leq 2^n$.

Depending on the manner in which flip-flops are triggered to count, there are two types of counters.

- Asynchronous counter
- Synchronous counter

In case of an asynchronous counter, the flip-flops are clocked sequentially, while in a synchronous counter, they are clocked simultaneously. Therefore, time delays of each flip-flop get added and their count action is much slower than in synchronous counters.

✦ Asynchronous Counters (Ripple Counters)

T-flip-flops are used in these counters. A 3-bit binary ripple counter is shown in Fig. 10.26. The small circles at the back of the clock input stand for the fact that these are triggered by the trailing edge of the input pulse (1 going 0). For toggling, all T inputs are kept high (1). From left to right, the first T flip-flop receives pulses from the counter. The other two receive pulses from the output of the preceding T flip-flop. The pulses kind of ripple through and hence, the name ripple counter. The truth table of a 3-bit ripple counter is presented in Table 10.18.

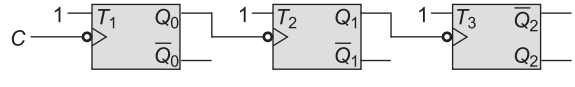


Fig. 10.26 3-bit (modules 8) ripple counter

Table 10.18 *Truth table of 3-bit ripple counter*

Q_2	Q_1	Q_0	Input pulse count
0	0	0	0
0	0	1	1
0	1	0	2
0	1	1	3
1	0	0	4
1	0	1	5
1	1	0	6
1	1	1	7
0	0	0	8

At 8^{th} pulse, the counter resets to 0.

The timing diagram of the counter is presented in Fig. 10.27. It is seen that pulse frequency gets divided by 2 at each stage.

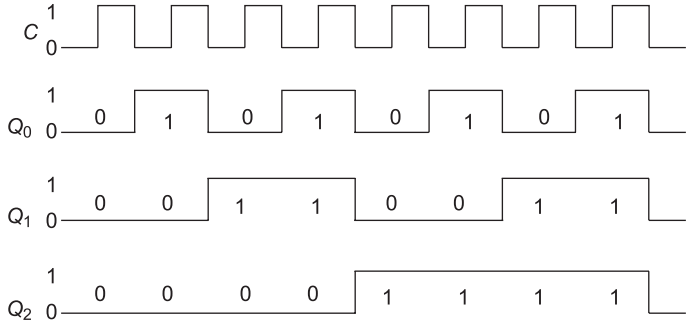


Fig. 10.27 Timing diagram of a 3-bit ripple counter

❑ ***Modulo-5 Ripple Counter*** As $2^3 = 8 > 5$, we need 3 *T* flip-flops. The counter circuit is drawn in Fig. 10.28, where the CLR (clear) terminals are also shown. As the pulse count reaches

$5 \rightarrow 1\ 0\ 1$

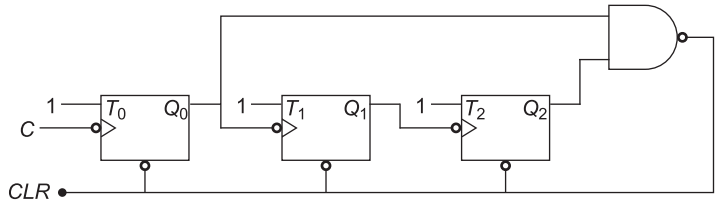


Fig. 10.28 Modulo 5-ripple counter

The counter must reset to zero (000). As outer 1's combination is unique to count 5, these are fed to a NAND gate, which produces 0 output to clear the three flip-flops. The counter is now ready for 0 recount.

✦ Synchronous Counter

In an asynchronous counter, as the pulse ripples through all the flip-flops, their time delay adds up. In synchronous counters, the clock pulse is applied to all the flip-flops simultaneously and so the time delay is that of one counter. Therefore, these are fast in operation.

For a modulo-16 synchronous counter, four-flip-flops are needed. We shall use the T flip-flop.

The truth table for output sequence for Q_3, Q_2, Q_1, Q_0 is listed in Table 10.19. We find from this table that

Q_0 continuously toggles, therefore $T_0 = 1$

Q_1 toggles when $Q_0 = 1$, therefore $T_1 = Q_0$

Q_2 toggles when $Q_0 = Q_1 = 1$, therefore $T_2 = Q_1Q_0$

Q_3 toggles when $Q_0 = Q_1 = Q_2 = 1$, therefore $T_3 = Q_2\, Q_1\, Q_0$.

Table 10.19 *Truth table*

Input (Clock)	Q_3	Q_2	Q_1	Q_0	T_3	T_2	T_1	T_0
0	0	0	0	0	0	0	0	1
1	0	0	0	1	0	0	1	1
2	0	0	1	0	0	0	0	1
3	0	0	1	1	0	1	1	1
4	0	1	0	0	0	0	0	1
5	0	1	0	1	0	0	1	1
6	0	1	1	0	0	0	0	1
7	0	1	1	1	1	1	1	1
8	1	0	0	0	0	0	0	1
9	1	0	0	1	0	0	1	1
10	1	0	1	0	0	0	0	1
11	1	0	1	1	0	1	1	1
12	1	1	0	0	0	0	0	1
13	1	1	0	1	0	0	1	1
14	1	1	1	0	0	0	0	1
15	1	1	1	1	1	1	1	1
0	0	0	0	0				

The corresponding $T = 1$ are indicated in the same table. To generate T_2 and T_3, two AND gates are needed. To meet these requirements, the circuit diagram of the synchronous counter is drawn in Fig. 10.29.

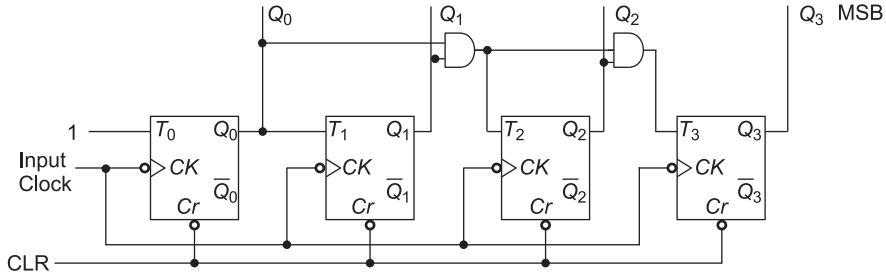


Fig. 10.29 Modulo-16 synchronous counter

10.8 GENERAL MODEL OF A SEQUENTIAL CIRCUIT

The circuit shown in Fig. 10.30 gives a general idea of a sequential circuit. During one clock cycle,

1. Flip-flops (may) change state,
2. Changes propagate through combinational logic, and
3. Combinational logic gates and flip-flop inputs become stable.

This cycle repeats indefinitely as long as the clock is running.

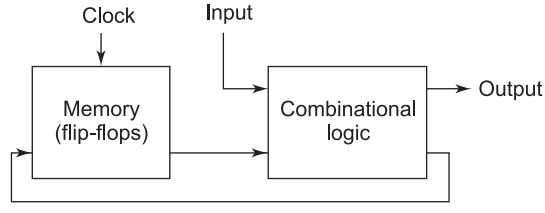


Fig. 10.30 Model sequential circuit

SUMMARY

- Number systems have been introduced. Digital and analog systems are compared.
- Digital circuits are classified. Sequential circuits have been introduced and discussed.

EXERCISES

Review Questions

1. Convert the given decimal numbers to binary:
   (i) $(258)_{10}$ (ii) $(137)_{10}$ (iii) (11.6875
2. Convert the hexadecimal $(8F6)_{16}$ to a decimal number.
3. Multiply $(11011)_2$ with $(101010)_2$.

4. Divide $(8EC2)_{16}$ by $(2F2)_{16}$.
5. What is the prime motivation for using Boolean algebra to simplify logical expressions?
6. What is the only input combination that will produce a HIGH at the output of a five-input AND gate?
7. What logic level should be applied to the second input of a two-input AND gate if the logic signal at the first input is to be inhibited from reaching the output?

➥ Problems

1. Find the decimal equivalents of the following binary numbers.
   (a) 1111111 (b) 11001.0101
2. Perform the following operations.
   (a) 10101.10101 + 110011 (b) 01110.1001 – 00011.1110
   Use 2's complement method.
3. Subtract in binary form:
   $(47)_{10} - (23)_{10}$
4. Express the following in 8-bit sign magnitude form:
   (a) +123 (b) –56
5. Decode the following into decimal form (H stands for Hexadecimal):
   (a) FCH (b) 9 AH
6. Add the following decimal numbers in 8-bit 2's complement form:
   (a) +45 –56 (b) + 67 –98
7. Find the 8-bit subtraction of the following decimal number in 2's complement form:
   (a) +54, +65 (b) –25, –66
8. Convert the following sign-magnitude numbers into decimal form:
   (a) 1001100110 (b) 100 1100, 01110101
9. Perform the following conversions:
   (a) 279 from decimal to octal number.
   (b) 36.125 from decimal to octal number.
   (c) 11000111.1101 from binary to octal number.
10. Convert the numbers given in Problem 9 to hexadecimal numbers.
11. Perform the indicated operations in binary:
    (a) $(32)_8 + (73)_8$ (b) $(175)_8 - (114)_8$ (c) $(7E)_{16} + (AD)_{16}$ (d) $(BC)_{16} - (F4)_{11}$
12. Encode the decimal numbers 43 and 295 in binary form:
    (a) Binary code (b) BCD code (c) Octal code (d) Hexadecimal code
13. Check the identity by truth table: $\overline{A + B} \pm \overline{A} \cdot \overline{B}$
14. Show that $(A + B) \cdot (A + \overline{B}) = A$
15. Using 2's complement, find
    1110.1001 – 0011.1110
16. Convert 1100 0111, 1101 to octal number.
17. Consider the Boolean expression
    $Y = AB + CD + E$
    Implement it using NAND gates
    Hint: Take complement, apply De Morgan's theorem and take complement again.
18. Design a circuit which can give AND operation of two variables with only a NAND gate.
19. Write the Boolean expression for the given circuit in Fig. 10.31; E = ?

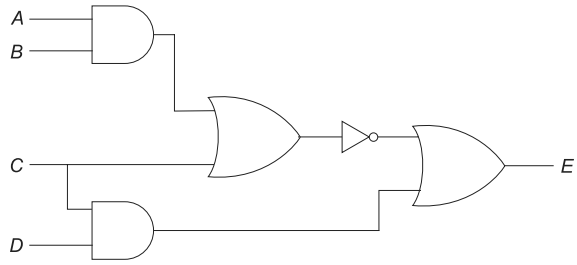


Fig. 10.31

20. For the logic circuit of Fig. 10.32, write the Boolean expression for Y in simple form.

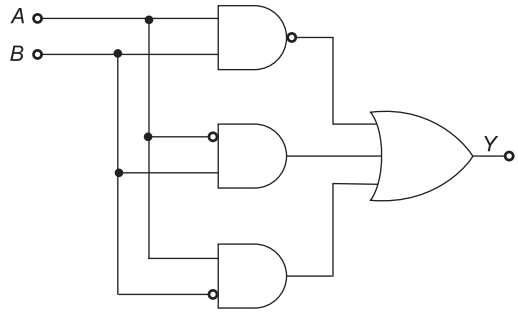


Fig. 10.32

21. Realise the following logic function using NAND gates only:

$Y = \overline{AB + \bar{A}\bar{B}} + \bar{A}B$

Hint: Use De Morgan's theorem.

22. Probe the following identities:

(a) $A + \bar{A}B = B$

(b) $(A + B)(\bar{A} + C) = AC + \bar{A}B$

(c) $(A + C)(A + D)(B + C)(B + D) = AB + CD$

23. Draw the logic circuit for

$Y = A\bar{B}C + ABC$

And then simplify and draw the simplified circuit.

24. Simplify the Boolean equation and then draw its logic circuit with appropriate gates.

$Y = (\bar{A} + B + C)(A + B + \bar{C})$

25. Write the truth table for the given circuit:

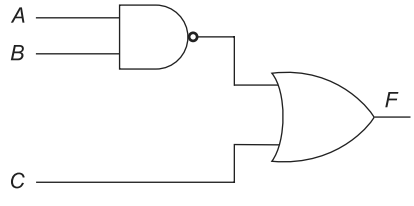


Fig. 10.33

26. Draw the output waveform for the following waveform input to the *SR* flip-flop.

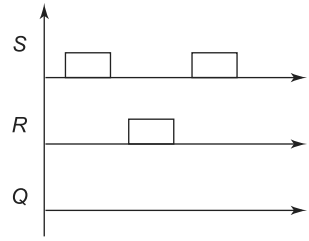


Fig. 10.34

27. Implement the 4:1 multiplexer using basic gates.

28. In the sequential circuit of Fig. 10.15, the NOR gates are replaced by NAND gates. Check that its operation agrees with the *SR* flip-flop truth table (Table 10.15).
Hint: In the NAND gate, the output is 0 only if both inputs are 1.

29. In an *SR* flip-flop (Fig. 10.19), *S* is connected to *R* through NOT gate and input (1/0) is given to *S*. Using the truth table 10.15, show that it acts as a *D*-flip-flop. Identify the output terminal.

30. A 1-to-4 de-multiplexer with one input, two selection lines and four outputs is given in Fig. 10.35.
Find the outputs (D_1 to D_4) for all possible combinations of selector inputs Present it in the form of a truth table. What conclusion can you draw?

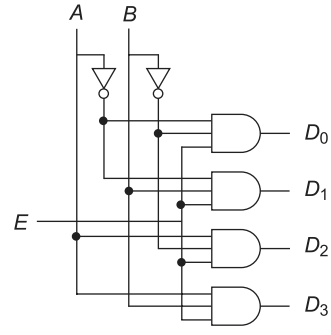


Fig. 10.35

31. For the de-multiplexer of Fig. 10.35, '*E*' is held at 1. Prepare a truth table linking *AB* to output. Is it a decoder. If *E* = 0, what would be the outputs? Are these dependent on *A*, *B*?

32. For the logic circuit of Fig. 10.36, prepare the truth table for various combinations (except *R* = *S* = 1) linking to Q_n and Q_{n+1}.

33. For the multiplexer of Fig. 10.37, prepare the truth table for *AB* and F. Write the logic expression for F for input (0, 1).
Hint: $F = \bar{A}$, $BI_2 = A\bar{B}\bar{C}$

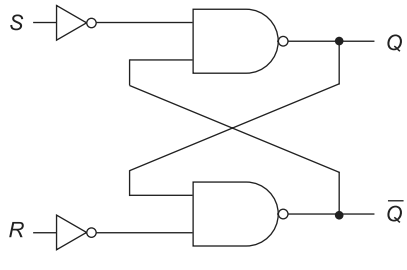


Fig. 10.36

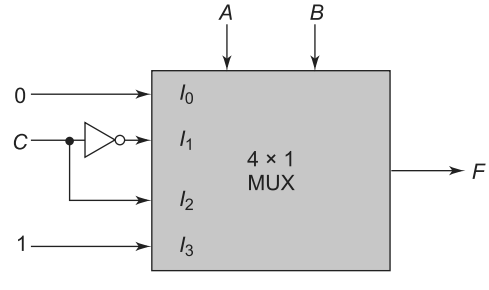


Fig. 10.37

For all possible inputs, all outputs are connected by OR. Thus, $Y = \bar{A}\bar{B}I_0 + \bar{A}BI_1 + A\bar{B}I_2 + A\bar{B}I_3$ or $Y = \bar{A}B\bar{C} - A\bar{B}C + AB$

➥ Multiple-Choice Questions

1. Binary representation of the decimal number 25 is
(a) 10001 (b) 11001 (c) 11101 (d) 10110

2. 2's complement representation is obtained by
(a) complementary the binary representation two times
(b) subtracting 1 from the 1's complement representation
(c) adding 1 to the 1's complement representation
(d) none of the above

3. According to De Morgan's theorem,
(a) $\overline{A+B} = \overline{A} \cdot \overline{B}$ (b) $\overline{A+B} = \overline{A} + \overline{B}$ (c) $\overline{A \cdot B} = \overline{A} \cdot \overline{B}$ (d) $\overline{A+B} = A + B$

4. In a combinational circuit the output
(a) depends upon the past output
(b) depends upon the present input and past output
(c) depends upon the combination of present input only
(d) does not depend upon any present or past input

5. A multiplexer is
(a) one-to-many converter (b) many-to-one converter
(c) a sequential circuit (d) all of these

6. The race-around condition is avoided by using
(a) *JK* flip-flop (b) *D* flip-flop
(c) *SR* flip-flop (d) master-slave *JK* flip-flop

7. Asynchronous counters are also known as
(a) ring counters (b) Johnson counters
(c) BCD counters (d) ripple counters

8. The sequential circuits are circuits whose output depends upon
(a) the present combination of input and past output
(b) on past output only
(c) on present input only
(d) none of the above

9. ABCD counter is
(a) Mod-10 counter (b) a combinational circuit
(c) Mod-12 counter (d) all of these

10. In toggling,
(a) the output remains same (b) the state changes occur
(c) the output is invalid (d) none of the above

11. According to De Morgan's theorem: BAR(X + Y) = BAR(X).BAR(Y). This means logically there is no difference between
(a) a NOR and an AND gate with inverted inputs
(b) a NAND and an OR gate with inverted inputs
(c) a AND and a NOR gate with inverted inputs
(d) a NOR and a NAND gate with inverted inputs

12. An AND gate with schematic "Bubbles" on its inputs performs the same function as _____ gate.
(a) NOT (b) OR
(c) NOR (d) NAND

13. Which Boolean algebra property allows us to group operands in an expression in any order without affecting the results of operation [for example, A + B = B + A]?

(a) Associative (b) Commutative
(c) Boolean (d) Distributive

ANSWERS

✦ Problems

1. (a) $(127)_{10}$ (b) $25375)_{10}$
2. (a) 111100.00001 (b) $(10.6875)_{10}$
3. 00011000
5. (a) $15 \times 16^1 + 12 \times (16)^\circ = 245$, (b) $9 \times 16^1 + 10 \times 16^\circ = 156$
9. (a) $(427)_8$ (b) $(44.1)_8$ (c) (307.64)
10. (a) $(117)_{16}$ (b) $24.2)_{16}$ (c) $(C7.D)_{16}$
11. (a) 1010.101 (b) 00011000 (c) 100101011 (d) $-(38)_{16}$
12. (a) 00101011 (b) 010.10, 0011 (c) 53 (d) 2B
13. (a) 0000000100100111
(b) 01001001.0101 (c) $1.447)_8$ (d) $(127)_{16}$
17. $(307.64)_8$
21. $\overline{AB}$
33. *SR* flip-flop

✦ Multiple-Choice Questions

1. (b) **2.** (c) **3.** (b) **4.** (c) **5.** (b) **6.** (d) **7.** (d) **8.** (a) **9.** (a) **10.** (b)
11. (a) **12.** (c) **13.** (b)

CHAPTER 11

Boolean Algebra and Combinational Circuits

GOALS AND OBJECTIVES

- Explanation of Boolean algebra
- Presentation of combinational circuits
- Discussion of binary functions

11.1 INTRODUCTION

Basic logic gates, their associated Boolean algebra, etc. were the subject matter of Chapter 10. The combination of logic gates to perform a certain logic function is known as a *combinational circuit*. The desired logic may be specified in the form of a truth table from which various Boolean logic functions can be written down and converted to reduced forms needing the least number of gates or even specified gates (say NAND gates only) for their implementation.

All these techniques and also the specified binary functions like addition, subtraction will be taken up in the chapter.

In order to prepare the reader for the above task, the chapter begins by discussing binary and other number systems employed in digital systems. Further details of Boolean algebra, basic theorems and function reduction techniques will then be exposed to the reader.

Memory cells (flip-flops) which can also be constructed from basic gates and the sequential circuits, which employ memory cells, will be the subject matter of the next chapter.

11.2 BINARY NUMBER SYSTEM

All of us are familiar with the decimal number system in which the symbols 0 through 9, known as *digits*, are used to specify any number. For example, the number 496.27 is nothing but

$$400 + 90 + 6 + 2/10 + 7/100$$

or $$4 \times 10^2 + 9 \times 10^1 + 6 \times 10^0 + 2 \times 10^{-1} + 7 \times 10^{-2}$$

The *base* of this number system is 10 and each position to the left or right of the decimal point corresponds to a power of 10.

When we speak of digital systems like computers, microprocessors, etc. it is the binary number system, which is employed. The word *binary* means two. As mentioned in the previous chapter, a digital system consists of only two distinct logic levels, 0 and 1. These are the two digits in the binary number system. Each digit is called a bit. Thus, binary numbers are merely strings of 0s and 1s. A string or sequence of 4 bits is called a *nibble*, an 8-bit sequence is a *byte* and a 16-bit sequence is a *word*. The binary number system, like decimal number system assigns a specific weight in n of 2. For example, the binary number 11011.101 can be written as

$$1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}$$

Here, the *base* is 2 and each digit corresponds to a power of 2.

11.2.1 Number Conversion

✦ Binary-to-decimal Conversion

To convert a binary number to its decimal equivalent, add the decimal equivalent of each position occupied by a 1.

For example,

$$\begin{aligned}(111001)_2 &= 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 2^5 + 2^4 + 2^3 + 0 + 0 + 1 \\ &= 32 + 16 + 8 + 1 = (57)_{10} \\ (101.01)_2 &= 2^2 + 0 + 2^0 + 0 + 2^{-2} \\ &= 4 + 1 + 0.25 = (5.25)_{10}\end{aligned}$$

The subscripts 2 and 10 respectively identify the base of binary and decimal number systems.

✦ Decimal-to-binary Conversion

A decimal number can be converted to its equivalent binary form by the inverse process, i.e. by expressing the decimal number as the sum of powers of 2. The *double-dabble* is a very popular method for decimal-to-binary conversion, in which the integers and the decimals are handled separately. It can be summarised as follows:

- To convert a decimal number to its binary equivalent, progressively divide the decimal number by 2, noting the remainders; the remainders taken in reverse order form the binary equivalent.
- To convert a decimal fraction to its binary equivalent, progressively multiply the fraction by 2, removing and noting the carries; the carries taken in forward order form the binary equivalent.

The double-dabble procedure is illustrated below:

EXAMPLE 11.1 Convert the decimal number 25.375 to its binary equivalent.

Solution Using double-dabble method on the integer part,

$$
\begin{array}{rl|l}
2 & 25 & \\
2 & 12 & 1 \uparrow \\
2 & 6 & 0 \\
2 & 3 & 0 \\
 & 1 & 1
\end{array}
\quad \text{remainders—read up}
$$

Therefore, the binary equivalent of 25 is 11001. Now consider the fraction,

$$
\begin{array}{rl}
0.375 \times 2 = 0.75 & 0 \\
0.75 \times 2 = 1.5 & 1 \\
0.5 \times 2 = 1.0 & 1 \\
0.0 \times 2 = 0 & 0 \downarrow
\end{array}
\quad \text{carries—read down}
$$

The binary equivalent of 0.375 is 0110 = 011. Therefore, the binary equivalent of 23.375 is 11001.011.

11.2.2 Binary Arithmetic

We are all familiar with the arithmetic operations like addition, subtraction, multiplication and division using decimal numbers. Such operations can be performed on binary numbers also and in fact, binary arithmetic is much simpler than decimal arithmetic, because here only two digits 0 and 1 are involved. We shall first take up the rules of binary addition before discussing other operations.

✦ Binary Addition

The rules for binary addition are

$$0 + 0 = 0$$

$$0 + 1 = 1$$

$$1 + 0 = 1$$

$$1 + 1 = 10$$

$$1 + 1 + 1 = 1 + 10 = 11$$

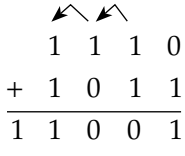
It is important to note that the sign "+" used here corresponds to arithmetic addition and not logical operation.

For large numbers, we add column by column, carrying where necessary into higher position columns.

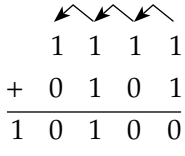
EXAMPLE 11.2 Add the binary numbers: (a) 1110 and 1011 (b) 1111 and 0101

Solution

(a)



(b)



Arrows indicate carry operation.

✦ Binary Subtraction

The rules of binary subtraction are

$$0 - 0 = 0$$

$$1 - 0 = 1$$

$$1 - 1 = 0$$

$$10 - 1 = 1$$

In subtraction, we subtract column by column, borrowing wherever necessary from higher position columns. In subtracting a large number from a smaller one, we can subtract the smaller from the larger and change the sign just as we do with decimals.

EXAMPLE 11.3 Subtract (a) 10 from 28, and (b) 11 from 6 using binary arithmetic.

Solution

(a)

28	1 1 1 0 0
–10	–0 1 0 1 0
18	1 0 0 1 0

(b)

6	0 1 1 0
–11	–1 0 1 1
–5	–0 1 0 1

Here, we subtract 6 from 11 and simply add a '–' sign to it.

11.2.3 Signed Numbers

In the decimal number system, we use '+' sign for denoting positive numbers and '–' sign for denoting negative numbers. (The absence of any sign indicates that the number is positive). As these signs are not available in the binary system, we shall see below how signs are attached to numbers.

There are three binary signed number systems. These are

1. Sign-magnitude representation
2. One's complement representation
3. Two's complement representation

✦ Sign Magnitude

In sign-magnitude representation, the most significant bit (MSB) is used to represent the sign (0 for positive and 1 for negative) and the remaining bits are used to represent the magnitude of the number. For example, the binary 6-bit number 011010 represents a positive number and its value is 26, whereas 111010 represents a negative number written as –26. In general, the maximum positive number that can be represented using sign-magnitude form is $+(2^{n-1} - 1)$ and the maximum negative number that can be represented is $-(2^{n-1} - 1)$, where n is the number of bits.

Unsigned n-bit numbers are 2^n

$$\text{Ratio}\left(\frac{\text{Signed}}{\text{Unsigned}}\right)\text{numbers} = \frac{2^{n-1}-1}{2^n} \approx \frac{1}{2}$$

It is seen that signed numbers are nearly ½ of unsigned numbers. Therefore, signed numbers should be used only when necessary.

For instance, the consider 8-bit number.

Unsigned range

00000000 → 0

11111111 → 255

Signed range

These reduce from 255 to 127.

Negative numbers

10000001 → –1

11111111 → –127

Positive numbers

00000001 → +1

01111111 → +127

✦ One's Complement Notation

In a binary number, if we replace each 0 by 1 and each 1 by 0, we obtain another binary number, which is the one's complement of the first binary number. In fact, both the numbers are complements of each other. For example, while $(0111)_2$ represents $(+7)_{10}$, $(1000)_2$ represents (–7). Here again, we note that the MSB denotes the sign of the number (0 for positive and 1 for negative). But, the magnitude of a negative number is obtained by taking the one's complement of the number. Thus, the maximum positive and negative numbers, that can be represented are = $(2^{n-1} - 1)$ and $-(2^{n-1} - 1)$ respectively. The major disadvantage here is that the number zero has two different representations, 0000 = + 0 and 1111 = – 0.

✦ Two's Complement Notation

By adding a 1 to the one's complement of a binary number, we get the two's complement of that binary number. This can also be used for representing negative numbers. For example, 0101 represents + 5, whereas its 2's complement 1011 (1's complement +1) represents –5. In this representation also, if the MSB is 0, the number is positive, whereas if the MSB is 1, the number is negative and its magnitude is obtained after taking its 2's complement. Here, for an n-bit number, the maximum positive number that can be represented is + $(2^{n-1} - 1)$ and the maximum negative number that can be represented is $-(2^{n-1})$. The reader may verify this by considering a 4-bit number.

✦ Binary Subtraction using Two's Complement

Two's complement representation is usually preferred over other representations because of the ease in binary subtraction using two's complement. It is based on the fact that adding the 2's complement of a number is equivalent to subtracting the number. For example, in evaluating $(A - B)$ (A and B are binary numbers of course), we subtract B from A, adding the 2's complement of B to A. If a final carry is generated, it is discarded and the answer is given

by the remaining bits, which is positive (A is greater than B). If the final carry is 0, the answer is negative (B is greater than A) and is in the 2's complement form.

EXAMPLE 11.4

(a) Represent $(-17)_{10}$ and $(-22)_{10}$ in 8-bit 2's complement notation.

(b) Perform (22 – 17) and (17 – 22) directly by 2's complement notation.

(c) Subtract (i) $(1010)_2$ from $(1111)_2$, (ii) $(111001)_2$ from $(101011)_2$, and (iii) (11010.10) from (10110.01) using 1's complement.

(d) Subtract (i) (011011) from $(100011)_2$, and (ii) $(101111)_2$ from $(10111)_2$ using 2's complement.

Solution

(a) By adding 1 to the one's complement, we get

$-17 \rightarrow -10001 \rightarrow -00010001 \rightarrow 11101110 + 1$

$\rightarrow 11101111$

$-22 \rightarrow -10110 \rightarrow -00010110 \rightarrow 11101001 + 1$

$\rightarrow 11101010$

Thus, 11101111 is the 2's complement representation of –17 in 8 bits. Similarly, 11101010 is the 2's complement representation of – 22. One important thing that can be observed is that the 2's complement is same as the number itself except that while scanning from LSB to MSB, we find that after the occurrence of the first 1, all the bits have been complemented. Based on the above observation, we can use the following rule for finding the 2's complement of a binary number: Scan the number from LSB to MSB and write the bits as they are up to and including the occurrence of the first 1 and complement all other bits.

(b) (i) 22 = 00010110

17 = 00010001

2's complement of 17 = 11101111

22	00010110	
–17	11101111	
+5	100000101	$\rightarrow (+5)_{10}$

Discard 1

MSB is 0. Therefore, the answer is positive.

(ii) 2's complement of 22 = 11101010

17	00010001
–22	11101010
–5	11111011

No carry.

MSB is 1. Therefore, the answer is negative and is in 2's complement form. Two's complement of 11111011 = 00000101 = $(5)_{10}$. Therefore, the answer is –5 in decimal representation.

(c) (i) $(1111)_2 - (1010)_2$

Taking 1's complement 1010 = 0101

By adding, we get

$$\begin{array}{r} 1111 \\ +0101 \\ \hline \textcircled{1}0100 \\ +1 \\ \hline 0100 \\ +1 \\ \hline 0101 \end{array}$$

Carry ← (1 added back)

⇒ Answer

(ii) $(101011)_2 - (111001)_2$

Taking 1's complement 111001 = 000110

By adding, we get

$$\begin{array}{r} 101011 \\ +000110 \\ \hline 110001 \end{array}$$

No carry →

Taking complement of 110001 and negative sign

⇒ –001110

or –1110 Answer

(iii) 10110.01 – 11010.10

Complement of 11010.10 = 00101.01

By adding, we get

$$\begin{array}{r} 10110.01 \\ 00101.01 \\ \hline 11011.10 \end{array}$$

No carry →

By taking complement of 11011.10 with negative sign

⇒ –00100.01

or –100.01 Answer

(d) (i) $(100011)_2 - (011011)$

Taking 2's complement of 011011, we get

$$\begin{array}{r} 100100 \\ 1 \\ \hline 100101 \end{array}$$

By adding,

$$\begin{array}{r} 100011 \\ 100101 \\ \hline 101000 \end{array}$$

No carry ↑

Answer: 01000 ⇒ 8

(ii) $(10111)_2 - (101111)_2$

Taking 2's complement of 101111, we get

$$\begin{array}{r} 010000 \\ +\quad\ 1 \\ \hline 010001 \end{array}$$

By adding,

$$\begin{array}{r} 10111 \\ +010001 \\ \hline 101000 \end{array} \quad \text{No carry} \rightarrow$$

Result is 2's complement of 101000 with negative sign, i.e.

$$\begin{array}{r} 010111 \\ +1 \\ \hline -011000 \end{array}$$

$= -24$ Answer

✦ Outflow

In 8-bit (or in general n-bit) arithmetic, it must the ensured that the result does not go beyond the range of numbers, i.e. it should be within the range of –127 to +127. If it exceeds this range, an outflow will occur into the sign bit, thereby changing it to incorrect sign. Two cases are illustrated below.

❑ ***Case I*** Let us add + 100 and + 50; the answer should be +150.

$$\begin{array}{rrr} 100 & 0110 & 0100 \\ +\ 50 & +0011 & 0010 \\ \hline +150 & 1001 & 0110 \end{array}$$

It is seen that sign bit has changed from 0 to 1, which means that the sum is negative.

❑ ***Case II*** Let us add two negative numbers.

$$\begin{array}{rrr} -85 & 1010 & 1011 \\ +(-97) & +1001 & 1111 \\ \hline & 1\ 0100 & 1010 \rightarrow 0100\ \ 1010 \end{array}$$

carry ignore ↑ ↑ positive sign

The sum has a positive sign but it should be negative.

✦ Binary Multiplication

Binary multiplication is similar to decimal multiplication. In binary multiplication, say $A \times B$, each partial product is either 0 or A itself (i.e. the multiplicand) as the multiplication is either by 0 or by 1 respectively. An example of binary multiplication is given below.

EXAMPLE 11.5 Multiply 1101 by 1010.

Solution

$$\begin{array}{rl} 1101 & \text{Multiplicand} \\ \times 1010 & \\ \hline 0000 & \\ 1101 & \text{Partial products} \\ 0000 & \\ 1101 & \\ \hline 10000010 & \text{Product} \end{array}$$

✦ Binary Division

Binary division is again performed in a way very similar to decimal division. An example of binary division is given below.

EXAMPLE 11.6 Divide 1011011 by 111.

Solution

$$\begin{array}{l} 111\overline{)1011011}\ 1101 \leftarrow \textit{Ans.} \\ \quad 0111 \\ \quad \overline{1000} \\ \quad\ 0111 \\ \quad \overline{000111} \\ \quad\quad\ 111 \\ \quad \overline{0} \end{array}$$

11.3 OCTAL NUMBER SYSTEM

Another popular number system is the octal number system. There are $8(2^3)$ combinations of 3-bit binary numbers. Therefore, sets of 3-bit binary numbers can be conveniently represented by octal numbers with base 8. These numbers are 0, 1, 2, 3, 4, 5, 6 and 7. This also is a positional number system and has two parts, integer and fractional. For example, $(1062 \cdot 403)_8$ is an octal number and can be written as

$$(1062 \cdot 403)_8 = 1 \times 8^3 + 0 \times 8^2 + 6 \times 8^1 + 2 \times 8^0 + 4 \times 8^{-1} + 0 \times 8^{-2} + 3 \times 8^{-3}$$
$$= 512 + 482 + 4/8 + 3/512$$
$$= (562 \cdot 50586)_{10}$$

The above procedure gives the decimal equivalent of an octal number.

EXAMPLE 11.7 Convert $(294 \cdot 6875)_{10}$ into octal.

Solution First, consider the integer part.

8 | 294		
8 | 36	6	↑
8 | 4	4	remainders—read up
0	4	

Therefore, $(294)_{10} = (446)_8$

Now, coming to the fraction

$0.6875 \times 8 = 5.50$	5	carries—read down
$0.5000 \times 8 = 4.00$	4	↓

Therefore, $(0.6875)_{10} = (54)_8$

From the above, it follows that $(294 \cdot 6875)_{10} = (446 \cdot 54)_8$

Octal-to-Binary and Binary-to-Octal Conversion

Octal numbers can be converted to equivalent binary numbers by replacing each digit by its 3-bit binary equivalent. Table 11.1 gives octal and binary equivalents for decimal numbers 0-7. For example,

$$(642 \cdot 71)_8 = (110100010 \cdot 111001)_2$$

Similarly, binary numbers can be converted into equivalent octal numbers by making groups of 3 bits starting from LSB and moving towards MSB for integer part. For example,

$$(101110011)_2 = 101110011 = (563)_8$$

For fractional parts, we start grouping from the bit next to the binary point and move towards right. For example,

$$(0.101010110)_2 = 0.101010110 = (0.526)_8$$

In forming the 3-bit groupings, sometimes, we may have to add 0's to complete the most significant digit group in the integer part and the least significant digit group in the fractional part.

Table 11.1 *Binary and octal equivalents of decimal numbers*

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

✦ Octal Arithmetic

Octal arithmetic rules are similar to the decimal or binary arithmetic. But in computing system's arithmetic operations, using octal numbers are performed by converting the octal numbers to binary numbers and then by using the rules of binary arithmetic. The result so obtained can be converted again to octal form.

11.4 | HEXADECIMAL NUMBER SYSTEM

Hexadecimal numbers are extensively used in association with microprocessors. Hexadecimal means 16. There are 16 combinations of 4-bit binary numbers and sets of 4-bit binary numbers can be entered in the microprocessor in the form of hexadecimal digits. The base (or radix) of a hexadecimal number is 16. This means that it uses 16 symbols to represent all numbers. These are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, *A*, *B*, *C*, *D*, *E*, *F*. Since both numbers as well as alphabets are used to represent the digits in a hexadecimal number system, it is also called the alphanumeric number system.

Table 11.2 shows the equivalences between hexadecimal, binary and decimal digits.

Table 11.2 *Binary and hexadecimal equivalences of decimal numbers*

Decimal	Binary	Hexadecimal
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7

(*Contd.*)

8	1000	8
9	1001	9
10	1010	*A*
11	1011	*B*
12	1100	*C*
13	1101	*D*
14	1110	*E*
15	1111	*F*

Since 16 digits are used, the weights are in powers of 16. The decimal equivalent of a hexadecimal string equals the sum of all hexadecimal digits multiplied by their weights. For example,

$$\begin{aligned}(F8E.\,B8)_{16} &= F \times 16^2 + 8 \times 16^1 + E \times 16^0 + 2 \times 16^{-1} + B \times 16^{-2} \\ &= 15 \times 16^2 + 8 \times 16^1 + 14 \times 16^0 + 2 \times 16^{-1} + 11 \times 16^{-2} \\ &= 3840 + 128 + 14 + 2/16 + 11/256 \\ &= (3982.16796875)_{10}\end{aligned}$$

Conversion from decimal to hexadecimal is similar to the procedure used in binary and octal conversion except that, here 16 is used in dividing for integer part and multiplying for fractional part.

EXAMPLE 11.8 Convert the following numbers to their hexadecimal equivalents.

(a) $(49.5)_{10}$ (b) $(972.625)_{10}$

Solution

(a) Integer part

$$\begin{array}{r|l} 16 & 49 \\ 16 & 3 \quad 1 \\ 0 & 3 \end{array}$$

Thus, $(49)_{10} = (31)_{16}$

Fractional part

$$0.5 \times 16 = 8.0$$

Thus, $(0.5)_{10} = (0.8)_{16}$

Therefore, $(49.5)_{10} = (31.8)_{16}$

(b) Integer part

$$\begin{array}{r|ll} 16 & 972 & \\ 16 & 60 & 12 = C \\ 16 & 3 & 12 = C \\ 0 & 3 & \end{array}$$

Thus, $(972)_{10} = (3CC)_{16}$

Fractional part

$0.625 \times 16 = 10.00$. The integral part is $10 \rightarrow A$.

Thus, $(0.625)_{10} = (0.A)_{16}$

Therefore, $(972.625)_{10} = (3CC.A)_{16}$

11.4.1 Hexadecimal-to-Binary and Binary-to-Hexadecimal Conversions

Hexadecimal numbers can be converted into equivalent binary by replacing each hexadecimal digit by its equivalent 4-bit binary number. For example,

$$(20E.CA)_{16} = (0010\ 0000\ 1110.1100\ 1010)_2$$
$$= (001000001110.11001010)_2$$

Similarly, binary numbers can be converted into hexadecimal numbers by making groups of four bits starting from LSB and moving towards MSB, for integers, and then replacing each group of four bits by its hexadecimal equivalents. Sometimes, in forming 4-bit groupings, 0's may be required to complete the most significant digit group in the integer part. For example,

$$(10100110111110)_2 = (0010\ 1001\ 1011\ 1110)_2$$
$$= (29BE)_{16}$$

For the fractional part, the above procedure is repeated starting from the bit next to the hexadecimal point and moving towards the right. Here again, in forming 4-bit groupings, 0's may be required to complete the least significant digit group. For example,

$$(0.00111110111101)_2 = (0.0011\ 1110\ 1111\ 0100)_2$$
$$= (0.3EF4)_{16}$$

11.4.2 Hexadecimal Arithmetic

The rules for hexadecimal arithmetic operations are similar to the rules for decimal, binary and octal systems. Since information can be handled only in binary form by a digital circuit, hexadecimal numbers are first converted into binary numbers and arithmetic operations are performed using rules for binary arithmetic.

11.5 CODES

Computers and other digital systems are required to handle data, which may be numeric, alphabets or special characters. Since digital circuits work only with binary digits; therefore, the numerals, alphabets and other special characters are to be converted into binary format. There are several ways of achieving this, and this particular process is called *encoding*. Numerous codes are in existence and different codes serve different purposes. At this stage, it

is important to realise that a series of 1's and 0's in a digital system may sometimes represent a binary number and at other times represent some other discrete information like alphabets, etc. as specified by a given binary code. Some commonly used binary codes are given in Table 11.3.

Table 11.3 *Binary codes*

Decimal number	Binary	BCD
0	0000	0000
1	0001	0001
2	0010	0010
3	0011	0011
4	0100	0100
5	0101	0101
6	0110	0110
7	0111	0111
8	1000	1000
9	1001	1001
10	1010	
11	1011	
12	1100	
13	1101	
14	1110	
15	1111	

✦ Binary Code

This is obtained by converting decimal numbers to their binary equivalents. The CPUs of computers process only binary numbers.

✦ BCD Code

This is a binary code in which decimal digits 0 through 9 are represented by their binary equivalents using four bits. As the weights in the BCD code are 8, 4, 2, 1, it is also known as 8421 code. For instance, the decimal 257 converts to BCD as follows:

2 5 7
↓ ↓ ↓
0010 0101 0111

Therefore, 0010 0101 0111 is the BCD equivalent of $(257)_{10}$. The reverse conversion is similar. For instance,

1001 1000 0110 = $(986)_{10}$

BCD numbers are very useful for input and output operations in digital circuits. They are used to represent decimal digits in systems like digital calculators, digital voltmeters, digital clocks, electronic counters, etc.

✦ The ASCII Code

To get information into and out of a computer, we need to use some kind of *alphanumeric* code (one for letters, numbers and other symbols). At one time, manufacturers used their own alphanumeric codes, which led to all kinds of confusion. Eventually, the industry settled on an input-output code known as the *American Standard Code for Information Interchange* (ASCII, pronounced ask'-ee). This code allows manufacturers to standardise computer hardware such as keyboards, printers and video displays.

❑ ***Using the Code*** The ASCII code is a 7-bit code whose format is

$$X_6X_5X_4X_3X_2X_1X_0$$

Table 11.4 *ASCII Code*

$X_3X_2X_1X_0$	010	011	100	101	110	111
0000	SP	0	@	P		p
0001	!	1	A	Q	a	q
0010	"	2	B	R	b	r
0011	#	3	C	S	c	s
0100	$	4	D	T	d	t
0101	%	5	E	U	e	u
0110	&	6	F	V	f	v
0111	‘	7	G	W	g	w
1000	(	8	H	X	h	x
1001	)	9	I	Y	i	y
1010	*	:	J	Z	j	z
1011	+	;	K	-	k	
1100	,	<	L	-	l	
1101	–	=	M	-	m	
1110	*	>	N	-	n	
1111	/	?	O	-	o	

where each X is a 0 or a 1. Use Table 11.4 to find the ASCII code for the uppercase and lowercase letters of the alphabet and some of the most commonly used symbols. For example, the table shows that the capital letter A has an $X_6X_5X_4$ of 100 and an $X_3X_2X_1X_0$ of 0001. The ASCII code for A is, therefore,

1000001

For easier reading, we can leave a space as follows:

100 0001 (A)

The letter 'a' is coded as

110 0001 (a)

11.6 BOOLEAN RELATIONS

As mentioned earlier, George Boole invented two-state algebra to solve logic problems. Today, Boolean algebra is the backbone of computer circuit analysis and design.

In other words, it may be stated that to obtain the inverse of any Boolean function, invert all variables and replace all OR's by AND's and vice versa.

The table giving a few identities of Boolean Algebra can be referred to the previous chapter, Table 10.5. The truth tables of AND, OR and NOT operations and DeMorgan's theorem can be referred to Section 10.3 of the previous chapter along with its truth table.

In AND operation, the output is 1 only if both A and B are 1. Otherwise, it is zero.

11.7 ALGEBRAIC SIMPLIFICATION

Sometimes it is possible to simplify the circuit obtained from the sum-of-products equation. One way to do this is with Boolean algebra. Starting with the sum-of-products equation, try to rearrange and simplify the equation to the extent possible using the Boolean rules of Section 11.6. The simplified Boolean equation means a simpler logic circuit. Thus, hardware for designing the circuit can be minimised.

The following examples clarify this procedure.

EXAMPLE 11.9 Factorise the following Boolean equations:

(a) $Y = A\overline{B} + AB$

(b) $Y = AB + AC + BD + CD$

(c) $Y = (B + CA)(C + \overline{A}\,B)$

(d) $Y = \overline{A}\,\overline{B}\,\overline{C}\,\overline{D} + \overline{A}\,\overline{B}\,\overline{C}\,D + A\overline{B}\,\overline{C}\,\overline{D} + A\overline{B}\,\overline{C}\,D$

Solution

(a) $Y = A\overline{B} + AB$

$= A\,(\overline{B} + B)$

$= A \cdot 1$

$= A$

(b) $Y = AB + AC + BD + CD$

$= A(B + C) + D(B + C)$

$= (A + D)(B + C)$

(c) $Y = (B + CA)(C + \bar{A}B)$

$= B(C + \bar{A}B) + CA(C + \bar{A}B)$

$= BC + B\bar{A}B + CAC + CA\bar{A}B$; $A\bar{A} = 0$, $BB = B$, $CC = C$

$= BC + \bar{A}B + AC$

(d) $Y = \bar{A}\bar{B}\bar{C}\bar{D} + \bar{A}\bar{B}\bar{C}D + A\bar{B}\bar{C}\bar{D} + A\bar{B}\bar{C}D$

$= \bar{B}\bar{C}(\bar{A}\bar{D} + \bar{A}D + A\bar{D} + AD)$

$= \bar{B}\bar{C}\,[\bar{A}\,(\bar{D} + D) + A(\bar{D} + D)]$

$= \bar{B}\bar{C}\,[\bar{A} \cdot 1 + A \cdot 1]$

$= \bar{B}\bar{C}\,(\bar{A} + A)$

$= \bar{B}\bar{C} \cdot 1 = \bar{B}\bar{C}$

EXAMPLE 11.10 Compare the AND/OR gate realisation of the Boolean function of Example 11.9(b) and its reduced form.

Solution It is immediately observed from Figs 11.1(a) and (b), that the realisation of the original function needs 5 gates (4 AND's and 1 OR), while the simplified function needs only 3 gates (1 AND and 2 OR's). On simplifying original function, no of gates needed for simplification gets reduced.

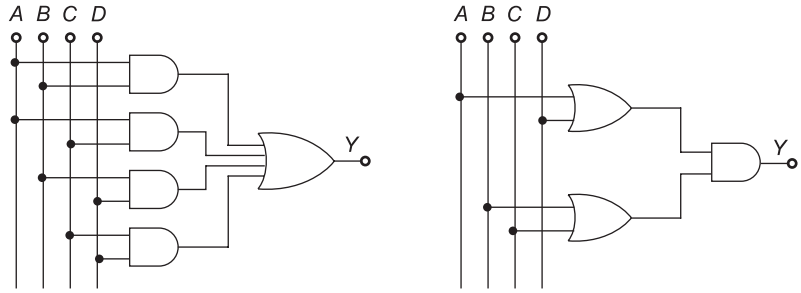


Fig. 11.1 Logic circuits for expression (b) of Example 11.9

EXAMPLE 11.11 Analyse the logic circuit of Fig. 11.2 and show that it can be replaced by a single NAND gate.

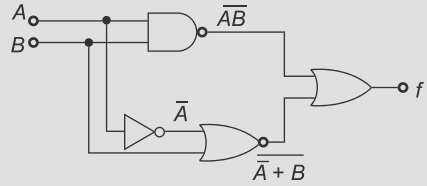


Fig. 11.2 Logic circuit analysis

Solution The sub-outputs are noted on the diagram. The simplification of the function is given below:

$$\begin{aligned} f &= \overline{AB} + \overline{\overline{A} + B} \\ &= (\overline{A} + \overline{B}) + A\overline{B} \quad \text{(De Morgan's law)} \\ &= \overline{A} + \overline{B}\,(1 + A) \\ &= \overline{A} + \overline{B} \cdot 1 \\ &= \overline{A} + \overline{B} \\ &= \overline{AB} \quad \text{(De Morgan's law)} \end{aligned}$$

Thus, the above circuit is simply equivalent to a NAND gate.

11.8 NAND AND NOR IMPLEMENTATION

It is an attractive preposition to realise combinational circuits using NAND/NOR gates (universal gates; see Sections 9.5.3 and 9.5.4 respectively, as only one type of gates would be needed).

11.8.1 NAND Implementation

NAND gates can be used to realise not only any combinational circuit but even sequential circuits can be constructed with three gates as we shall see in Chapter 16. The realisation of NOT, AND and OR using NAND's already discussed in Example 9.2, is illustrated in Fig. 11.3.

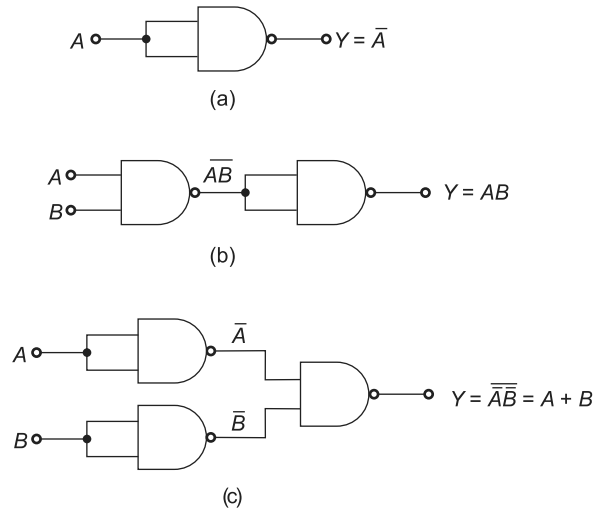


Fig. 11.3 Realisation of (a) NOT, (b) AND, and (c) OR using NAND gates

The implementation of a Boolean function with NAND gates requires that the function be simplified in the sum-of-products form. After this, we apply the De Morgan's theorem. For example, consider the function

$$Y = AB + CD + E$$

$$\overline{Y} = \overline{AB} \cdot \overline{CD} \cdot \overline{E}$$

Therefore,

$$Y = \overline{\overline{AB} \cdot \overline{CD} \cdot \overline{E}}$$

This can be implemented in two levels by using only NAND gates as shown in Fig. 11.4.

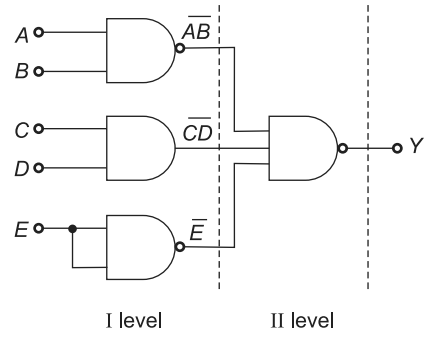


Fig. 11.4 Realisation of $Y = AB + CD + E$

EXAMPLE 11.12 Given the Boolean function

$F = (A + \overline{B})(CD + E)$

Obtain AND-OR implementation and also its implementation using NAND gates only.

Solution We assume that signals corresponding to each literal are available, i.e. the variables are available in complemented as well as in the uncomplemented form.

The realisation of the function using AND and OR gates is shown in Fig. 11.5.

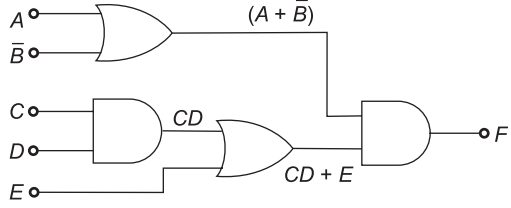


Fig. 11.5 AND/OR implementation

Consider again, the function

$$F = (A + \overline{B})(CD + E)$$

We can write

$$A + \overline{B} = \overline{\overline{A} \cdot B}$$

And similarly,

$$CD + E = \overline{\overline{C}\,\overline{D} \cdot \overline{E}}$$

Therefore,

$$\overline{F} = \overline{\overline{A}B \cdot \overline{C}\,\overline{D} \cdot \overline{E}} = \overline{(A + \overline{B})(C\,D + E)}, \text{ i.e., output of gate (4).}$$

The NAND realisation of the above equation is shown in Fig. 11.6. The output of the NAND gate 4 is $\overline{F}$. So in order to obtain F, we invert the output of gate 4 by using an inverter NAND gate (5).

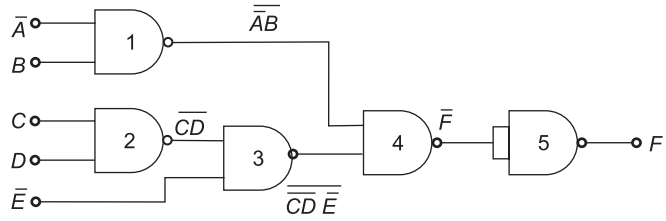


Fig. 11.6 NAND implementation of Example 11.12

Any given function on being expressed in sum of product form can also be realised using NAND gates. Here again, 5 NAND gates are required; interested reader may verify this.

$$\bar{F} = \overline{(A + \bar{B})(CD + E)} = \overline{(A + \bar{B})} + \overline{(CD + E)} = (\bar{A}\,B) + (\bar{C}\,\bar{D} \cdot \bar{E})$$

Then, $$F = (\bar{A}\,B) + (\bar{C}\,\bar{D} \cdot \bar{E}) = (\overline{\bar{A}B}) \cdot (\bar{C}\,\bar{D} \cdot \bar{E}) = (A + \bar{B})(CD + E)$$

11.8.2 NOR Implementation

The NOR function is the dual of NAND function. Hence, all the procedures and rules for NOR logic are the dual of corresponding procedures and rules of NAND logic.

The implementation of NOT, OR and AND function using NOR gates is shown in Fig. 11.7.

The realisation of any Boolean expression using only NOR gates can be achieved by writing the equation in product-of-sums form, For example, the expression.

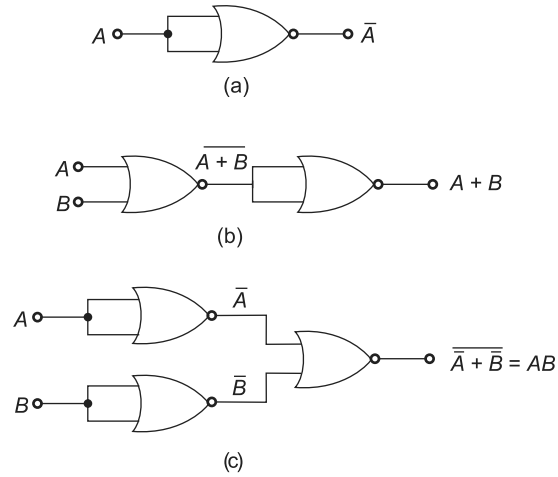


Fig. 11.7 Implementation of (a) NOT, (b) OR, and (c) AND using NOR gates

$$Y = (A + B)(C + D)\,(A + \bar{D}) \tag{11.1}$$

is in product-of-sums form. This can be realised using OR-AND gates as shown in Fig. 11.8.

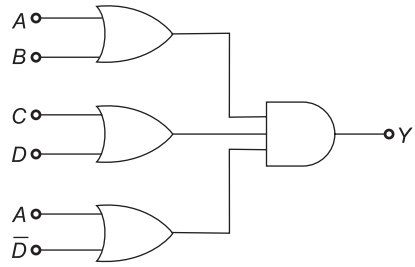


Fig. 11.8 Realisation using OR-AND gates

Using De Morgan's theorem, we can write the above equation as

$$\overline{Y} = \overline{(A+B)(C+D)(A+\overline{D})} \tag{11.2}$$

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

$$= Y_1 + Y_2 + Y_3 \tag{11.3}$$

where $Y_1 = \overline{A+B} \rightarrow$ NOR 1

$Y_2 = \overline{C+D} \rightarrow$ NOR 2

$Y_3 = \overline{A+\overline{D}} \rightarrow$ NOR 2

Therefore,

$$Y = \overline{Y_1 + Y_2 + Y_3} \rightarrow \text{NOR 3 inputs} \tag{11.4}$$

The reader can draw the diagrams.

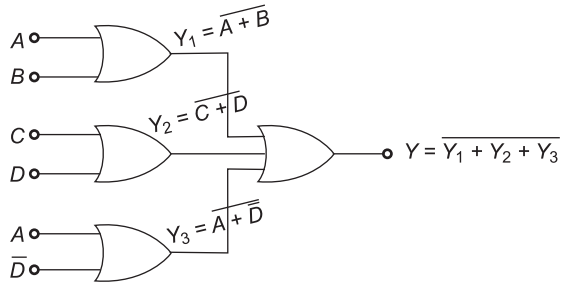
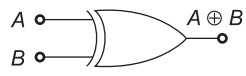


Fig. 11.9 Calculations

✦ XOR Gate

The symbol of XOR gate, its truth table are given in Fig. 1110(a) and (b). It immediately follows that,

$$A \oplus B = \overline{A}B + A\overline{B}$$



(a)

A	B	A ⊕ B
0	0	0
0	1	1
1	0	1
1	1	0

(b) Truth table

Fig. 11.10 XOR gate

✦ Half Adder (HA)

This circuit adds two binary variables, yields a carry but does not accept carry from another circuit (adder). The truth table of a half adder is given in Fig. 11.11(a) from which we can write

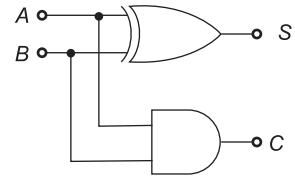
$$S = \bar{A}B + A\bar{B} = A \oplus B$$

$$C = AB$$

The half-adder logic circuit is drawn in Fig. 11.11.

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

(a) Truth table:
S = Sum; C = Carry



(b) Circuit using XOR

Fig. 11.11 Half Adder (HA)

✦ Full Adder

This circuit can add two binary numbers, accept a carry and yield a carry. Such a circuit can easily be visualised by means of two half adders (HA) and an OR as in Fig. 11.12. The reader is advised to check against the truth table.

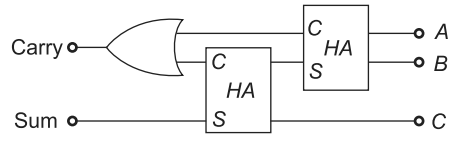


Fig. 11.12 Full adder using two half adders

❑ *Full Adder Direct Synthesis*

Table 11.5 *Full adder truth table*

S = Sum, C_i = Carry in , C_o = Carry out; Inputs are A, B, C
Outputs are S, C_o

Row No.	A	B	C_i	S	C_o
1	0	0	0	0	0
2	0	0	1	1	0
3	0	1	0	1	0
4	0	1	1	0	1
5	1	0	0	1	0
6	1	0	1	0	1
7	1	1	0	0	1
8	1	1	1	1	1

It is observed from the truth table that $C_o = 1$ for rows which have two 1's otherwise it is 0. Its Boolean equation is

$$C_o = AB + BC_i + C_iA$$

It can be implemented by three AND and one OR gates

We further observe that $S = 1$ for rows with one 1 and three 1's,
i.e. odd number of 1's. This is implemented by a three input XOR. Accordingly, the logic diagram is drawn in Fig. 11.13.

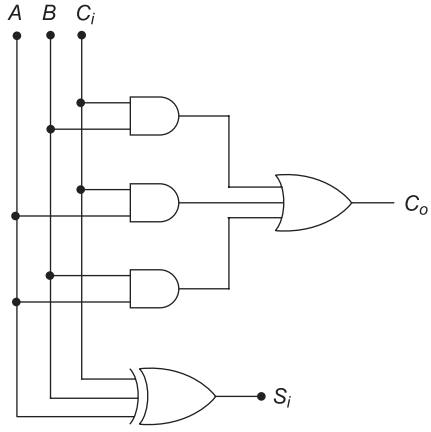


Fig. 11.13 Full adder

The reader is advised to verify.

SUMMARY

➢ The principle of boolean algebra and various combinational circuits have been explained.

EXERCISES

➥ Review Questions

1. Check the identity by truth table: $\overline{A+B} = \overline{A}\,\overline{B}$

2. Show that
$(A - B) \cdot (A + \overline{B}) = A$

3. Using 2's complement, find
01110.1001 – 0011.1110

4. Convert 11000111.1101 to octal number.

5. Consider the Boolean expression
$Y = AB + CD + E$
Implement it using NAND gates.
Hint: Take complement, apply De Morgan's theorem and take complement again.

➥ Multiple-Choice Questions

1. The binary equivalent of 24 is
(a) 11001 (b) 11000 (c) 10110 (d) 10111

2. The total variables used in the hexadecimal system is
(a) 15 (b) 14 (c) 16 (d) none of these

3. $(AB2)_H$ is represented in binary by
(a) $(101010110010)_2$ (b) $(010111010110)_2$ (c) $(101011000010)_2$ (d) $(101101101010)_2$

4. According to the De Morgan's theorem,
(a) $\overline{A \cdot B} = \overline{A} + \overline{B}$ (b) $\overline{A} \cdot \overline{B} = A + B$ (c) $\overline{A+B} = \overline{A} \cdot \overline{B}$ (d) both (a) and (c)

5. Which of the following is not true?
(a) $a.0 = 0$ (b) $a.1 = 0$ (c) $a + 1 = 9$ (d) $a + 0 = a$

6. A full adder has
(a) 2 inputs and 2 outputs
(b) 3 inputs and 2 outputs
(c) 3 inputs and 3 outputs
(d) 3 inputs and 1 output

7. The Boolean expression for the circuit of Fig. 11.14 is
(a) 0 (b) 1
(c) $xy + z$ (d) $xy + z + z^{-}$

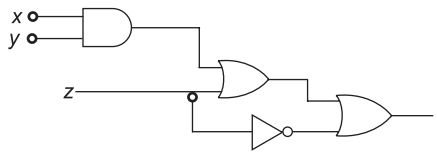


Fig. 11.14

8. According to distributive law,
(a) $A + BC = (A + B) \cdot (A + C)$ (b) $A(B + C) = AB + AC$
(c) $\overline{AB} = \overline{A} + \overline{B}$ (d) both (a) and (b)

9. Which of the following is correct?
(a) $a.0 = a$ (b) $a + 1 = 1$ (c) $a + \overline{a} = 0$ (d) $a.\overline{a} = 9$

10. A full adder is
(a) a combinational circuit
(b) a sequential circuit
(c) made up of 2 half-adders
(d) all of these

ANSWERS

✦ Multiple-Choice Questions

1. (b) **2.** (c) **3.** (a) **4.** (d) **5.** (c) **6.** (b) **7.** (b) **8.** (d) **9.** (b) **10.** (a)

CHAPTER 12

Flip-Flops

GOALS AND OBJECTIVES

- ❏ Introduction to flip-flops and their conversions
- ❏ IEEE logic symbols
- ❏ Operation of NAND gate latch/NOR gate latch and *RS* flip-flop
- ❏ Functional truth tables of preset and clear (flip-flop) and clocked *D*-flip-flop
- ❏ Explanation of edge triggering, edge triggered *RS* flip-flop and *JK* master-slave flip-flop
- ❏ Understanding of importance of memory
- ❏ Explain the various types of memory

12.1 | INTRODUCTION

A flip-flop is an electronic circuit, which has memory. If its output is 1/0, it remains same unless input changes. On change in input, its output changes (flips) to 0/1 and then remains constant. A flip-flop is a basic element of all sequential systems. A flip-flop when operated with clock pulse is called a *latch*.

12.2 | IEEE LOGIC SYMBOLS

Symbols of the logic gate that we have memorized are the traditional ones and being recognised by all workers in the electronics industry. These symbols are very useful in that they have distinctive shapes. Manufacturers' data manuals include traditional logic symbols and are recently including the newer *IEEE functional logic symbols*. These newer symbols are in accordance with ANSI/IEEE Standard 91-1984 and IEC Publication 617-12. These newer IEEE symbols are commonly referred to as "*dependency notation*." For simple gating circuits, the traditional logic symbols are probably preferred, but the IEEE standard symbols have advantages as ICs become more complicated. Most military contracts call for the use of *IEEE standard symbols*.

Figure 12.1 shows the traditional logic symbols and their IEEE counterparts. All IEEE logic symbols are rectangular. There is an identifying character or symbol inside.

12.3 | NAND GATE LATCH/NOR GATE LATCH

Two NAND gates are cross connected as shown in Fig. 12.2. Its truth table is given in Fig. 12.2(c). It is verified below.

✦ $\bar{R} = \bar{S} = 1$

Let the prior output be $Q = 1$ ($\bar{Q} = 0$). The inputs to the gate 1 are 1, 0, so its output is $Q = 1$ (no change). The inputs to the gate 2 are (1, 1), so its output in $\bar{Q} = 0$. So no change. Same happens if $Q = 0$, $\bar{Q} = 1$. The inputs to gate 1 are (1, 1), so output is $Q = 0$. The inputs to gate 2 are (1, 1), output is $Q = 0$ and $Q = 1$. So no change.

❑ ***Note*** Replace NAND gate by NOR gate for achieving NOR latch.

✦ $\bar{R} = 1, \bar{S} = 0$

Previous output $Q = 1$, $\bar{Q} = 0$. The gate 1 input is (1, 0) and output is $Q = 1$, $\bar{Q} = 0$, no changes.

Previous output $Q = 0$, $\bar{Q} = 1$, the gate 1 inputs are (0, 1) which cause its output to change to $Q = 1$ and correspondingly output of gate 1 changes to $\bar{Q} = 0$.

Thus, the stable state is $Q = 1$ ($\bar{Q} = 0$).

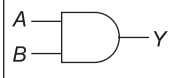
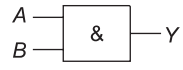
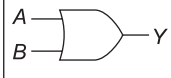
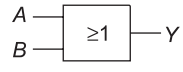
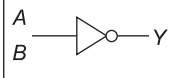
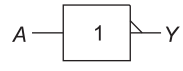
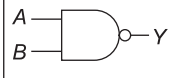
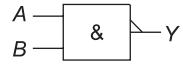
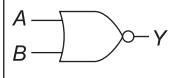
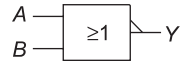
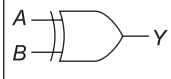
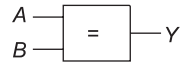
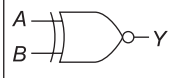
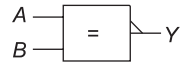
✦ $\bar{R} = 0, \bar{S} = 1$

Stable state is $Q = 0$ ($\bar{Q} = 1$), which is easily verified. If the output is $Q = 1$, it will change to $Q = 0$.

✦ $\bar{R} = 0, \bar{S} = 0$

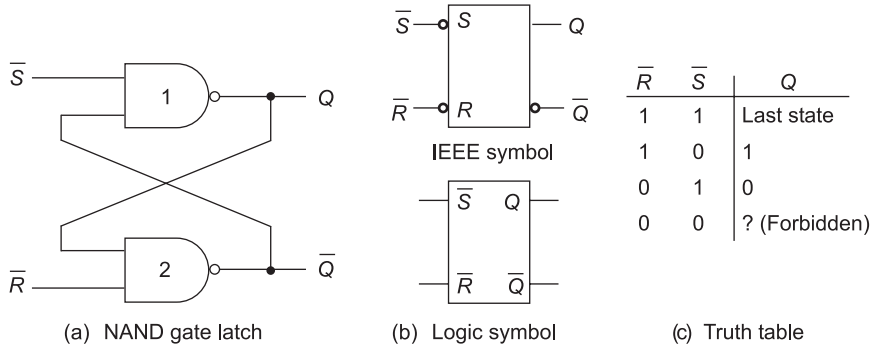
It would try to make $Q = 1$, $\bar{Q} = 1$, which is not possible. So there is a conflict.

The symbols of *SR* latch are given in Fig. 12.2(b).

Logic Function	Traditional Logic Symbol	IEEE Logic Symbol*
AND		
OR		
NOT		
NAND		
NOR		
XOR		
XNOR		

*ANSI/IEEE Standard 91-1984 and IEC Publication 617-12.

Fig. 12.1 Comparing traditional and IEEE logic gate symbols



$\overline{R}$	$\overline{S}$	Q
1	1	Last state
1	0	1
0	1	0
0	0	? (Forbidden)

(c) Truth table

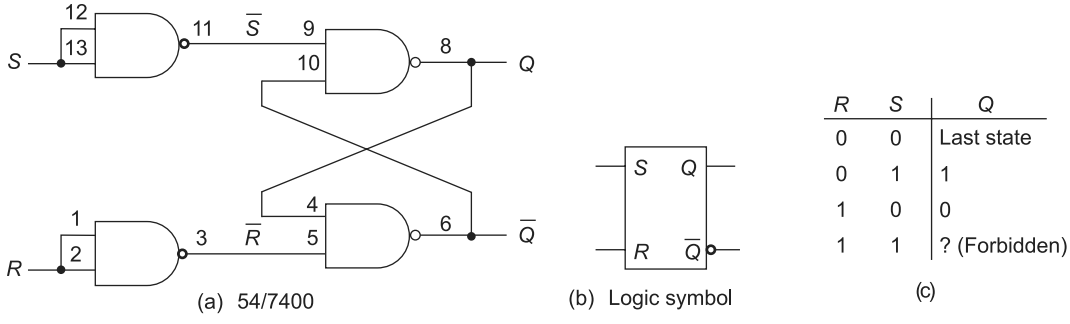
Fig. 12.2 NAND gate latch

12.4 | *RS* FLIP-FLOP

In *RS* flip-flop, we want to

- Set $S = 1 \Rightarrow Q = 1$
- Reset $R = 1 \Rightarrow Q = 0$

It means that we want to reverse the entries of the truth table of Fig. 12.2(c) and $\overline{R} \rightarrow R$, $\overline{S} = S$. This is shown in the truth table of Fig. 12.3(c). This is achieved simply by obtaining $\overline{S}$ from S and R from $\overline{R}$ by means of an inverter; we shall use NAND gate inverter. The complete circuit is drawn in Fig. 12.3(a) and its symbol in Fig. 12.3(b).



R	S	Q
0	0	Last state
0	1	1
1	0	0
1	1	? (Forbidden)

(c)

Fig. 12.3 RS flip-flop latch

12.5 | GATED FLIP-FLOPS

RS flip-flops (or latch) are said to be *transparent* as any change in R or S is immediately transmitted to the outputs Q and $\overline{Q}$. Whether R, S change will be reflected in Q, is decided with help of two NAND gates.

✦ Clocked RS Flip-flop

A clock signal is fed to both input NAND gates as shown in Fig. 12.4(a).

- When the clock is HIGH, the output of the two-input NAND gate is $\overline{S}$, $\overline{R}$. So the latch operates normally.

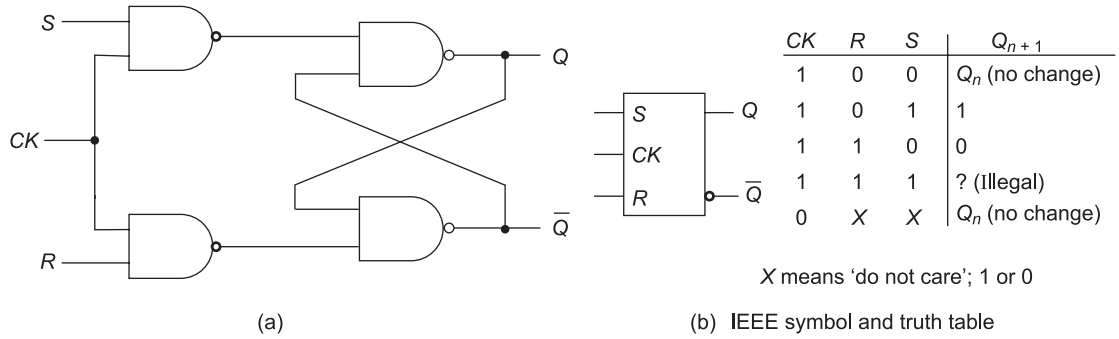


Fig. 12.4 Clocked RS flip-flop

- When the clock is LOW, the output of the two input NAND gates is (1, 1) and the latch is disabled and output does not change.

The truth table and symbol of a clocked *RS* flip-flop are shown in Fig. 12.4(b).

12.6 | PRESET AND CLEAR

When power is turned on, the flip-flop assumes a random state depending upon the time delay of the gates. It is often desirable to predefine the starting state (set or reset) of the flip-flop. This is achieved by two inputs, called *preset Pr* and *clear Cr*, as shown in Fig. 12.5. These inputs are applied in the interval between two clock pulses.

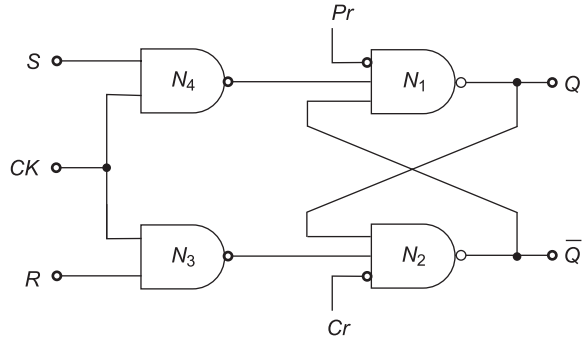


Fig. 12.5 Present and clear

Assume $CK = 0$. When $Pr = 0$ and $Cr = 1$, Q is 1, so $\overline{Q} = \overline{1.1.1} = 0$; the flip-flop is set. When $Pr = 1$ and $Cr = 0$, $\overline{Q}$ is 1 and $Q = \overline{1.1.1} = 0$; the flip-flop is in reset state or cleared. When Pr and

Cr are both 1, the circuit functions like a normal clocked *SR* flip-flop. If *Pr* = *Cr* = 0, both *Q* and $\bar{Q}$ will become 1, which is not desired. So *Pr* = *Cr* = 0 is not allowed.

Pr and *Cr* inputs are applied directly, not in synchronism with the clock pulse. Thus, they are asynchronous inputs and are also called direct set and direct reset respectively. Since the desired function is performed when the corresponding input is LOW (*Pr* = 0, *Cr* = 1 sets and *Cr* = 0, *Pr* = 1 resets clear the flip-flop); they are active low inputs (indicated by placing a bubble at these inputs). The truth table with active low preset and clear is shown below.

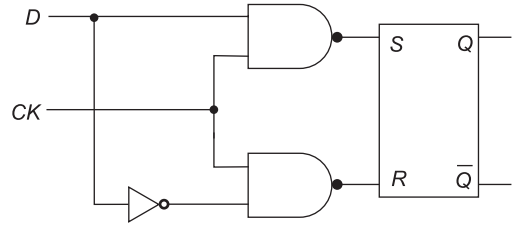
Table 12.1 *Truth table of preset and clear*

CK	*Pr*	*Cr*	*S*	*R*	Q_{n+1}
1	1	1	0	0	Q_n
			0	1	0
			1	0	1
			1	1	Not allowed
0	0	1	×	×	1 (Set
0	1	0	×	×	0 (Reset)

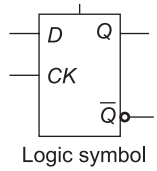
12.7 CLOCKED *D* FLIP-FLOP

In an *RS* flip-flop, 1 is transferred to *Q* if *S* = 1, *R* = 0 and a 0 is transferred to *Q* if *S* = 0, *R* = 1. It means that while transferring 1/0 to *Q*, *R* and *S* should be complements of each other. This is achieved by an inverter and two AND gates as shown in Fig. 12.6(a). When *CK* is high, *D* = 1000 are entered into *S* and its complement to *R*. It means *D* value (1/0) appears at *Q*. Thus, data *D* is transferred to *Q*. When *CK* is low, system is disabled and *Q* remains unchanged.

The truth table and logic symbol are given in Fig. 12.6(b).



(a)



Logic symbol

CK	*D*	Q_{n+1}
0	*X*	Q_n (Last state)
1	0	0
1	1	1

(b) Truth table

Fig. 12.6 A clocked *D* flip-flop

12.8 EDGE TRIGGERING

In edge triggering, short-duration pulses are generated from the clock signal of the same frequency, by passing the clock signal through certain gate arrangement. The clock signal of corresponding pulses are shown in Fig. 12.7. Positive pulses are produced when the clock signal goes from LOW to HIGH and negative pulses are produced when the clock goes from HIGH to LOW. So we have two types of edge triggering, positive edge (PT) and negative edge (NT) as shown in Fig. 12.7.

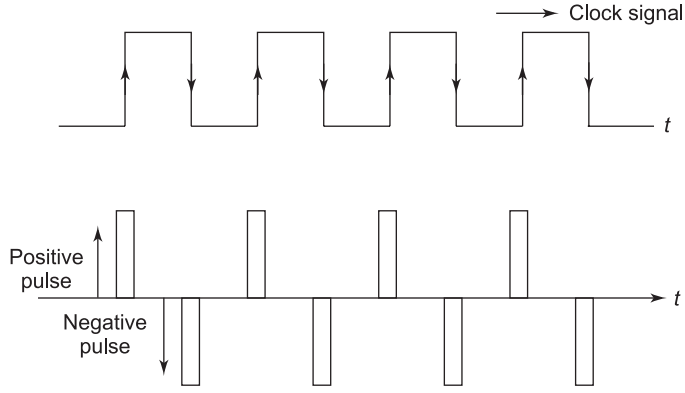


Fig. 12.7 Edge triggering

Pulse Production

✦ PT (Positive Transition)

The clock signal is applied to NAND and AND. The output of NAND which acts as inverter is the second signal fed to AND as shown in Fig. 12.8(a). It is to be noted here that the output of an AND gate is HIGH only when both inputs are HIGH.

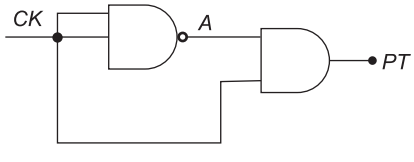


Fig. 12.8(a) Positive transition

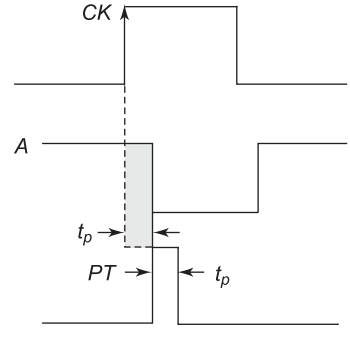


Fig. 12.8(b) Pulse after delay

Because of signal inversion and direct input to AND is HIGH and the other is LOW; so AND gate output is LOW. However, the signal delay of t_p in the inverter causes both signals fed to AND gate to be HIGH for time t_p. This is low-to-high transition; a positive pulse. The pulse appears as the AND gate after a delay of t_p as shown in Fig. 12.8(b).

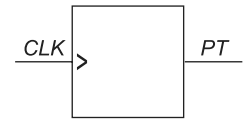


Fig. 12.8(c) Symbol for positive-edge-triggered circuit

✦ NT (Negative Transition)

The output AND gate is replaced by an OR gate as shown in Fig. 12.9(a). It is reminded here that the output OR gate is LOW when both inputs are low.

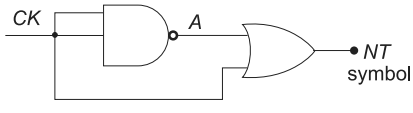


Fig. 12.9(a) Negative transition

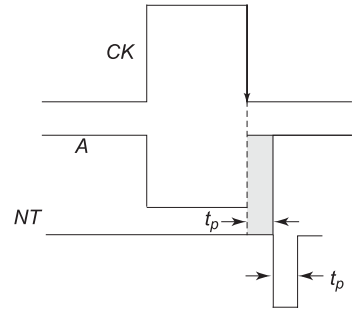


Fig. 12.9(b) Inverted signal waveform

The clock and its inverted signal A waveform are drawn in Fig. 12.9(b). Because of inverter delay t_p, when this clock, must become LOW, the signal A remains LOW for time t_p. Thus, for time t_p, the output of the OR, which is high, and then becomes low. This is the equivalent for HIGH to LOW, a negative pulse (NT).

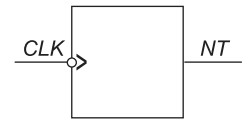


Fig. 12.9(c) Symbol for negative-edge-triggered circuit

Note:

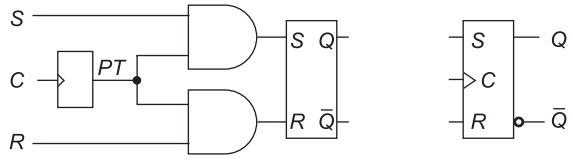
In case of edge triggering, $CK = 1$ will be replaced by arrows.

$\uparrow$ for PT and $\downarrow$ for NT

12.9 EDGE-TRIGGERED *RS* FLIP-FLOP

✦ Positive-edge-triggered

The logic diagram, symbol and truth table are given in Figs. 12.10(a), (b) and (c).



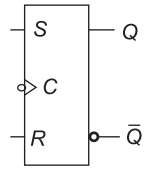
C	S	R	Q_{n+1}	Action
↑	0	0	Q_n	No change
↑	0	1	0	Reset
↑	1	0	1	Set
↑	1	1	?	Illegal

(a) Logic diagram (b) IEEE symbol (c) Truth table

Fig. 12.10 Positive-edge-triggered RS flip-flop

✦ Negative-edge-triggered

The symbol and truth table are shown in Fig. 12.11.



C	S	R	Q_{n+1}	Action
↓	0	0	Q_n	No change
↓	0	1	0	Reset
↓	1	0	1	Set
↓	1	1	?	Illegal

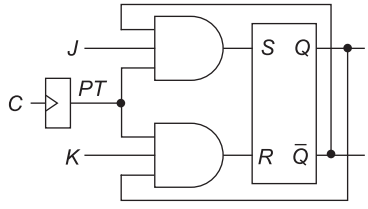
(a) IEEE symbol (b) Truth table

Fig. 12.11 Negative-edge-triggered RS flip-flop

12.10 | POSITIVE-EDGE-TRIGGERED *JK* FLIP-FLOPS

In Fig. 12.12, the pulse-forming box changes the clock into a series of positive pulses, and thus this circuit will be sensitive to PTs of the clock. The basic circuit is identical to the previous positive-edge-triggered *RS* flip-flop, with two important additions:

1. The Q output is connected back to the input of the lower AND gate.
2. The $\bar{Q}$ output is connected back to the input of the upper AND gate.



C	J	K	Q_{n+1}	Action
↑	0	0	Q_n (Last state)	No change
↑	0	1	0	Reset
↑	1	0	1	Set
↑	1	1	$\bar{Q}_n$ (toggle)	Toggle

(a) One way to implement a *JK* flip-flop (b) Truth table

Fig. 12.12 Positive-edge-triggered JK flip-flop

Table 12.2 *Development truth table*

		Previous state				
J	K	Q_n	$\bar{Q}_n$	S	R	Q_{n+1}
0	0	1	0	0	Q	Q_n
		0	1	0	0	Q_n
0	1	1	0	0	1	0 Resets
		0	1	1	0	0 no change
1	0	1	0	0	0	1 no change
		0	1	0	1	1 set } Toggle
1	1	1	0	0	1	0 resets } Toggle
		0	1	1	0	1 sets

The above table conforms the truth table of the *JK* flip-flop.

✦ Race-Around Condition

If the triggering pulse is longer in time (t_p) than the delay in toggle time (f_g), i.e. $t_g < t_p$, triggering occurs again, so the five outcomes are certain. This problem is remedial by master-slave condition of true *JK* flip-flop if such situation is lost likely in edge triggering.

12.11 | *JK* MASTER-SLAVE FLIP-FLOP

Two *JK* flip-flops are connected as shown in Fig. 12.13.

The master feeds the slave such that

$$J_s = Q_m,\ K_s = \bar{Q}_m$$

The master is triggered by the positive end of clock pulse and the slave follows with delay, triggered by the negative end of the pulse. Q_m and $\bar{Q}_m$ of the master as triggers are in accordance with the *JK* truth table. If $Q_m = 0$ and $\bar{Q}_m = 1 \Rightarrow J_s = 0$. $K_s = 1$. Upon triggering up the slave, the output is $Q = 0$ and $Q = 1$, same as the master. Similarly, $Q_m = 1$ and $\bar{Q}_m = 0 \Rightarrow J_s = 0, K_s = 1$. Upon triggering of the slave, the output is $Q = 1$, $\bar{Q} = 0$; same as the master. It means the slave simply copies the master with a delay.

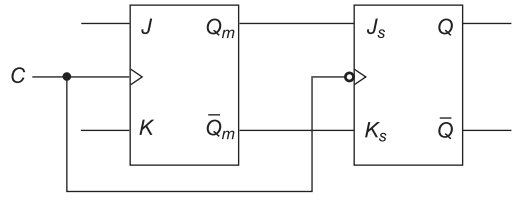
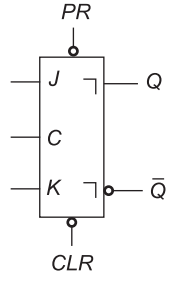


Fig. 12.13 Master-slave flip-flop

✦ 7476 JK Master Slave

The symbol and truth table are given in Fig. 12.14. It is a pulse-triggered-flip-flop. Presetting *PR* low, Q = H, and when CLR is low Q = L.

❑ ***Operation*** When the clock C is high, master Q_m, $\bar{Q}_m$ changes according to JK, which therein changes during high period of pulse (PT). The contents of the master are shifted to the slave at NT of clock pulse. This is indicated by the symbol ⊓⊔. The symbol ⌝ at output indicates postponed output changes in JK should be in low part of the pulse.



(a) Symbol

C	J	K	Q_{n+1}	Action
⊓⊔	L	L	Q_n	No change
⊓⊔	L	H	L	Reset
⊓⊔	H	L	H	Set
⊓⊔	H	H	$\bar{Q}_n$	Toggle

(b) Truth table

Fig. 12.14 Operations of JK master-slave flip-flop

12.12 | *JK* CLOCKED CONVERSIONS

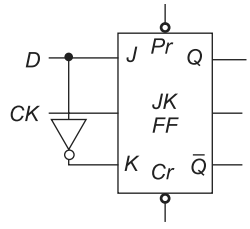
✦ D Flip-Flop [Fig. 12.15]

J and K are made complementary by connecting these through an inverter.

$D = 0 \Rightarrow J = 0, K = 1 \rightarrow Q = 0$

$D = 1 \Rightarrow J = 1, K = 0 \rightarrow Q = 1$

Hence, the truth table on clock point D appears at Q; data is transferred as shown in Fig. 12.15(b).



(a)

Truth table

D	Q_{n+1}
0	0
1	1

(b)

Fig. 12.15 JK conversions of D flip-flop

✦ T Flip-Flop [Fig 12.16]

J and K are connected

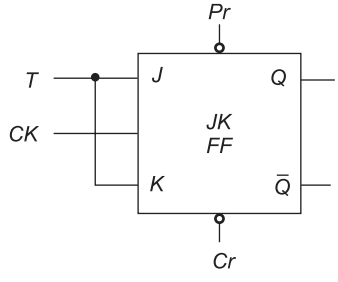
$T = 0, J = K = 0 \rightarrow Q_n$ (no change)

$T = 1, J = K = 1 \rightarrow \bar{Q}_n$ (flips)

On positive clock pulse,

$T = 0, Q_{n+1} = Q$
$T = 1, Q_{n+1} = \bar{Q}_n$ } Toggle

as shown in Truth table of Fig. 12.16.



Truth Table

T	J	K	Q_{n+1}
0	0	0	Q_n
1	1	1	$\bar{Q}_n$

Fig. 12.16 JK conversions of *T* flip-flop

✦ Summary of Excitation Tables of Flip-flops

In problems involving flip-flops, a table listing the required inputs for a certain change of state, called excitation table, is very helpful. The excitation table for various kinds of flip-flops are given in Table 12.3.

Table 12.3 *Excitation table for various kinds of flip-flops*

(a) *SR*

Q_n	Q_{n+1}	S	R
0	0	0	X
0	1	1	0
1	0	0	1
1	1	X	0

(b) *D*

Q_n	Q_{n+1}	D
0	0	0
0	1	1
1	0	0
1	1	1

(c) *JK*

Q_n	Q_{n+1}	J	K
0	0	0	X
0	1	1	X
1	0	X	1
1	1	X	0

(d) *T*

Q_n	Q_{n+1}	T
0	0	0
0	1	1
1	0	1
1	1	0

EXAMPLE 12.1 Determine the functional behavior of the circuit given assume that input W is driven by a square wave signal.

Solution

Time interval	FF_0			FF_1		
	J_0	K_0	Q_0	J_1	K_1	Q_1
Clear	1	1	0	0	1	0
t_1	1	1	1	1	1	0
t_2	0	1	0	0	1	1
t_3	1	1	0	0	1	0
t_4	1	1	1	1	1	0

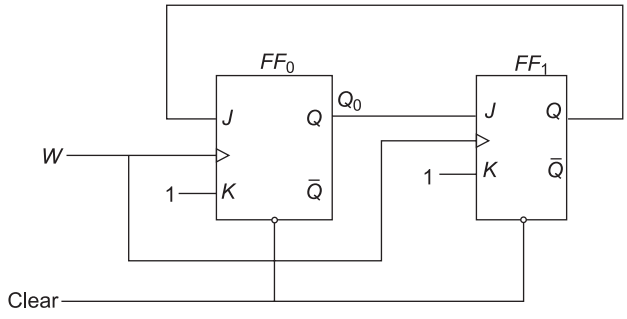


Fig. 12.17

EXAMPLE 12.2 If the clock frequency driving this flipflop is 240 Hz, what is the frequency of the flip-flops output signals?

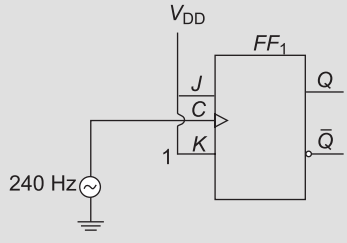


Fig. 12.18

Solution

$$F_{out} = 120 \text{ Hz}$$

EXAMPLE 12.3 The flip-flop circuit shown here is classified as synchronous because both flip-flops receive clock pulses at the same time.

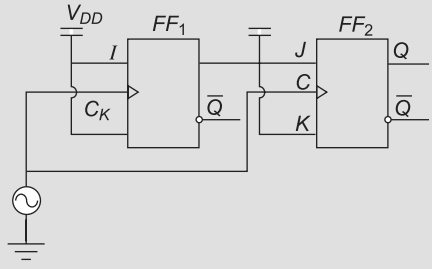


Fig. 12.19

Define the following parameters:

- Set-up time
- Hold time
- Propagation delay time
- Minimum clock pulse duration

Solution The clock frequency must be slow enough that there is adequate set-up time before the next clock pulse. The propagation delay time of FF_1 must also be larger than the hold time of FF_2.

The pulse width of the clock signal must be long enough for both flip-flops to reliable "Clock".

EXAMPLE 12.4 Although the toggle function of the JK flip-flop is one of its most popular uses, this is not the only type of flip-flop capable of performing a toggle function. Behold the surprisingly versatile D-type flip-flop configured to do the same things.

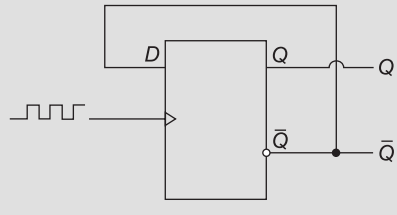


Fig. 12.20

Explain how the circuit performs the toggle function more commonly associated with JK flipflops

Solution At each clock pulse, the flip-flop must switch to the opposite state because *D* receives inverted feedback from [□Q].

EXAMPLE 12.5 Identify the flip-flop in Fig. 12.20.

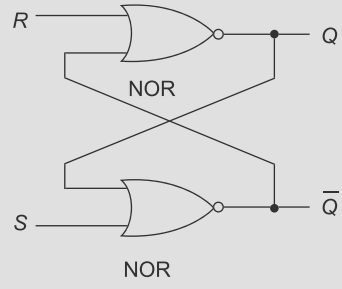


Fig. 12.21

Solution

R	S	Q	$\bar{Q}$	Q_{n+1}	
0	0	0	1	0	no change
		1	0	1	
0	1	0	1	1	set
		1	0	1	
1	0	0	1	0	Reset
		1	0	0	
1	1	0	1	–	not allowed
		1	0	–	

EXAMPLE 12.6 In Fig. 12.21, If $Q = 1$, what are R and S?

Solution If $Q = 1$, the possible combinations of RS are

R	S	Q	Q_{n+1}
0	0	1	1

Hence, $R = 0$ and $S = 0$.

EXAMPLE 12.7 In Fig.12.21. What is Q?

Solution

R	S	Q	Q_{n+1}
1	0	0	0
		1	0

Hence, for $S = 0$, $R = 1$, the output Q will be reset.

EXAMPLE 12.8 For the NOR base flip-flop of Fig. 12.22, if $R = 0$, $\bar{S} = 1$, what would be Q?

Hint: $\bar{Q} = \overline{S + Q}$,
$Q = \overline{R + Q}$

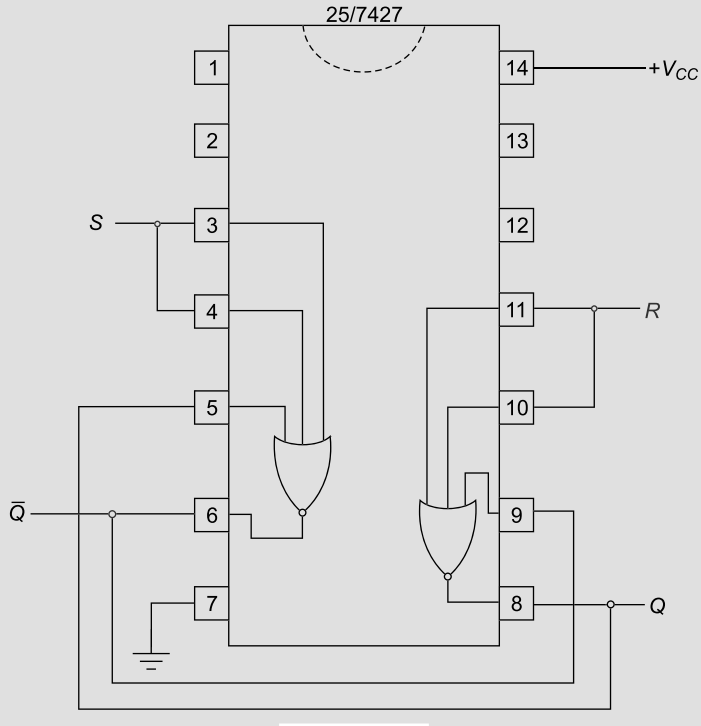


Fig. 12.22

Solution

S	R	Q	$\bar{Q}$	Next State Q_{n+1}
0	0	0	1	0
		1	0	1

From Fig. 12.22, output of NOR gate at pin number 8 is

$$\overline{Q} = \overline{R + R + \overline{Q}} = \overline{R + \overline{Q}}$$

Given, $R = 0$, let $Q = 0$, then $\overline{Q} = 1$.

Hence, the given equation is satisfied.

Output of NOR gate at pin number 6 is Q which is given as

$$\overline{Q} = \overline{S + S + Q} = \overline{S + Q}$$

Given, $\overline{S} = 1$, let $Q = 0$, then $\overline{Q} = 1$.

Hence, when $R = 0$, $S = 1$, output $Q = 0$.

EXAMPLE 12.9 For the Fig. 12.22, if $\overline{Q} = 1$, what would be $\overline{R}$, $\overline{S}$?

Solution

Given $\overline{Q} = \overline{S + S + Q} = \overline{S + Q}$

Given, $\overline{Q} = 1$, hence $Q = 0$

Therefore, from the equation – $\overline{Q} = \overline{S + Q}$

$1 = \overline{S + 0}$

$\therefore$ $\overline{S} = 1$ is the solution to satisfy the above equation.

As $Q = \overline{R + \overline{Q}}$

Given, $\overline{Q} = 1$, $\therefore Q = 0$

$O = \overline{R + 1}$

If $R = 1$, only then the above equation is satisfied.

So $\overline{R} = 0$

$\therefore$ For $\overline{Q} = 1$, $\overline{R} = 0$, $S = 1$.

EXAMPLE 12.10 Show that the *D* flip-flop excitation table connected as in Fig. 12.23 acts as an *RS* flip-flop.

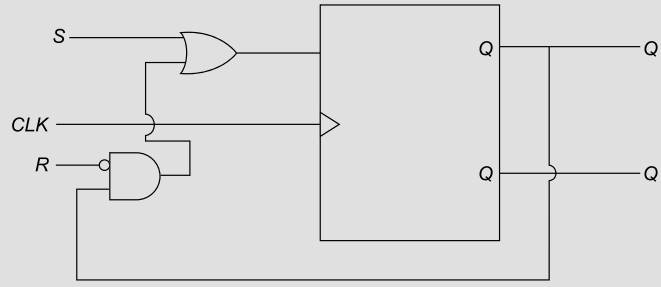


Fig. 12.23

Solution The truth table of the *RS* flip-flop is

R	S	Q	$\overline{Q}$	$\overline{Q}_{n+1}$	
0	0	0	1	0	no change
		1	0	1	
0	1	0	1	0	Reset
		1	0	0	
1	0	0	1	1	set
		1	0	1	
1	1	0	1	–	illegal – not valid
		1	0	–	

Hence, verified that *D* flip-flop for the above connection acts as *RS*.

12.13 MEMORY

Memory is the storage space in any electronic machine or computer where data and instructions are stored. Memory usually refers to a form of semiconductor storage to save information known as random access memory (RAM).

There are three main types of memory, given as follows:

1. Primary Memory/Main Memory
2. Cache Memory
3. Secondary Memory

We will explain the different types one by one in the following section.

12.13.1 Primary Memory (Main Memory)

Primary memory saves computer's presently working data and instructions. It has a smaller space and the data vanishes away when the computer is turned off. In the main memory, the data and the instructions which need to be processed are stored. Primary memory is of two types: RAM and ROM. Main memory is usually volatile and faster than the secondary memory.

(a) RAM or Random Access Memory

RAM is the memory where currently used data, all currently running programs, and operating system resides temporarily and data can be accessed by the computer's processor. RAM is volatile as the contents are available only till the power is on. The contents of memory are no more accessible once the power is turned off. It is expensive to handle, and hence UPS (uninterrupted power supply) is used with computers. Size of RAM is small and it can handle only a small amount of data. There are two types of RAM: Static RAM (SRAM) and Dynamic RAM (DRAM).

❑ ***Static RAM (SRAM)*** Static Ram is the memory that retains its contents as long as the computer is on or power supply is available. If the power gets down then the contents are lost. RAM uses more chips than DRAM and hence has extra space available. SRAM is volatile in nature, has high manufacturing cost, long life and faster (can be used as cache), but at the same time it is expensive and consume more power.

❑ ***Dynamic RAM (DRAM)*** DRAM is very economic so it uses most of the system's memory. DRAMs are made up of one capacitor and one transistor (memory cell). The transistor and capacitor used here are of very small size; hence millions can be placed on one single chip. DRAM has short life span, consumes less power, and is slower than SRAM. To maintain the data, DRAM is to be continuously refreshed.

(b) ROM or Read Only Memory

ROM is the memory that can be only read and the contents remain as it is when the power is turned off. This feature makes it nonvolatile in nature. ROM generally contains manufacturer's information and instructions. It stores bootstrap loader (the initial program). As the power is turned on, bootstrap loader starts computer operation. The information is infused permanently in memory during the manufacture only. ROM chips are used in various electronic items like microwave, dishwasher, washing machine, etc. Typically ROM are not read only device but updates are also possible. If ROM is to be rewritten then large portions of memory have to be erased.

There are different types of ROM as follows:

- ***MROM (Masked ROM)*** MROM contains a pre-programmed set of instructions or data. It is very expensive to handle.
- ***PROM (Programmable Read only Memory)*** PROM is read-only memory which can be altered only once. PROM is bought blank and desired content can be written on it using PROM program. The content entered once cannot be erased.
- ***EPROM (Erasable and Programmable Read Only Memory)*** As the name suggests, this type of ROM can be removed by exposing it to UV light for some time (40 minutes). While programming, an electrical charge is trapped in the gate region. The charge can stay there for more than a decade as there is no path for leakage. UV light is bombarded through an LED (quartz crystal window) to erase the charge.
- ***EEPROM (Electrically Erasable and Programmable Read Only Memory)*** EEPROM can be erased and programmed electrically and can be reprogrammed again for thousand times. It is very fast and takes a fraction of millisecond to erase the content and reprogram. EEPROM is location independent and has flexible programming unlike EPROM, as to erase one byte of memory there is no need to blank the entire chip.

12.13.2 Cache Memory

Cache memory is used to increase the performance of a computer. In cache, most frequently used functions and programs are saved. CPU accesses the instructions and programs which resides in the cache. Cache memory is faster than the main memory and also has less access time, but it lacks in terms of capacity and cost (very expensive).

12.13.3 Secondary Memory (Backup Memory)

Also known as external memory, secondary memory is non-volatile in nature. Thus, it can be used to store data permanently. It is slower than the main memory. Secondary memories are accessed by input and output devices. Data is stored in secondary memory and then shifted to primary memory, and from there, CPU accesses that data. Examples of secondary memory are DVD, hard disk, CD-ROM, etc.

12.14 DIFFERENCE BETWEEN RAM AND ROM

Both RAM and ROM are the parts of primary or internal memory of a computer system but they differ each other in the role played by them in the working of a computer system. The following Table 12.14 presents some points of differences between RAM and ROM.

Table 12.4 *Difference between RAM and ROM*

S. No.	RAM	ROM
1	Information can be read as well as written on it. It randomly selects and uses location of the memory to access the information.	Information can only be read and cannot be written.
2.	It stores information temporarily.	It stores information permanently.
3.	It is a volatile memory, i.e. it's contents get lost once the power supply is turned off.	It is a nonvolatile memory, i.e. its content does not get lost even if the power supply is turned off.
4.	It is used to hold the data and instructions currently being processed.	It holds startup instructions like BIOS.
5.	It holds application programs which are currently being used.	It holds system software like operating system.
6.	Types of RAM are DRAM and SRAM.	Types of ROM are PROM and EPROM.

SUMMARY

➢ In this chapter flip-flops are introduced and their conversions are described. IEEE logic symbols are explained NAND gate and RS flip flop are discussed. Functional truth tables of preset and clear and clocked D flip-flop explained. Edge triggering, edge trigged RS flip-flop JK master-slave flip-flop are also explained.

EXERCISES

➥ Review Questions

1. What is a latch?
2. Why are universal gates preferred for making flip-flops?
3. Clear the comparative changes in various flip-flops.
4. How does a sequential circuit differ from a combinational circuit?
5. Explain the term *edge triggering*.
6. Draw and explain with excitation table the *RS*, *JK*, *D* and *T* flip-flop.
7. RAM erases itself every time you shut down the computer. Explain.
8. Why does getting more RAM make the system run faster?
9. Explain the functioning of RAM.
10. Explain the differences between RAM and ROM.
11. Explain the differences between RAM and Cache memory.

➥ Problems

1. For pin diagram of Fig. 12.20, is it a flip-flop? If so what kind?
   Hint: Two inputs of the two NORs make three inputs; NORs as two input NORs.
2. Draw the symbols of *RS* flip-flop for various types of triggering.
3. For the NOR gate flip-flop of Fig. 12.15, label the logic levels (1/0) on LED pin for
   (a) $R = S = 0$ (b) $R = S = 1$ (c) $R = 0, S = 1$ (d) $R = 1, S = 0$

Pin No.1	(a)	(b)	(c)	(d)
1	0	1	0	1
2	0	1	0	1
3	1	0	1	0
5	1	0	1	0
12	0	1	1	0
13	0	1	1	0
11	1	0	1	0
9	0	0	1	0
4	1/0	(last state) conflict	1	0
8	1/0		1	0
6	0/1		0	1
10	0/1		0	1

4. Establish the equivalence of the flip-flop of Fig. 12.24. What happens when the clock is low?

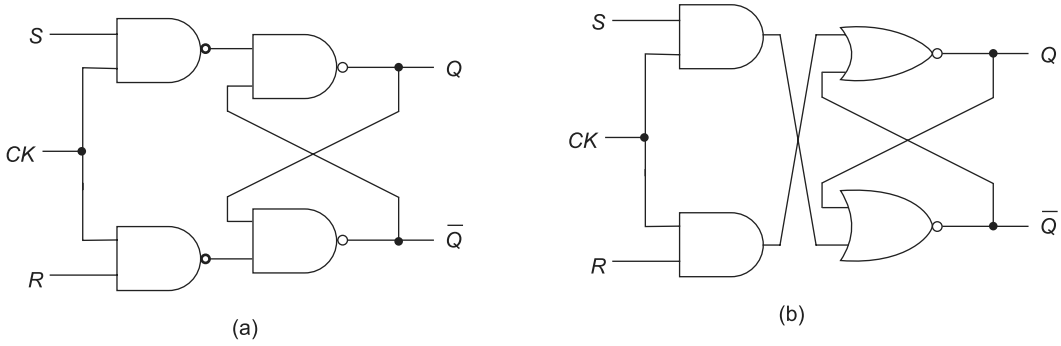


Fig. 12.24

5. Check the given excitation table of various flip-flops.

$Q_n \rightarrow Q_{n-1}$		*S*	*R*	*J*	*K*	*D*	*T*
0	*0*	0	×	0	×	0	0
0	1	1	0	1	×	1	1
1	0	0	1	×	1	0	1
1	1	×	0	×	0	1	0

Hint: Look at the truth table of the flip-flops in the reverse way.

6. Show that the *D* flip-flop excitation table connected as in Fig. 12.25 acts as an *RS* flip-flop.
Hint: A bubble at the input of the AND gate is the symbol of the inverter.

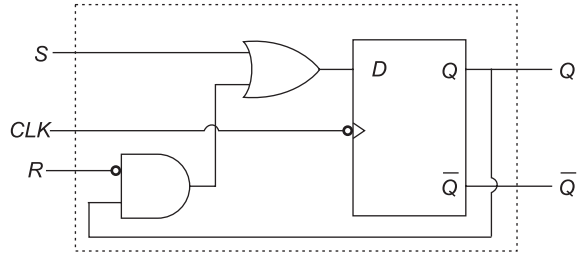


Fig. 12.25

7. Figure 12.26 shows the input waveform *R*, *S*, and *EN* applied to a clocked *RS* flip-flop. Below these waveforms, draw the timed output *Q* waveform.

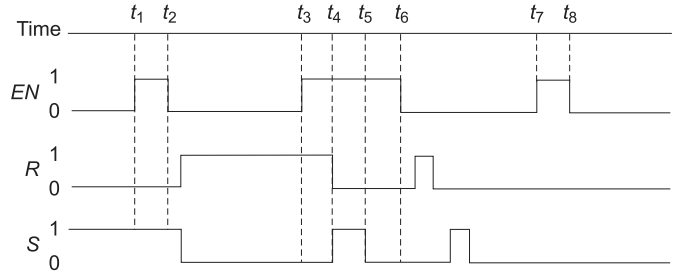


Fig. 12.26

8. Draw the output waveform for the following waveform input to the *SR* flip-flop (Fig. 12.27).

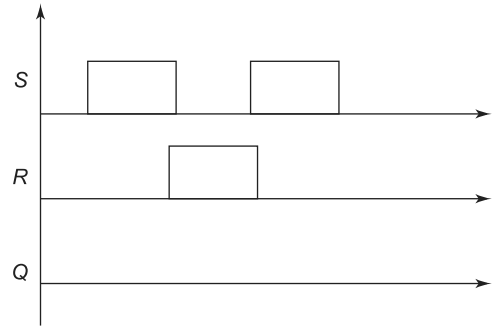


Fig. 12.27

9. Implement the 4:1 multiplexer using basic gates.

10. In the sequential circuit of Fig. 12.16, the NOR gates are replaced by NAND gates. Check that its operation agrees with the *SR* flip-flop truth table (Table 12.2).
Hint: NAND gate the output is 0 only if both inputs are 1.

11. In an *SR* flip-flop (Fig. 12.20), *S* is connected to *R* through a NOT gate input (1/0) given to *S*. Using the truth table 12.2, show that it acts as D flip-flop. Identify the output terminal.

➥ Multiple-Choice Questions

1. A flip-flop is
(a) a combinational circuit (b) a sequential circuit
(c) an analog to digital converter (d) a digital to analog converter

2. The race-around condition occurs in
(a) *SR* flip-flop (b) *D* flip-flop (c) *T* flip-flop (d) *JK* flip-flop

3. Race-around condition is avoided by using
(a) master-slave flip-flop (b) set reset flip-flop
(c) level-triggered flip-flop (d) *D* flip-flop

4. The *D* flip-flop is mainly used in
(a) counters (b) serial abler (c) shift registers (d) all of these

5. Which property of the flip-flop is used in counter designing?
(a) Shifting (b) Toggling (c) Set-Reset property (d) None of these

6. Which of the following table is used for determining input if given present state and next state?
(a) Characteristics table (b) Truth table
(c) Excitation table (d) None of these

7. The race-around condition in *JK* flip-flop occurs when
(a) $J = 0, K = 1$ (b) $J = 1, K = 0$ (c) $J = 1, K = 1$ (d) $J = 0, K = 0$

8. In a master-slave *JK* flip-flop,
(a) the slave follows the output of the master (b) the master follows the output of the slave
(c) both work independently (d) none of these

9. The reason for race-around condition to occur is
(a) the clock period is less than the propagation delay of the gates
(b) the clock period is greater than the propagation delay of the gates
(c) because of the improper feedback connection
(d) because of the circuit delay involved

10. The *SR* flip-flop processes illegal output only when
(a) $S = 0, R = 0$ (b) $S = 0, R = 1$ (c) $S = 1, R = 0$ (d) $S = 1, R = 1$

11. Which of the following statements is true about EEPROM?
(a) It can be modified with memory write operation.
(b) It cannot be modified.
(c) It is lost when the power is turned off.
(d) It can be erased by using high voltage.

12. RAM and ROM differ as _____.
(a) RAM has a read/write operation but ROM doesn't
(b) RAM is volatile, whereas ROM is volatile
(c) RAM holds currently used instructions, whereas ROM holds system software
(d) All of these

13. DRAM is better than SRAM as it is _____.
(a) cost efficient (b) fast
(c) more storage capacity (d) Both (a) and (c)

ANSWERS

✦ Problems

1. (a) $(127)_{10}$ (b) $25375)_{10}$
2. (a) 111100.00001 (b) $(10.6875)_{10}$
3. 00011000
5. (a) $15 \times 16^1 + 12 \times (16)^\circ = 245$, (b) $9 \times 16^1 + 10 \times 16^\circ = 156$
9. (a) $(427)_8$ (b) $(44.1)_8$ (c) (307.64)
10. (a) $(117)_{16}$ (b) $24.2)_{16}$ (c) $(C7.D)_{16}$
11. (a) 1010.101 (b) 00011000 (c) 100101011 (d) $-(38)_{16}$
12. (a) 00101011 (b) 010.10, 0011 (c) 53 (d) 2B
13. (a) 0000000100100111 (b) 01001001.0101 (c) $(1.447)_8$ (d) $(127)_{16}$
17. $(307.64)_8$ **21.** $\overline{AB}$
33. *SR* flip-flop

✦ Multiple-Choice Questions

1. (b) **2.** (d) **3.** (a) **4.** (c) **5.** (b) **6.** (c) **7.** (c) **8.** (a) **9.** (b) **10.** (d)
11. (d) **12.** (d) **13.** (d)

CHAPTER 13

Registers and Counters

GOALS AND OBJECTIVES

- Introduction of shift registers—serial in serial out (SISO), serial in parallel out (SIPO), parallel in serial out (PISO), parallel in parallel out (PIPO)
- Explanation of counters—asynchronous counter and synchronous counters
- Introduction to microprocessors and microcontrollers

13.1 INTRODUCTION

A register can store data, which is entered into it. The data can be taken out when needed.

Counters count the input pulses, which have known frequencies. So these can measure time.

13.2 SHIFT REGISTERS

Data are entered and stored in a shift register. It can then be taken out when needed. A 4-bit shift register is shown in Fig. 13.1. It comprises 4 *JK* flip-flops in which data are entered.

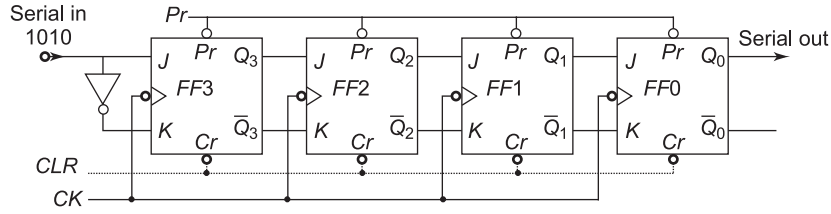


Fig. 13.1 4-bit shift register

FF3 J, K terminals are connected through an inverter so that it acts as D flip-flop through which the data are entered; the flip-flops are cleared by placing 0 at CLR. During a normal operation, the CLR is held at 1. Also, *Pr* is held at 0 so that these do not interfere with normal operation of the registers.

✦ Operation

All the flip-flops are triggered by common clock pulses (CK), data (1010) is now fed in Least Significant Digit (LSD) first and Most Significant Digit (MSD) last as shown in Table 13.1.

Table 13.1 *Operation of shift registers*

Clock pulse	Serial in	Q_3	Q_2	Q_1	Q_0 (serial out)	
0	0	0	0	0	0	
1	1	1	0	0	0	
2	0	0	1	0	0	
3	1	1	0	1	0	Data entered
4	0	0	1	0	1	
5	0	0	0	1	0	
6	0	0	0	0	1	
7	0	0	0	0	0	1 Register cleared

13.3 REGISTER TYPE

- Serial In Serial Out (SISO)
- Serial In Parallel Out (SIPO)
- Parallel In Serial Out (PISO)
- Parallel In Parallel Out (PIPO)

A 4-bit shift register which can operate as all-fuse type is shown in Fig. 13.2. The SISO operation is same as illustrated in Fig. 13.1. All we need to explain is the parallel operation.

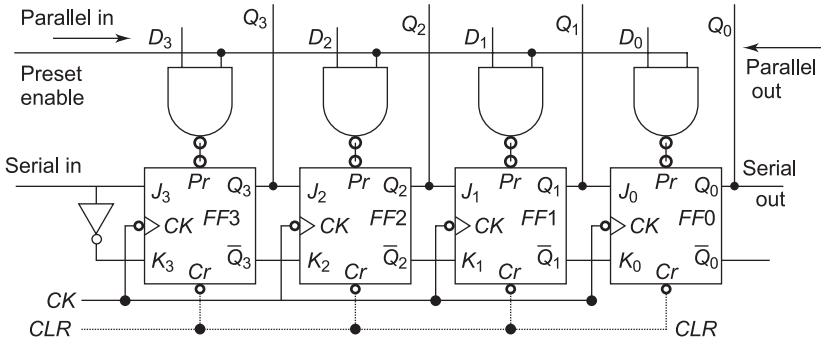


Fig. 13.2 4-bit shift register

✦ Parallel Out

Connections are directly taken from Q_3 to Q_0 outputs of flip-flops.

✦ Parallel In

After clearing the register, 'parallel enable' input is made. This results in NAND gates to act as inverters.

Any D_i input causes

$D_i = 0$, $P_r = 1$ and $Q_i = 0$

$D_i = 1$, $P_r = 0$ and $Q_i = 1$

Data is thus entered in parallel (simultaneously) at Q_is. It can then be read at $Q_3 - Q_0$ or it can be serially output by applying four pulses.

✦ Shifting of Data to Left

In the register of Figs 13.1 and 13.2, the data is shifted from left by one bit at clock pulse by feeding output of FF0 to input of FF1, and FF1 to FF2, FF2 to FF3. On the occurrence of a pulse, the data word will shift one bit to left; this logic is shown in Fig. 13.3.

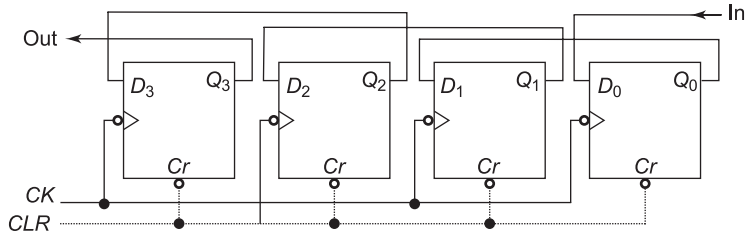


Fig. 13.3 Register shift left

✦ Bidirectional Register

By combining Figs 13.1 and 13.3 through AND-OR-logic, we can build a bidirectional shift register. This logic is shown in Fig. 13.4 in which one AND gets 1 input, while the other AND gets 0, and the connection in one is Q, while the other will be $\bar{Q}$.

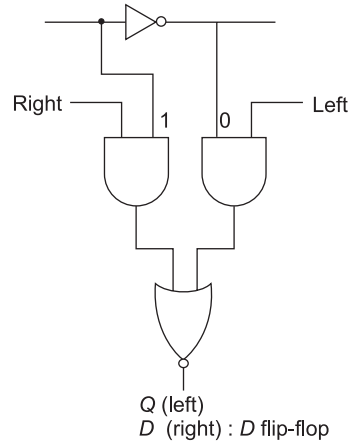
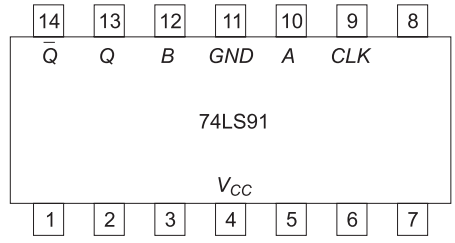


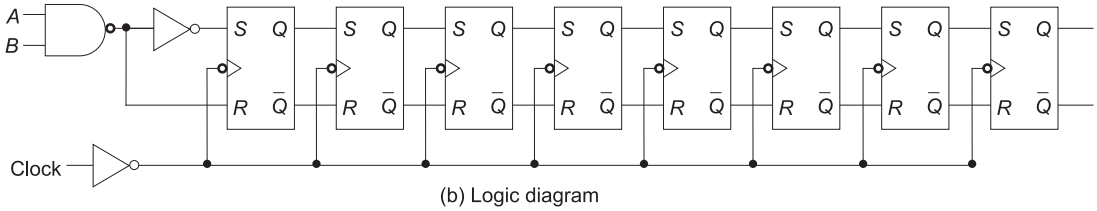
Fig. 13.4 Bidirectional register

13.4 | COMMERCIALLY AVAILABLE TTL MSI/LSI REGISTERS

Serial In Serial Out—54/74 LS91 8 bits
Serial In Parallel Out—54/74164 8 bits
Parallel In Serial Out— 54/74165 8 bits
Parallel In Parallel Out—54/74198 8 bits
An example of 74LS91 shift register is shown below in Fig. 13.5.



(a) DIP pinout



(b) Logic diagram

Fig. 13.5 74LS91 8-bit shift register

The NAND gate with inputs A and B simply provides a gating function for the input data stream if desired. If B is set at 1, the NAND acts as inverter for input at A.

The two inverters feed the input at S of first flipflop. A_1 as data sets $S = 1$ and $R = 0$. A_0 as data resets $S = 0$ and $R = 1$.

If gating is not desired, simply connect the pins 10 and 12 together and apply the input data stream to this connection. If $A = B$, NAND gate acts as inverter.

13.5 | COUNTERS

Binary counters are one of the most important and widely used of all digital circuits. A counter is a circuit that counts the number of occurrences of an input (in terms of positive or negative edge transitions in the case of a binary input). Each count, a binary number, is called a state of the counter. Hence, a counter counting in terms of n bits has 2^n different states. The number of different states of a counter is also known as the *modulus* of the counter. Thus, an n-bit counter is a modulo 2^n counter.

Counter circuits are primarily constituted of flip-flops, which along with combinational elements are used for the generation of control signals. Depending on the manner in which the flip-flops are triggered, counters can be divided into two major categories:

1. Asynchronous counter (ripple counter)
2. Synchronous counter

In case of an *asynchronous counter*, not all the flip-flops are clocked simultaneously, whereas in a *synchronous counter*, all the flip-flops are clocked simultaneously.

If the counter counts in such a way that the decimal equivalent of the output increases with successive clock pulses, it is called an UP counter. If it decreases, it is called a DOWN counter. An UP/DOWN counter can also be designed which can count in any direction depending upon the control input.

A decoding circuit can be connected to the output of a counter in such a way that the output of the circuit goes high (or low) only when the counter contents are equal to a given state.

13.6 | ASYNCHRONOUS COUNTERS

Figure 13.6 shows a 3-bit (modulo 8) asynchronous counter. As J and K are tied together to 1 input, the JK, flip-flop act as T-flip-flop. The Q-output of one flip-flop feeds the CK trigger of the next flip-flop.

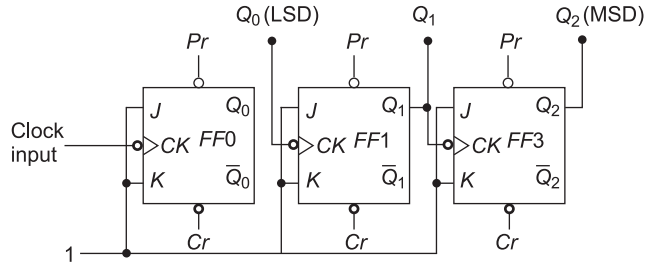


Fig. 13.6 3-bit asynchronous counter (UP)

✦ Operation

All the flip-flops are initially cleared. The clock pulses are then applied with

each clock pulse NT (1 to 0), Q toggles. For each NT (1 to 0) of Q_0, Q_1 toggles; similarly 1 to 0 transition of Q_2 toggles. The transition and count are shown in Table 13.2. Thus, the counter counts from 0 to 7(111), these all the three transit from 1 to 0 and the counter returns to 0. As the pulses get applied to flip-flops in sequence, it is also called a *ripple counter*.

Table 13.2 *Operations of asynchronous counter*

Input (Clock)	Q_2	Q_1	Q_0
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1
8	0	0	0

✦ Observation

The clock transits (NT) 8 times, Q_0 4 times, Q_1 2 times, and Q_3 1 time. The wave frequency reduces by a factor of 2 at each stage (frequency division).

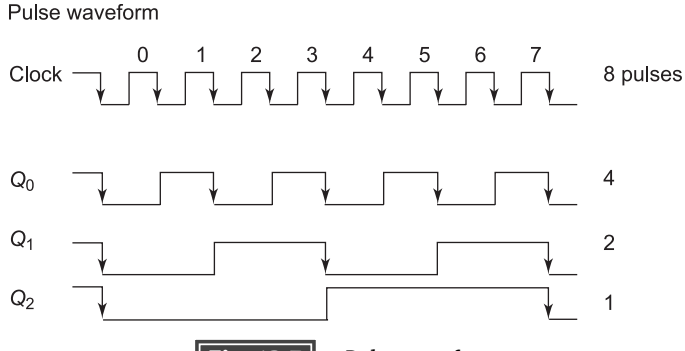


Fig. 13.7 Pulse waveform

❑ ***Down Counter*** $\bar{Q}$s are connected to the next flip-flop. As $\bar{Q}$ toggles from 1 to 0, the corresponding Q toggles 1 to 0, which is count down.

❑ ***Up-Down Counter*** CK is connected to Q for up-down and $\bar{Q}$ through AND-OR logic.

❑ ***Preset*** The counter through *Pr* input is set to desired value before the clock pulses are applied.

13.7 | SYNCHRONOUS COUNTER

The clock pulses are applied synchronously (at the same time) to all the flip-flops. This avoids the ripple delay of asynchronous counter and logic complexity increases.

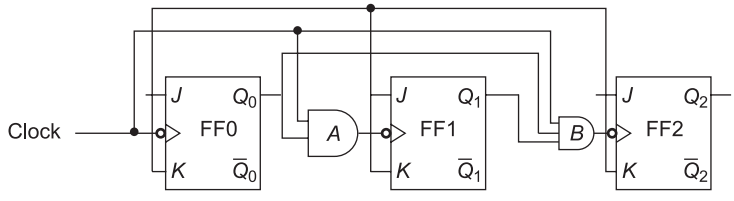


Fig. 13.8 Synchronous counter (up)

✦ Connected Diagram and Operation

Clock is applied to all the three flip-flops simultaneously. However, clock NT for FF_1 and FF_2 must pass through AND gates *A* and *B* respectively. AND gate A passes NT at NT of clock, and of Q_0 (1 to 0), then FF_1 toggles.

- AND gate *B* passes NT at NT of clock, and of Q_0 and Q_1 (1 to 0), FF_2 then toggles.
- FF0 toggles at every NT of clock, at clock frequency.

Values of Q_2, Q_1, Q_0 against count are shown in Table 13.3, with 1 to 0 transitions also indicated.

Table 13.3 *Operation of synchronous counter*

Count	Q_2	Q_1	Q_0
0	0	0	0
1	0	0	1 ↓
2	0	1	0 ↓
3	0	1 ↓	1
4	1	0 ↓	0 ↓
5	1	0	1 ↓
6	1	1	0 ↓
7	1 ↓	1 ↓	1
8	0 ↓	0 ↓	0 ↓

✦ Count Down

The A gate is for CK and $\bar{Q}_1$ and the B gate is fed CK, $\bar{Q}_2$. The counter is preset to $Q_3 = Q_2$. $Q_1 = 1$. The transition table of $\bar{Q}_3$, $\bar{Q}_2$, $\bar{Q}_1$ and the countdown is given in Table 13.4.

Table 13.4 *Countdown*

Count	Q_3	Q_2	Q_1
7	1	1	1
6	1	1	0
5	1	0	1
4	1	0	0
3	0	1	1
2	0	1	0
1	0	0	1
0	0	0	0

EXAMPLE 13.1 How many flip-flops are needed in a shift register to store

(a) 6-bit binary numbers?

(b) Decimal numbers up to 24?

(c) Hexadecimal numbers up to E?

Solution

(a) 6

(b) $24 \rightarrow 11000$, five

(c) $E \rightarrow 14 \rightarrow 1110 \rightarrow$ four

EXAMPLE 13.2 A shift register has eight flip-flops. What is the largest binary, decimal and hexadecimal numbers that can be stored in it?

Solution

(a) 11111111

(b) 255

(c) FF

EXAMPLE 13.3 Name the four types of shift registers. Draw their block diagrams.

Solution

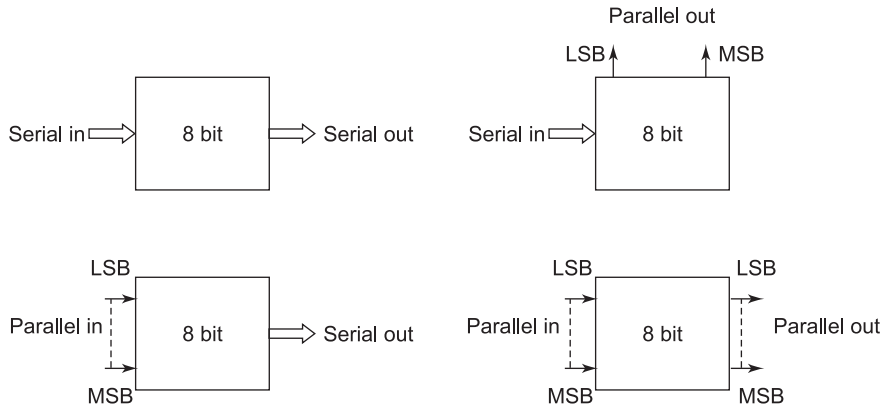


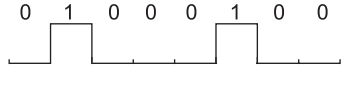
Fig. 13.9

EXAMPLE 13.4 How is the decimal number 68 is shifted into the register 74LS91 of Fig. 13.4? Prepare a table giving the state of each flip-flop on the clock pulses applied. Draw the waveform when the number has been shifted in the register. How many pulses are needed to clear the register?

Solution 68 in binary is 01000100.

Number to be entered	Pulse no.	State register 8 bits	
		MSD	LSD
	0		
01000100	1	00000000	
0100010	2	00000000	
010001	3	00000000	
01000	4	10000000	
0100	5	01000000	
010	6	00100000	
01	7	00010000	
0	8	10001000	
		01000100	

Waveform after 8th pulse



No. of pulses to clear the registers = 7

EXAMPLE 13.5 How long will it take to shift 8-bit binary data into 54/74164 of Fig. 13.10 if the clock is (a) 1 MHz (b) 5 MHz?

Solution

(a) 1 MHz = 10^6 cycles/second, Time for 8 cycles = $8 \times 10^{-6} = 8\ \mu s$

(b) 5 MHz, Time for 8 cycles = $8 \times (10^{-6}/5) = 1.6\ \mu s$

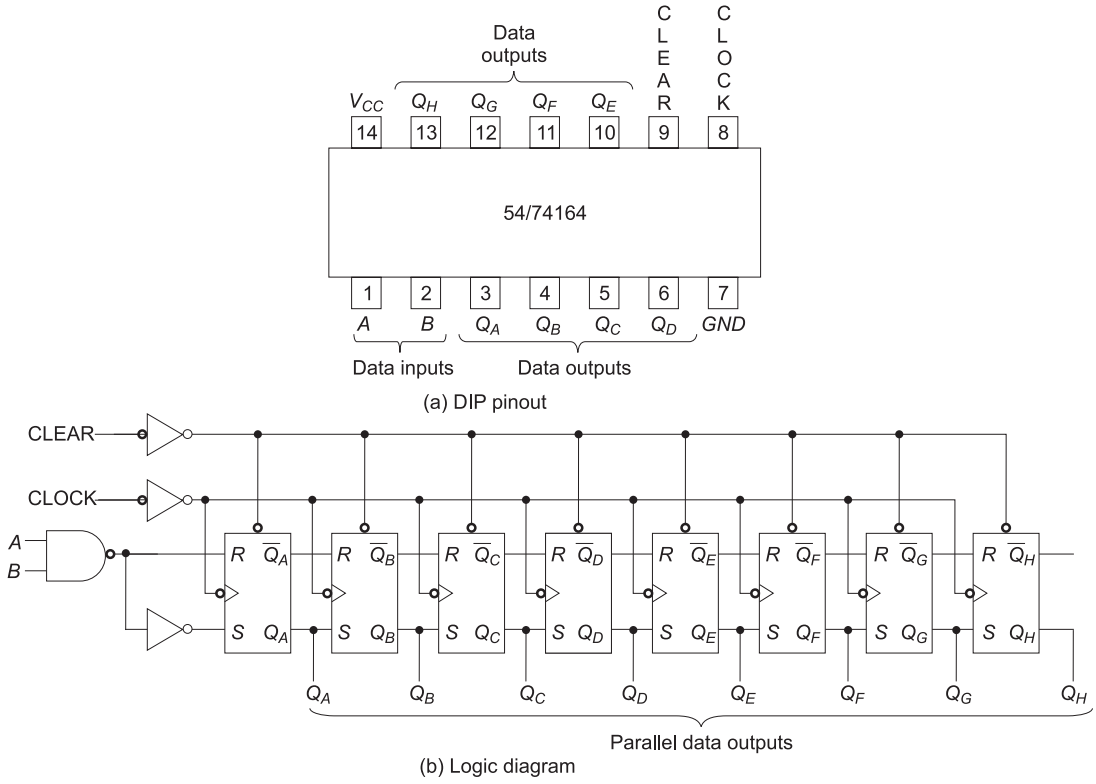


Fig. 13.10 54/74164 8-bit shift register

The register is enabled when $B = 1$, data at A will go to S of the first flip-flop. If $B = 0$, the register is disabled as $S = 0$ for data 1 or 0.

EXAMPLE 13.6 How many flip-flops are needed for (a) Mod-128 counter, (b) Mod-64 counter, and (c) Mod-32 counter? What is the largest decimal number and largest hexadecimal number that can be stored in the Mod-64 counter?

Solution

(a) $128 = 2^7$, 7 flip-flops

(b) $64 = 2^6$, 6 flip-flops
(c) $32 = 2^5$, 5 flip-flops
Mod-64 is 6 flip-flops
Largest binary number that can be stored is 111111 → decimal 63 → Hexadecimal 63
Hexadecimal number AE

EXAMPLE 13.7 What is the Mod of the counter to store Hexadecimal number *AE*.

Solution

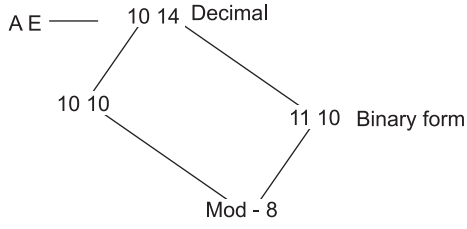


Fig. 13.11

EXAMPLE 13.8 A 6-bit ripple down counter has the following states:

(a) 101010 (b) 001100 (c) 110000

Determine the number of flip-flops that would be complemented after the application of one clock pulse.

Solution Subtract 1 in each case.

(a) 101010
$\underline{\quad -1}$
101001 First and second from right are complemented.

(b) 001100
$\underline{\quad -1}$
001011 Three

(c) 110000
$\underline{\quad -1}$
101110 Four

EXAMPLE 13.9 It is desired to reduce the frequency of a pulse signal to one half. Suggest a suitable circuit for the signal.

Solution
All it needs is a single *T* flip-flop with negative edge triggering. See Fig. 13.12.

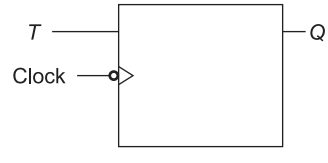


Fig. 13.12

EXAMPLE 13.10 Design a mod-3 counter (non-linear).

Solution

$2^1 < 3 < 2^2$

We need 2 flip-flops. $\bar{B}$ is connected to J of FF1 with synchronous triggering.

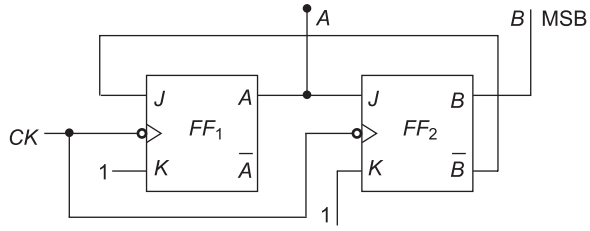


Fig. 13.13

- $A = 0, B = 0, \quad J_1 = \bar{B} = 1, J_2 = A = 0$, count 0
- On negative edge triggering (NT), count 1
  FF_1 toggles $\rightarrow A = 1, B = 0, \bar{B} = 1 = J_1, J_2 = A = 1$
- On next NT Count 2
  $A = 0, B = 1, \quad J_1 = \bar{B} = 0, \; J_2 = A = 0$
- On next NT, count 3
  $A = 0, B = 0$
- Next cycle starts

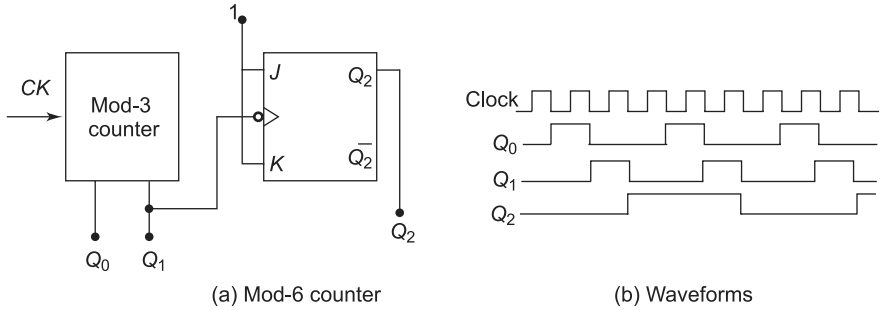
Truth table

Count	B	A
0	0	0
1	0	1
2	1	0
2	0	0 Next cycle

EXAMPLE 13.11 Convert Mod-3 counter to a Mod-6 counter.

Solution

At the output of the Mod-3 counter, connect a JK flip-flop in toggle mode as shown in Fig. 13.14. The result is a 2 × 3 = Mod 6 counter.



(a) Mod-6 counter (b) Waveforms

Fig. 13.14

Counter state

Count	Q_2	Q_1	Q_0
0	0	0	0
1	0	0	1
2	0	1	0
3	1	0	0
4	1	0	1
5	1	1	0
6	0	0	0 repeats

13.8 INTRODUCTION TO MICROPROCESSOR AND MICROCONTROLLERS

13.8.1 Microprocessor

Microprocessor is a multipurpose, programmable logic device which performs computations and controls all the operations of automated devices, mini-computers, etc. It acts as a central processing unit mounted over a single chip interconnected with external memory (including ROM, RAM, SDRAM, etc.), several input-output devices and peripherals, interfacing ICs, etc. The basic block diagram of a microprocessor is shown in Fig. 13.15.

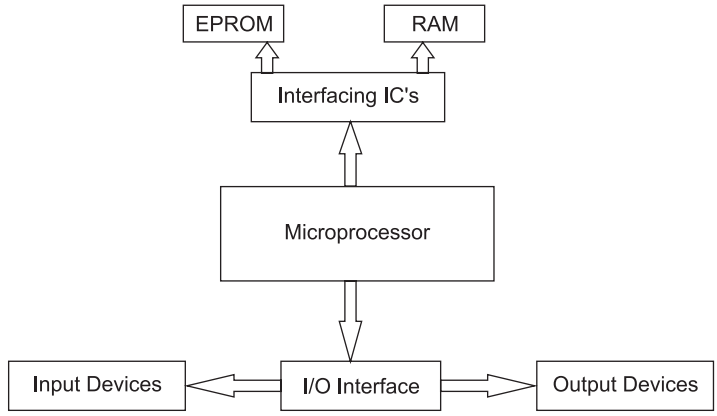


Fig. 13.15 Block diagram of microprocessor

The microprocessor includes:

1. **Arithmetic Logic Unit (ALU):** It is responsible for various mathematical computations like addition, subtraction, multiplication, etc. and logical operations like AND, OR, NOT, Exclusive-OR, etc. being performed by the microprocessor.
2. **Control Unit:** It provides all the necessary timing and control signals for smooth and synchronized functioning of the processor.

3. **Registers:** They are basically flip-flops used for data storage and may or may not be accessible to a programmer. There are different types of registers for each processor type. It includes accumulator, general purpose registers, instruction registers, temporary registers, index registers, etc.

Input devices include the components used to enter data into the processor. Typical examples are keyboards, mouse, toggle switches, etc. Output devices are used for result display or communication from the processor to the peripherals/end users. The commonly used devices are LEDs, printers, CRT displays, etc. The storage of data is done in externally attached memories. It may include RAMs, ROMs, EPROMs, SDRAM, floppy disks, etc. Different interfacing ICs are often required to act as buffer IC for protection, better, and efficient interfacing for memory and peripheral devices. Some of the devices include 8255 (PPI), 8257 (DMA Controller), 8279 (keyboard and display controller), 8259 (Programmable Interrupt Controller), 8253, 8254 (Programmable interval timer), etc.

All the components work in conjunction with each other to act as a system. The hardware part consists of physical set of equipment like transistors, resistors, capacitors, memory banks, and logic devices, etc. which are made operational to perform certain tasks through a sequence of instructions, called programs. The different units are connected through bus. The buses may be classified in accordance with direction of data flow, or type of signal being carried.

Buses are categorized as follows:

According to direction of data flow

1. **Unidirectional:** Here, data travel in one direction only, i.e. either from processor to device or peripheral, and vice versa.
2. **Bidirectional:** In this type, data may flow in either direction, i.e. to the processor or from the processor.

According to signal being carried

1. **Control Bus**: The control bus carries the entire control signal necessary for smooth functioning of system.
2. **Data Bus:** The data bus carries the data and decides the *word size* of microprocessor. A microprocessor with bus capable enough to handle 8-bit data at one time is said to have word size of 8 bit.
3. **Address Bus:** The address bus carries the address of memory or any of the peripheral devices called upon by the processor. It also represents the range of physical memory that can be addressed by the processor. The processor having n address lines can address maximum of 2^n memory locations.

All the components may be addressed using either von Neumann architecture or Harvard architecture.

1. **von Neumann Architecture:** It is the architecture commonly used in microprocessors and computers. Here, the processor consists of CPU, memory, and input-output devices. The actions to be performed by the processor are stored sequentially in memory and executed with the help of some special registers, like program counter, which stores the address of next instruction to be executed. The only limitation of this

architecture is that all the information must flow back and forth between processor and memory through a single channel which has a fixed bandwidth. When this bandwidth is fully occupied, the processing speed is reduced. To overcome this issue, parallel connected CPU's are used, with each CPU being the basic building block of Neumann processors.

2. **Harvard Architecture:** It involves the physical set-up where the memory is divided into two dedicated sections. One section is dedicated for storing the instructions, while the other one for data. This offers the advantage of increased speed due to increased bandwidth available.

The basic operation of processor involves fetching the instructions from the memory, decoding and executing them for performing certain actions. The instructions are written in fixed assembly language or higher level languages like C, C++, etc. They are stored in system memory in sequential manner. The microprocessor fetches first instruction into the processor by converting them into American Standard Code for Information Interchange (ASCII) code or Extended Binary Coded Decimal Interchange Code (EBCDIC). Processor executes the instructions and sends the data into memory. The cycle of fetch, decoding, and executing, continues till Halt instruction is received.

The instructions of microprocessor may be written in three different languages:

1. **Machine Language:** It is the basic level of programming where the instructions are available in binary codes and are understandable by processors. The instructions written by the programmer in any of the format are first converted into machine language before they are being fetched and executed.
2. **Assembly Language:** It consists of mnemonics (symbolic representation) of the functions to be performed by the microprocessor. They are usually shortened initials of english words for operation performed by the instruction. For example, mnemonic for addition is ADD, subtraction is SUB, etc. It must be translated into machine language by an assembler. The assembler does this work for the processor as it recognizes the syntax and writes the corresponding machine code for each instruction. The original assembly language program is known as the source code and corresponding program in machine language is called the object code.
3. **High Level Language:** The programming may be done using any high level language such as BASIC, C, C++, etc. A compiler is used to translate higher level language programmes into memory and then execute.

The instruction written in assembly language consists of two parts namely, **opcode and operand.** Opcode contains the information about the function to be performed, while operand contains the data on which the action is to be performed. The instructions may be classified as one-byte, two-byte, or three-byte instructions depending upon the number of consecutive spaces required for their storage. The necessary steps carried out to perform a fetch, a read, or a write, constitute a machine cycle. The necessary steps that a processor performs to fetch an instruction or data from memory (known as fetch cycle), and then execute (execute cycle) constitute an instruction cycle. It consists of several machine cycles.

The data in a microprocessor is often represented in fixed point, floating point, or decimal formats.

1. The **fixed point data** implements binary format through signed or unsigned numbers. The rightmost bit is referred as Least Significant Bit (LSB) and leftmost as Most Significant Bit (MSB). MSB acts as a sign bit for signed operations (0 for positive and 1 for negative numbers).
2. **Floating point representation** contains sign of mantissa, mantissa, sign of exponent, and exponent. It is used for convenience as sometimes it is easier to express a number in scientific notations (e.g., it is easier to express 1000000 in 1.0×10^6 form). Also, 32-bit number is called single precision floating point representation, while 64-bit number is called double precision floating point representation.
3. **Decimal representation** consists of 8-bit representation, four bits for zone (1111-EBCDIC and 0101 for ASCII-8) and the next four bits for decimal numbers.

13.8.2 Evolution of Microprocessors

The evolution of microprocessors started with 4-bit processor chips and have gradually grown into 8, 16, 32, 64-bit processors. Among the different processor manufacturing industries, like Intel, Motorola, Zilog, etc. Intel is the most popular and dominating one. The different stage of evolution of Intel processors is summarized in Fig. 13.16.

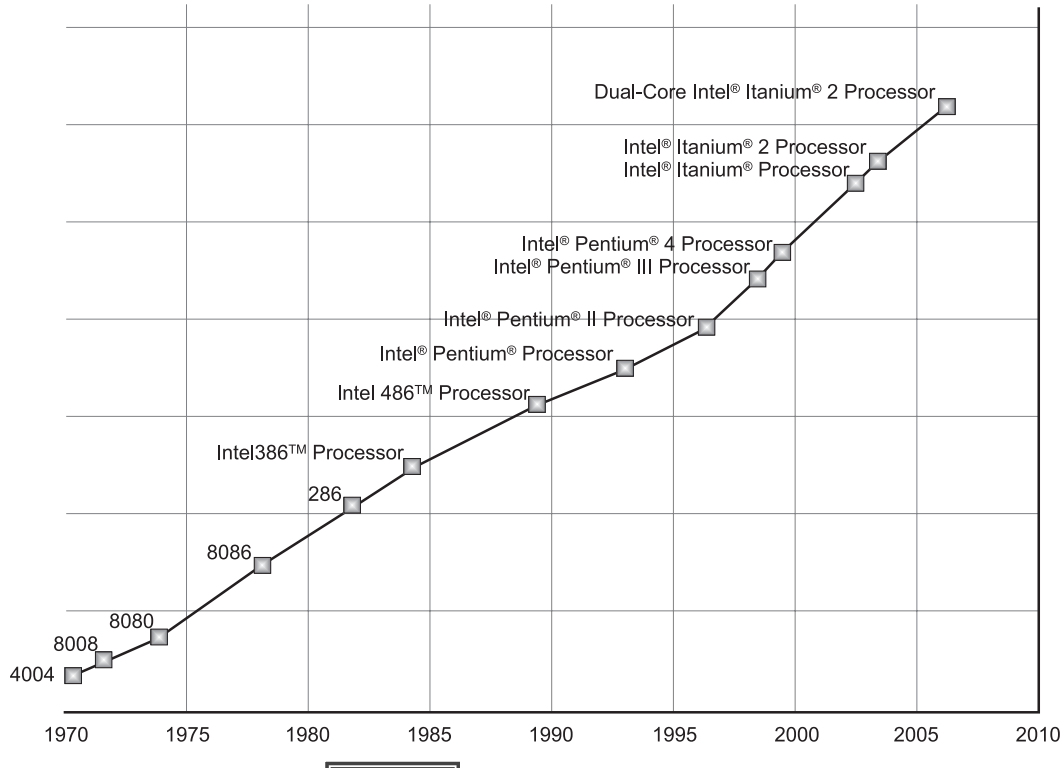


Fig. 13.16 Evolution of microprocessors

13.8.3 Microcontrollers

A microcontroller is, in many ways, analogous to microprocessor, but includes all the logic circuitry, memory, etc. on a single chip. It is designed to perform the specific tasks of embedded systems like displaying microwave's information, receiving remote signals, etc. The basic block diagram of a microcontroller is shown in Fig. 13.17.

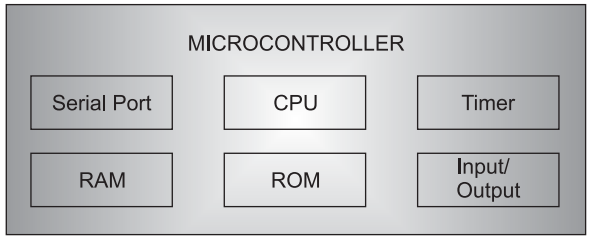


Fig. 13.17 Block diagram of microcontroller

The basic differences between a microprocessor and a microcontroller are discussed in Table 13.5.

Table 13.5 *Difference between microprocessor and microcontroller*

Microprocessor	Microcontroller
It has only CPU on chip, while all the other memory and peripheral devices are to be interfaced externally.	It has memory, serial ports, timer, etc. inbuilt on single chip only in addition to CPU.
Microprocessors are used for unspecific applications like software, games, etc.	Microcontrollers are used to execute a single task within an application.
The designing and hardware cost is high.	The designing and hardware cost is low.
Not so easy to replace.	Easy to replace.
Its power consumption is high because it has to control the entire system.	It is built with CMOS technology which requires less power to operate.
Cannot be used in compact systems and hence inefficient.	Can be used in compact systems and hence it is an efficient technique.
Since memory and I/O components are all external, each instruction will need external operation, hence it is relatively slower.	Since components are internal, most of the operations are internal instructions, hence speed is fast.
Microprocessors are based on von Neumann architecture where program and data are stored in same memory module.	Microcontrollers are based on Harvard architecture where program memory and data memory are separate.

13.8.4 Applications

The microprocessors find potential applications and implementation in computers as well as in non-computerised automatic household appliances like tools, toys, car control keys, lights, alarms, circuit breakers DVD players, cellular phones, HDTV broadcasting systems, etc. The

different applications of microprocessors may be categorised as reprogrammable control applications and embedded systems. Reprogrammable applications include mass storage devices, peripheral controls systems, industrial applications, etc. The embedded systems include the end user devices which will never be reprogrammed like printer, photocopier, washing machines, etc. The main reason behind the popularity of microprocessors and microcontrollers is the ability to revise the embedded software/control program easily. This helps in upgradations in the existing system with minimum requirement of redesigning and thereby negligible cost factors. Also, the practical application of various strategies would have been next to impossible without them or using traditional electromechanical devices. Nowadays, modern microprocessors can perform extremely sophisticated operations in areas such as meteorology, aviation, nuclear physics and engineering, and take up much less space as well as deliver superior performance.

SUMMARY

- Shift registers have been introduced and synchronous and asynchronous counters have been explained.
- The basic concept of microprocessors and microcontrollers, including block diagram, architecture used, and their applications, has been discussed.

EXERCISES

Review Questions

1. Explain shift registers with examples.
2. Distinguish between synchronous and asynchronous counters.
3. Draw the block diagram of a microprocessor based system and explain the function of each of the following components: microprocessor, memory, and input/output devices.
4. Explain the difference between machine language and assembly language of a computer.
5. Explain the difference between microprocessors and microcontrollers.
6. Give the different applications of microprocessors and microcontrollers in daily life.
7. Give the functions of different microprocessor buses.

Problems

1. How many flip-flops would be required for a shift register to store
   (a) 0.7-bit number (b) decimal number up to 64 (c) hexadecimal numbers up to I
2. A shift register has 12 flip-flops. What is the largest (a) decimal number, and (b) hexadecimal number that can be stored in it?
3. The hexadecimal number *AB* is stored in 54/74LS91 in Fig. 13.4. Show the waveform at the output, assuming that the clock is allowed to run for eight cycles and that $A = B = 0$. What hexadecimal number will remain stored if the clock is run for 4 more cycles.
4. Four *D* flip-flops are connected to form a shift register. The register is initially empty. The number 1001 is shifted into the register.

(a) Complete the following table.

Data	Pulse No.	Q_3	Q_2	Q_1	Q_0
1001					

(b) Also draw the waveform at third clock transition.

5. For S4/74164 of the Fig. 13.12, $B = 1$, clear = 1, and a 1 MHz clock is used to shift the decimal number 200 into the register at A. How much time will it take? Propose the table as below.

Data	Pulse count	Register state
8-bit		

6. Assume that binary number 1100 (left 1 MSB and right CS0) is present in the shift register of Fig. 13.2. It is desired to reverse the order of the stored bits. How can this be done? If the clock frequency is 1 MHz, what would be the time required to do so? Assume that parallel entry of data is not allowed.

7. Using four *D* flip-flops design a shift-left register and explain its operation.

8. Getting the idea from Fig. 13.7, draw the schematic of four *JK* flip-flops. Design an asynchronous down counter and write its truth table.

9. Getting the idea from Fig. 13.9, draw the schematic of a model synchronous down counter using *T* flip-flops and prepare its truth table.

10. The synchronous up/down counter, mod-16 by additional logic circuitry is preset at 1010. Write its truth table as clock pulses are applied for up count from this state.

11. An up counter, 6-bit binary, has the following states:
(a) 1101010 (b) 001100 (c) 110000
Determine the number of flip-flops which would be complemented after application of one clock pulse.

12. How many memory locations can be addressed with a processor of 24-bit address bus?

13. If 1048576 memory locations, each capable of storing an 8-bit data, are to be addressed by a CPU, calculate the number of address and data lines required.

➥ Multiple-Choice Questions

1. Which of the following is a type of registers?
(a) Serial in serial out (b) Serial in parallel out (c) Parallel in serial out (d) All of these

2. Asynchronous counters are also called
(a) ring counters (b) ripple counter (c) BCD counter (d) Mod-6 counter

3. A Mod-10 counter has a total of—states
(a) 10 (b) 15 (c) 9 (d) ∞

4. A counter can count in the direction
(a) up (b) down (c) up-down (d) all of these

5. A counter is
(a) a group of registers (b) a memory storage device
(c) a group of transistors (d) none of these

6. A 4-bit shift register will have
(a) 4 flip-flops (b) 8 flip-flops (c) 2 flip-flops (d) none of these

7. A shift register is used for
(a) storing and shifting binary information (b) to give specific sequence of binary number
(c) performing some special Boolean functions (d) none of these

8. Which of the following shift register will require minimum number of clock pulses to give the output?
(a) Serial In Serial Out (SISO) (b) Parallel In Serial Out (PISO)

(c) Serial In Parallel Out (SIPO) (d) Parallel In Parallel Out (PIPO)

9. Which of the following are special counter?
(a) Johnson counter (b) BCD counter (c) Ring counter (d) Both (a) and (c)

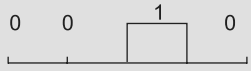
10. The total number of states a counter can count before reaching the first state is called
(a) modulus of the counter (b) mod number of the counter
(c) count of the counter (d) both (a) and (b)

ANSWERS

✦ Problems

3. A

4. (b)



6. Feedback serial output to serial input. Apply two clock pulses, 0011, 2 μs.

11. Answer (a) 1 (b) 1 (c) 1

12. 2^{24}

13. 20 address lines

✦ Multiple-Choice Questions

1. (d) **2.** (b) **3.** (a) **4.** (d) **5.** (a) **6.** (a) **7.** (a) **8.** (d) **9.** (d) **10.** (a)

CHAPTER 14

Digital to Analog (D/A) and Analog to Digital (A/D) Converters

GOALS AND OBJECTIVES

- ❏ Introduction of Analog to Digital (A/D) and Digital to Analog (D/A) converters
- ❏ Functional diagrams of converters (A/D and D/A)
- ❏ Calculating the accuracy and resolution of D/A and A/D converters with examples

14.1 INTRODUCTION

Digital-to-Analog (D/A) and Analog-to-Digital (A/D) conversion form two very important aspects of digital-data processing. Digital-to-analog conversion involves translation of digital information into equivalent analog information. As an example, the output of a digital system

might be changed to analog form for the purpose of driving a pen recorder and servomotors which drive the cursor arms of a plotter. In this respect, a D/A converter is sometimes considered a *decoding device*.

The process of changing an analog signal to an equivalent digital signal is accomplished by the use of an A/D converter. For example, an A/D converter is used to change the analog output signals from transducers (measuring temperature, pressure, vibration, etc.) into equivalent digital signals which can be fed to a digital system for digital processing. An A/D converter is often referred to an encoding device since it is used to encode signals for entry into a digital system.

Digital-to-analog conversion is a straightforward process and is considerably easier than A/D conversion. In fact, a D/A converter is usually an integral part of any A/D converter. For this reason, we consider the D/A conversion process first.

14.2 | DIGITAL-TO-ANALOG CONVERTER (DAC)

The analog output voltage V_A of an N-bit straight binary DAC converter is related to the digital equation

$$V_A = K(2^{N-1} b_{N-1} + 2^{N-2} b_{N-2} + \cdots + 2b_1 + b_0) \tag{14.1}$$

where K = proportionality factor

$$\left.\begin{aligned} b_j &= +1 \text{ if } j\text{th bit of input is } 1 \\ &= 0 \quad \text{if } j\text{th bit of input is } 0 \end{aligned}\right\}; \; j = 0 \text{ to } (N-1)$$

✦ Weight Resistor DAC

As the D/A conversion involves a weighted sum corresponding to the input to the converter, the summing circuit of Fig. 14.1 can be used as a DAC. In this circuit,

$$I_i = I_{N-1} + I_{N-2} + \cdots + I_1 + I_0 \tag{14.2}$$

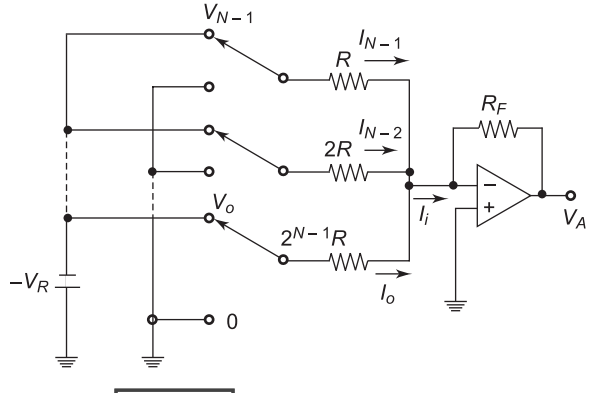


Fig. 14.1 Weighted resistor DAC

These currents can be expressed in terms of voltages as

$$I_{N-1} = V_{N-1}/R$$

$$I_{N-2} = V_{N-2}/2R$$

or $$I_0 = V_0/2^{N-1}\,R$$

Let $$V(0) = -b_0\,V_R$$

$$V(1) = -b_1\,V_R$$

Then $$V_{N-1} = -b_{N-1}\,V_R$$

The switches in the circuit are *digitally controlled* causing b_n to acquire values 0 or 1.

We can now express Eq. (14.2) as

$$I_i = -V_R\left[\frac{1}{R}b_{N-1} + \frac{1}{2R}b_{N-2} + \cdots + \frac{1}{2^{N-1}R}b_0\right] \tag{14.3}$$

The analog output of the circuit is given by

$$V_A = -R_F I_i$$

or $$V_A = V_R\left[\frac{R_F}{R}b_{N-1} + \frac{R_F}{2R}b_{N-2} + \cdots + \frac{R_F}{2^{N-1}R}b_0\right]$$

$$= K[2^{N-1}\,b_{N-1} + 2^{N-1}\,b_{N-1} + \ldots + 2^1 b_1 + 2^0 b_0] \tag{14.4}$$

where $$K = \frac{V_R R_F}{2^{N-1}R} \tag{14.5}$$

Hence, the circuit is a DAC.

For accurate conversion, the input voltages would have to be precisely known values, a condition that is not required in a digital system. The various weighing resistors would also have to be precisely formed, a very difficult requirement in an IC. In practice, the *R*-2*R* ladder discussed below is used.

✦ R-2R Ladder Network

An *R*-2*R* ladder network is shown in Fig. 14.2(a). It uses resistors of only two values, *R* and 2*R*. The inputs to the resistor network are applied through digitally controlled switches.

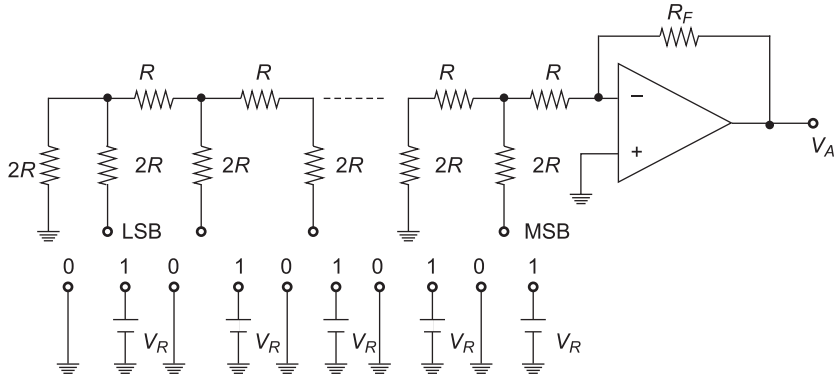
To analyse this circuit, consider a 3-bit *R*-2*R* ladder DAC as shown in Fig. 14.2(b), wherein the input is assumed as 001. Applying Thevenin's theorem at *XX'*, *YY'* and *ZZ'* we obtain circuits of Fig. 14.3 (a), (b) and (c) respectively. It is immediately seen that for input† 001, voltage $V_R/2^3$ is applied through 3*R* to the inverting terminal of the OP-AMP.

Similarly, for the digital inputs of 010 and 100, the equivalent voltages are $V_R/2^2$ and $V_R/2^1$ respectively with resistance 3*R* in each case. Therefore, we obtain equivalent circuit of Fig. 14.3(d) wherein the voltage V_A is given by

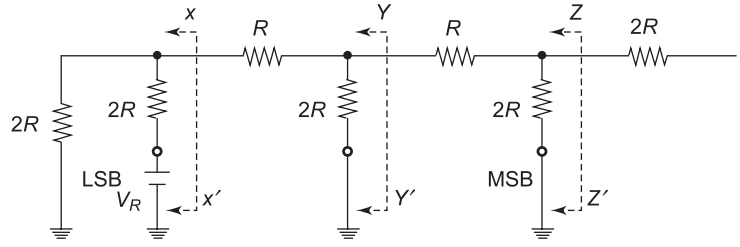
$$V_A = -\left[\frac{R_F}{3R}\frac{V_R}{2^3}b_0 + \frac{R_F}{3R}\frac{V_R}{2^2}b_1 + \frac{R_F}{3R}\cdot\frac{V_R}{2^1}b_2\right]$$

† Note that binary digit are in reverse order, i.e. LSB to MSB.

$$= -\left(\frac{R_F}{3R}\right)\left(\frac{V_R}{2^3}\right)\left[2^2 b_2 + 2^1 b_1 + 2^0 b_0\right]$$



(a) *R*-2*R* ladder DAC



(b) 3 bit R-2R ladder DAC network

Fig. 14.2 R-2R ladder network

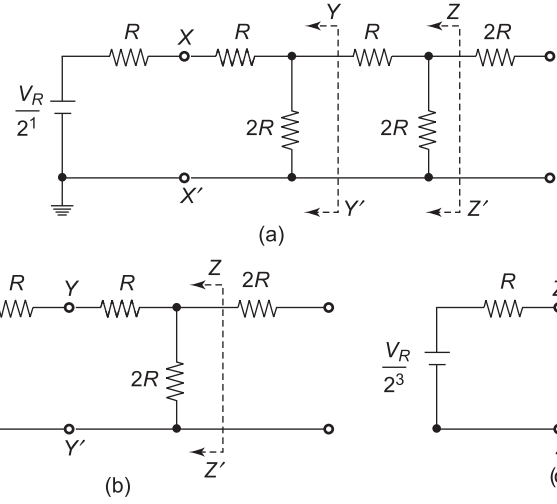


Fig. 14.3 *Contd.*

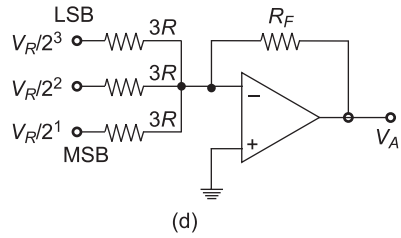


Fig. 14.3 Equivalent circuit of Fig. 14.2

The equation shows that the analog output voltage is proportional to the digital input. In general, for N-bit DAC,

$$V_A = V[2^N b_{N-1} + 2^{N-2} b_{N-2} + \ldots + 2^1 b_1 + 2^0 b_0] \quad (14.6)$$

where $R_F = 3R$ and $V_R = -2^N V$

The number of resistors required for an N-bit DAC is $2N$ in the case of an R-$2R$ ladder, whereas it is only N in case of a weighted-resistor network. But because of wide spread in the resistance values for large N, a weighted resistor DAC is not widely used.

✦ Implementation

The R-$2R$ ladder D/A converter is implemented by the circuit of Fig. 14.4. When the strobe pulse is high, the input bits are fed to the register (4 SR flip-flops). The register output is fed to the level amplifier connected to the position voltage above (+10 V). For the register output of $b = 1$, the level amplifier output of + 0 V is fed to the ladder. Hence, V_A = equivalent of form bits b_3, b_2, b_1, b_0.

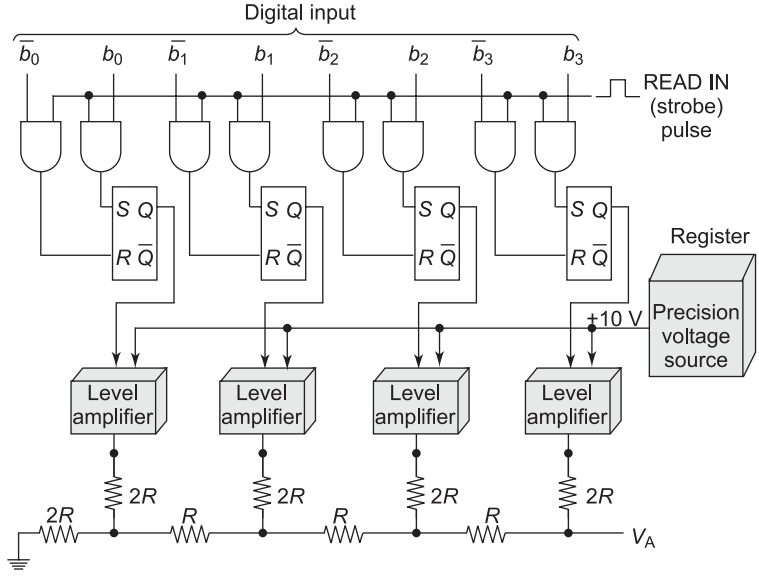


Fig. 14.4 4-bit D/A converter

$$V_A = 10\left\{\frac{b_3}{2^1} + \frac{b_2}{2^2} + \frac{b_1}{2^3} + \frac{b_0}{2^4}\right\};\ b_3 \text{ is MSB and } b_0 \text{ is LSB}$$

or
$$V_A = \frac{10}{2^4}(2^3\, b_3 + 2^2 b_2 + 2^1 b_1 + 2^0 b_0) \tag{14.7}$$

In general, for N bits,

$$V_A = \frac{10}{2^N}\ (2^{N-1} b_{N-1} + 2^1 b_1 + 2^0 b_0);\ b_{N-1} \text{ is MSB} \tag{14.8}$$

Figure 14.5 shows
Case I: V_A is connected to non-inverting unit gain op-amp.
Case II: V_A is connected to inverting op-amp

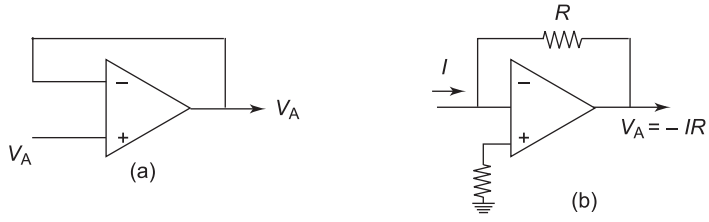


Fig. 14.5 V_A connected to non-inverting and inverting op-amps

14.3 ANALOG-TO-DIGITAL CONVERTER (ADC)

It is often required that data taken in a physical system should be converted into digital form. Such data would normally be in electrical analog form (voltage or current) that is picked up at the output of a transducer (like temperature, pressure, displacement, etc.) or electric voltage or current source. There are various ways in which electrical analog data can be converted to digital form. We shall consider following two methods:

1. The counting ADC
2. The parallel comparator ADC
3. Successive approximation

✦ Counting ADC

This system is shown in Fig. 14.6(a).

The clear pulse resets the counter to the zero count. The counter then records the number of pulses from the clock line in binary form. The clock is the source of pulses equally spaced in time. Since the number of pulses counted increases with time, the binary word representing this count is used as an input to the DAC, whose output is the staircase waveform shown in Fig. 14.6(b). The comparator has an output which is 'HIGH' and the AND gate is open for transmission of clock pulses. When V_d exceeds V_a, the comparator output goes 'LOW' and

AND gate is disabled. This stops the counting at the time when $V_a = V_d$ and the counter can be read as the digital output representing analog input voltage.

An improved version of the counting ADC called tracking or servo converter is shown in Fig. 14.7. To understand the operation of the system, assume initially that the output of the DAC is less than the analog input V_a. Then the positive comparator output causes the counter to read UP. The DAC output increases until it exceeds V_d. The UP-$\overline{\text{DOWN}}$ control line changes state, so that it now counts down and the count decreases by one count. This causes the control to change to count UP and the count increases by 1 LSB. This process keeps repeating so that digital output bounces back and forth by ±1 LSB around the correct value. This is also known as *a continuous converter*.

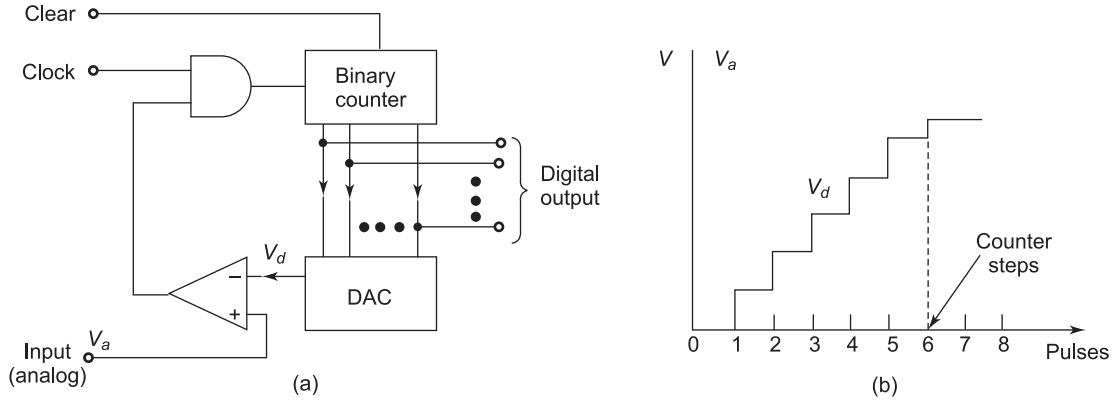


Fig. 14.6 Counting ADC

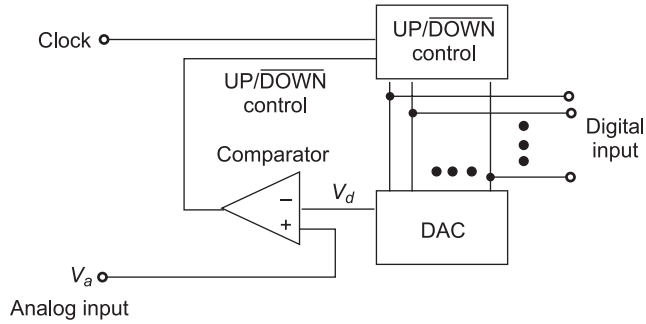


Fig. 14.7 A tracking ADC

14.4 SUCCESSIVE APPROXIMATION CONVERTER

The successive-approximation converter (SAC) is most useful. The block diagram for this type of converter is shown in Fig. 14.8. The converter operates by successively dividing the voltage ranges in half. The counter is first reset to all 0's, and the MSB is then set. The MSB is then left

in or taken out (by resetting the MSB flip-flop) depending on the output of the comparator. Then the second MSB is set in, and a comparison is made to determine whether to reset the second MSB flip-flop. The process is repeated down to the LSB, and at this time, the desired number is in the counter. Since the conversion involves operating on one flip-flop at a time, beginning with the MSB, a ring counter may be used for flip-flop selection.

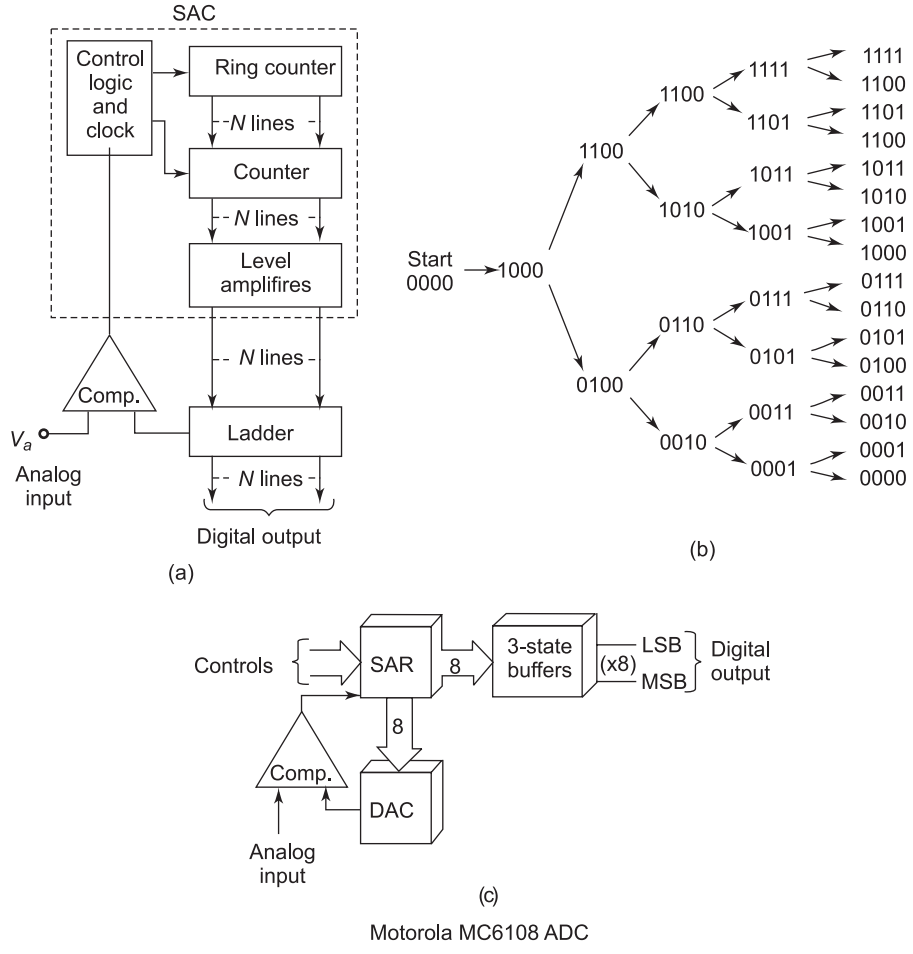


Fig. 14.8 A/D converter-successive approximation method

The successive-approximation method thus is the process of approximating the analog voltage by trying 1 bit at a time beginning with the MSB. The operation is shown in diagram in 14.8. It can be seen from this diagram that each conversion takes the same time and requires one conversion cycle for each bit. Thus, the total conversion time is equal to the number of bits N times the time required for one conversion cycle. One conversion cycle normally requires one cycle of the clock. As an example, a 10-bit converter operating with a 1 MHz clock has a conversion time of $10 \times 10^{-6} = 10^{-5} = 10\ \mu s$.

14.5 COMMERCIALLY AVAILABLE CONVERTERS

✦ D/A Converter

The block diagram with pin numbers is shown in Fig. 14.9.

$$I_{ref} = V_{ref}/R_{ref}$$

The current into the pin 4 is

$$I_0 = I_{ref}\left[\frac{A_1}{2^1} + \frac{A_2}{2^2} + \cdots + \frac{A_8}{2^8}\right]$$

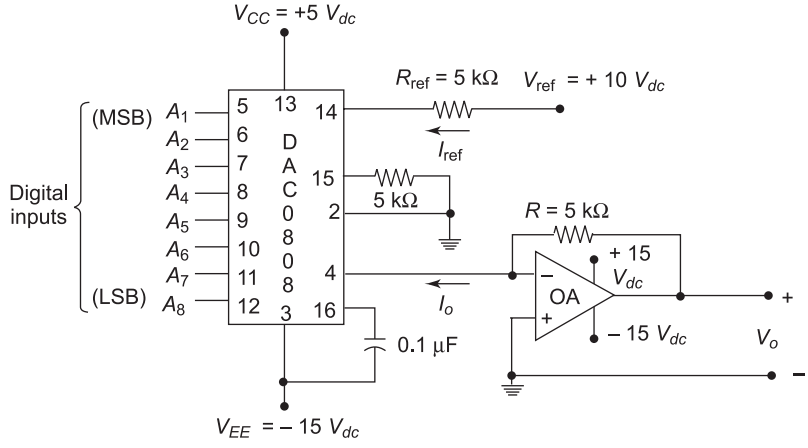


Fig. 14.9 DAC0808 D/A converter

The output voltage is

$$V_0 = I_0 R = V_{ref}\left(\frac{R}{R_{ref}}\right)\left[\frac{A_1}{2^1} + \frac{A_2}{2^2} + \cdots + \frac{A_8}{2^8}\right]$$

Choosing $R = R_{ref}$

$$V_0 = V_{ref}\left[\frac{A_1}{2^1} + \frac{A_2}{2^2} + \cdots + \frac{A_8}{2^8}\right] \tag{14.9}$$

If all $A = 0$, $V_0 = 0$

If all $A = 1$ $V_0 = 0.996\ V_{ref}$; it does reach V_{ref}

✦ A/D Converter

The ADC0804 is an inexpensive and very popular A/D converter. It is an 8-bit CMOS successive approximation converter, which can digitise 0 V to +5 V dc. Its diagram is shown in Fig. 14.10. The controls can be circuited for continuous conversion-free running mode.

The 10 kΩ resistor and 150 pF capacitor establish the operating frequency $f = 1/1.1\,(RC)$. In this case,

$$f = \frac{1}{1.1 \times (10\ \text{k}\Omega)} \frac{1}{(150\ \text{pF})} = 607\ \text{kHz}$$

To begin the operation, the start switch is momentarily activated.

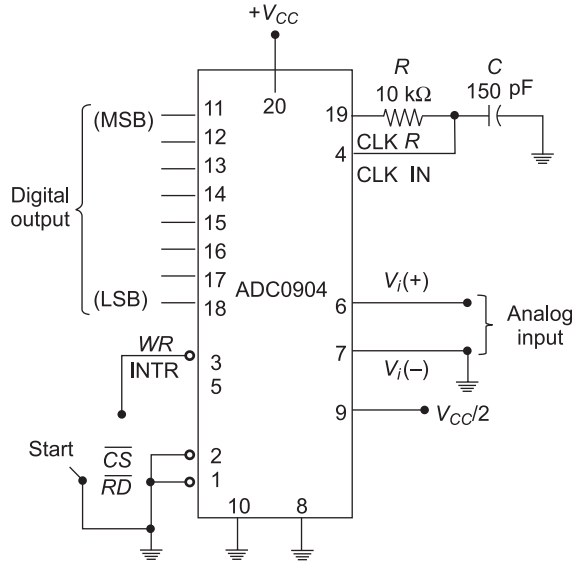


Fig. 14.10 ADC0804 A/D converter

14.6 ACCURACY AND RESOLUTION

The accuracy and resolution of D/A and A/D converter are explained here.

✦ Digital-to-Analog Accuracy and Resolution

❑ ***Accuracy*** It is primarily a function of the accuracy of the precision resistors of the ladder and the precision of the reference voltage supply. Accuracy is a measure of how close the actual voltage is to the theoretical output value.

$$\text{Accuracy} = \frac{\text{Theoretical voltage} - \text{Actual voltage}}{\text{Theoretical voltage}}$$

$$\text{Accuracy (say)} = \frac{10 - (9.9 \text{ or } 10.1)}{10} = 0.01 \text{ or } 1\ \%$$

❑ ***Resolution*** It is the smallest change in voltage that can be discerned. It is a function of the number of bits and corresponds to voltage change caused by LSD.

For 4-bit input, this is $\frac{1}{16}$ of the input voltage corresponding to LSD. For input voltage of 16 V, it is 1 V. For better resolution, more input bits are needed. For 10-bit input, the resolution is

$$\frac{1}{2^{10}} = \frac{1}{1024}$$

For full-scale output of 10 V, the minimum voltage change is

$$= \frac{10}{1024} \approx 1\,\text{mV}$$

✦ Analog-to-Digital Accuracy and Resolution

An A/D converter is a feedback system. Assuming that the components are precise and function properly; there are two sources of error. These are:

- **DAC resolution**, which is determined by LSD weight called *quantizing error*.
- **Comparator error**, the output ladder voltage which causes a variation in switching point. The comparator senses the voltage difference by a small (very) amount before it switches state. Switching is also affected by ripple and noise.

Differential linearity is a measure of variation in step change that causes the DAC to switch state.

The quantizing error, ladder error and comparator error may be assumed to be additive. For an 88-bit converter, quantisation error = $\frac{1}{256}$ = 0.4 percent. To achieve an overall accuracy of 1 percent, converter accuracy 0.5 percent, ladder accuracy is 0.1 percent and comparator accuracy of 0.2 percent is acceptable.

EXAMPLE 14.1 Find the output voltage for a 5-bit ladder that has digital input 10011. Assume input to be +10 V.

Solution Proceeding from MSB to LSB,

$$V_A = 10\text{ V}\left(\frac{1}{2^1} + \frac{0}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} + \frac{1}{2^5}\right)$$

$$V_A = \frac{10}{2^5}(2^4 + 0 + 0 + 2^1 + 2^0) = \frac{10}{32}(16 + 2 + 1)$$

$$V_A = \frac{10}{32} \times 19 = 5.9375\text{ V}$$

EXAMPLE 14.2 An 8-bit A/D converter-type inverter has 500 kHz clock. Find (a) maximum conversion time, (b) average conversion, and (c) maximum conversion rate.

Solution

$$\text{Counter cycle time} = \frac{10^6}{500 \times 10^3} = 2\ \mu s$$

Counter advance by 1 count is 2 μs
Maximum count = $2^8 = 256$

(a) Maximum conversion time = $256 \times 1 = 512\ \mu s$

(b) Average conversion time = $\frac{1}{2} \times 512 = 256\ \mu s$

(c) Maximum conversion rate = $\frac{10^6}{512} = 1953$ conversions/s

EXAMPLE 14.3 What is the resolution of a 9-bit D/A which uses ladder network? Express it as percent. What is the resolution in volts if the full-scale converter output is 10 V?

Solution The LSB in a 9-bit system has a weight of $\frac{1}{2^9} = \frac{1}{512}$. Thus, the resolution is 1 part in 512; as a percent $\frac{100}{512} = 0.2$.

$$\text{Resolution in voltage} = \frac{10}{512} \times 10^3 = 20\ \text{mV}$$

EXAMPLE 14.4 What is the quantizing error of a 10-bit A/D converter? What overall accuracy do you expect if the analog part error is 0.1 percent?

Solution Quantizing error corresponds to LSD = $\frac{1}{2^{10}} = \frac{1}{1024} \approx 0.1$ percent

Analog part accuracy = 0.1 percent
Expected overall accuracy = 0.1 + 0.1 = 0.2 percent

EXAMPLE 14.5 Consider the resistor divider of Fig. 14.11; determine V_A, the node voltage.

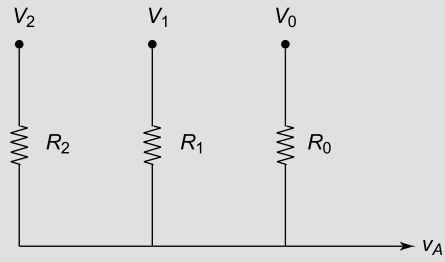


Fig. 14.11

Solution Sum of nodal current is

$$(V_0 - V_A)/R_0 + (V_1 - V_A)/R_1 + (V_2 - V_A)/R_2 = 0$$

or $$\frac{V_0}{R_0} + \frac{V_1}{R_1} + \frac{V_2}{R_2} = V_A\left(\frac{1}{R_0} + \frac{1}{R_1} + \frac{1}{R_2}\right)$$

or $$V_A = \frac{V_0/R_0 + V_1/R_1 + V_2/R_2}{1/R_0 + 1/R_1 - 1/R_2}$$

This result follows directly the application of *Millman's theorem* which states that the node voltage is obtained by the sum of currents entering the node, assuming node voltage to be zero which is decided by the sum of conductances connected to the node.

EXAMPLE 14.6 How can the output voltage V_A be made to represent the binary number $b_2b_1b_0$?

Solution

Let $R_1 = R_0/2$, $R_2 = R_0/4$

$V_0 = b_0V$, $V_1 = b_1V$, $V_2 = b_2V$; b_i's can be 0 or 1.

We can then write

$$V_A = \frac{V(2^0b_0 + 2^1b_1 + 2^2b_2)/R_0}{\frac{1}{R_0}(2^0 + 2^1 + 2^2)}$$

$$V_A = V\left[\frac{2^0b_0 + 2^1b_1 + 2^2b_2}{(2^3 - 1)}\right]$$

Or, in general,

$$V_A = V\left[\frac{(2^{N-1}b_{N-1} + \cdots + 2^1b_1 + 2^0b_0)}{(2^N - 1)}\right] \tag{14.10}$$

EXAMPLE 14.7 In a resistor divider of 6 bits, what is the weight of each bit?

Solution

$$N = (2^N - 1) = 2^6 - 1 = 63$$

The weights of 6 bits are

	$\frac{1}{63}$,	$\frac{2}{63}$,	$\frac{4}{63}$,	$\frac{8}{63}$,	$\frac{16}{63}$,	$\frac{32}{63}$
Bits input (say)	0	1	1	1	0	1

SUMMARY

- A/D and D/A converters are introduced with their functional diagrams.
- Their accuracy and resolution with examples are discussed.

EXERCISES

➥ Review Questions

1. Distinguish between weighted resistor DAC and *R*-2*R* ladder DAC. What is preferred and why?
2. What is the purpose of a level amplifier in DAC?
3. Draw the block diagram of the DAC of Fig. 14.4. Briefly explain its operation.
4. What are the various types of A/D converters? Which is the fastest? What limits its use?
5. What is a parallel ADC? How many comparators does this need? How is the comparator output converted to digital form?
6. Compare successive-approximation A/D converter, with counter-type converter. Which requires less conversion time and why?
7. What determines the resolution of a D/A counter?
8. What is meant by quantizing error in a A/D converter?

➥ Problems

1. Draw the schematic of a 6-bit resistor divider. If the divider has a full-scale output of +10 V, find the following:
   (a) Change in output due to change in LSB
   (b) The output voltage for an input of 1001110
2. What is the full-scale output of 6-bit binary ladder if 0 = 0 V and 1 = +10 V? What is the output for input of 110001?
3. A 12-bit D/A converter has full-scale output of 10 V. What are its percent resolution and voltage resolution?
4. Find the following for a 12-bit converter-type ADC using a 1 MHz clock:
   (a) Maximum conversion time
   (b) Average conversion time
   (c) Maximum conversion rate

5. What is the full-scale output of an *R*-2*R* ladder for
   (a) 2-bit number (b) 5-bit number
   Hint: Take 0 = –10 V and 1 = +10 V
6. What is the output of an *R*-2*R* ladder for the following 4-bit input patterns?
   (a) 1010 (b) 0101
   Hint: Take $R_F = 3R$, 0 = +0 V, 1 = +10 V
7. What clock frequency must be used with an 8-bit counter-type A/D converter which can make 5000 conversions per second?
8. What is the conversion time of a 12-bit successive-approximation-type ADC using a 1 MHz clock?

➥ Multiple-Choice Questions

1. An 8-bit converter is used for a dc range of 0 to 10 V. Find the weight of LSB.
   (a) 39 mV (b) 78 mV (c) 39.2 mV (d) None of these
2. A successive-approximation A/D converter has a resolution of 20 mV. What will be its digital output for an analog input of 2.17 V?
   (a) 01101100 (b) 01101101 (c) 01101011 (d) None of these
3. An analog transducer has a range of 0–10 V. Calculate bits of an ADC if the resolution is 5 mV.
   (a) 10 (b) 9 (c) 11 (d) None of these
4. A 0–10 V ADC has to have a resolution of 0.025%. Find the rms value of quantizing error.
   (a) 176 μV (b) 352 μV (c) 705 μV (d) 1410 μV
5. A 5-bit DAC has a current output. The digital input is 10100 and the output current corresponding to this is 10 mA. What will be the output current for digital input 11101?
   (a) 10.5 mA (b) 29 mA (c) 14.5 mA (d) None of these

➥ Fill in the Blanks

1. The resolution of a 12-bit D/A converter is _____ of the full-scale output.
2. The maximum quantizing error for an A/D converter is _____ of maximum analog input voltage.
3. The number of comparators required for an 8-bit parallel comparator A/D converter is _____.
4. The linearity of a D/A converter is specified as _____LSB.
5. The maximum sampling time T_s of a 10 kHz sinusoidal voltage for conversion to digital form is _____.

ANSWERS

✦ Problems

1. (a) $\frac{10}{63}$ V, (b) $\frac{10}{63}(2 + 4 + 8 + 32)$ **2.** 9.844 V, 7.656 V

3. $\frac{1}{4096}$ or 0.0244% and 2.44 mV **4.** (a) 4.096 ms, (b) 2.078 ms, (c) 2.44 s

5. (a) –7.5 V, (b) –9.6875 V **6.** (a) = 5.25 V (b) –3.125 V

7. 1.28 MHz **8.** 12 μs

✦ Multiple-Choice Questions

1. (a) **2.** (a) **3.** (c) **4.** (c) **5.** (c)

✦ Fill in the Blanks

1. 1/4095 **2.** 1/510 **3.** 255 **4.** Less than $\pm\frac{1}{2}$ **5.** 0.05 ms

CHAPTER 15

Transducers

GOALS AND OBJECTIVES

- ❑ Mechanical transducers
- ❑ Passive electrical transducers
- ❑ Active electrical transducers

15.1 | INTRODUCTION

A transducer is a device or combination of elements, which responds to the physical condition or chemical state of a substance and converts it into an output signal. The output signal may be an electrical or mechanical parameter, which can be easily measurable. If a transducer produces mechanical signal as its output then it is called a *mechanical transducer*. The transducers that produce electrical signals as output are called *electrical transducers*. Sometimes two transducers connected in cascade may produce an electrical quantity in the output terminals as shown in Fig. 15.1.

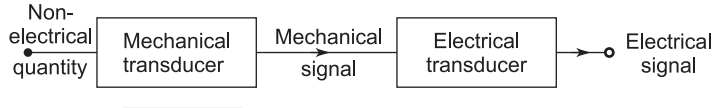


Fig. 15.1 Transducers connected in cascade

A *sensor*, or primary sensing element, is the first element, which is not directly coupled to the system under examination. A transducer may be considered to have accomplished the function of measurement by drawing an insignificant amount of power and energy from the system under study whereas a sensor does it by standing away without getting into physical contact with the medium or system under examination. Intensity and luminance of a source may be measured by sensors, whereas temperature is measured by transducers.

Based on the role of transducers, it can be classified into input and output transducers. An input transducer can be used as a measurement device and is known as an *instrument transducer*. An output transducer is known as a *power transducer*; it delivers output signals like force, torque, pressure or displacement when the electrical signal is applied as an input.

Based on the operation, transducers are classified into active and passive transducers. *Active transducers* develop the voltage and current as the output signal from the physical quantities being measured. But external energy is required for *passive transducers* to create the electrical output signals.

Thermo-couple, piezoelectric transducers, photoelectric cell and photovoltaic cell are the examples of the active transducer. Resistance strain gauge, thermistor, Linear Variable Differential Transformer (LVDT), Hall Effect sensor and photomultiplier tube are the examples of passive transducers.

Transducers should have satisfactory static and dynamic characteristics. The accuracy, precision, repeatability, reproducibility, stability, sensitivity and linearity are the static characteristics. Dynamic error, fidelity, bandwidth and speed of response are the dynamic performance characteristics.

15.2 MECHANICAL TRANSDUCERS

Mechanical quantities like temperature, pressure, flow, density, speed, velocity, acceleration, altitude and distance are the most important quantities to be controlled in any industry. Accuracy of the control of the above quantities is not possible without the accuracy of measurement. Table 15.1 lists transducer with the mechanical quantities to be measured.

Table 15.1 *Mechanical transducers*

Mechanical quantity to be measured	Transducer
Temperature	• Bimetallic element
Pressure	• Metallic diaphragms
	• Bourdon tubes
Force	• Spring balance
	• Cantilever
	• Diaphragms
	• Hydraulic load cells
Torque	• Torsion bar

(*Contd.*)

	• Flat spiral springs
	• Gyroscope
Density	• Hydrometer
	• U-tube weighting system
Liquid level	• Float elements
	• Manometer
Viscosity	• Capillary tube
Flow rate	• Pitot tube
	• Wave system
Displacement	• Flapper nozzle system
Vehicle attitude	• Gyroscope

A **transducer** is a device that converts a signal in one form of energy to another form of energy. The term *sensor* should be distinguished from *transducer*. The latter is a converter of one type of energy into another, whereas the former converts any type of energy into *electrical*. Transducers are widely used in measuring instruments.

The mechanical transducers are the mechanical elements that are used for converting one form of energy into other form that can be measured easily.

✦ Strain Gauge

If a resistor of fine wire is distorted then its resistance changes as a result of dimensional change. The result is a resistor for which the resistance is related to strain. Such a device is called a strain gauge. Strain gauges find extensive use in mechanical and biological measurements. Often the fine wire is attached to some flexible insulating substrate such that, when the substrate is flexed the actual wire itself is also flexed. Some strain gauges are made by depositing thin metal film directly onto the substrate. Since the resistance change is usually a very small fraction of the total resistance, it is very important to use a very sensitive resistance measurement technique. A second difficulty is that, small changes in temperature of the gauge can masquerade as a change in strain. One most often uses a Wheatstone bridge configuration with a temperature controlled strain gauge in order to maximize the sensitivity of the measurement.

There are two main types of Displacement Sensors. Linear Variable Differential Transformer (LVDTs) are differential devices that can detect motion and position with incredible sensitivity and linearity. The second type is a capacitive position sensor – this device has a smaller dynamic range than the LVDT but has even higher sensitivity. LVDT's internal structure consists of a primary winding centered between a pair of identically wound secondary windings, symmetrically spaced about the primary. The coils are wound on a one-piece hollow form of thermally stable glass reinforced polymer, encapsulated against moisture, wrapped in a high permeability magnetic shield, and then secured in cylindrical stainless steel housing. This coil assembly is usually the stationary element of the position sensor. The moving element of an LVDT is a separate tubular armature of magnetically permeable material called the core, which

is free to move axially within the coil's hollow bore, and mechanically coupled to the object whose position is being measured. This bore is typically large enough to provide substantial radial clearance between the core and bore, with no physical contact between it and the coil. In operation, the LVDT's primary winding is energized by alternating current of appropriate amplitude and frequency, known as the primary excitation. The LVDT's electrical output signal is the differential AC voltage between the two secondary windings, which varies with the axial position of the core within the LVDT coil. Usually this AC output voltage is converted by suitable electronic circuitry to high level DC voltage or current that is more convenient to use. As an alternative, to avoid the use of magnetic materials one can place the primary coil on the moving object with the two secondary fields picking up its driving field.

There are number of mechanical transducers, some of the commonly used ones are described below.

❑ ***Bellows*** These are the elastic elements, that convert the air pressure into displacement and it is commonly used for the measurement of pressure.

❑ ***Bourdon Tube*** This elastic tube converts air pressure to the rotary motion of the pointer used to indicate the pressure.

❑ ***Spring*** The spring tends to expand when force is applied to them; thus they are used for the measurement of force.

❑ ***Proving Rings*** Like the springs, the proving rings also convert applied force to the displacement.

❑ ***Diaphragm*** It converts applied pressure to the displacement.

❑ ***Manometer*** The manometer converts the applied pressure into variable displacement of the liquid within it, enabling to measure the pressure.

❑ ***Thermocouple*** Thermocouple is the device that produces electric current when one of its end is heated. The current produce by the device can be measured, which can be calibrated against the temperature enabling us to measure the temperature of the body.

❑ ***Bimetals*** These are the bimetallic strips comprising of two different metals having different coefficient of thermal expansion, joint together. When the strip is heated, one metal expands lesser while the other metal expands more leading to the deflection of the bimetallic strip, which is converted into the rotary motion of the pointer that indicates the temperature.

❑ ***Hydropneumatic Transducers*** These include devices like orifice, venturi, pitot tube, vanes and turbines that are used for measurement of pressure, velocity, flow rate and force of water.

Apart from the mechanical transducers mentioned above there are many others like seismic mass, pendulum scale, float etc. Most of the mechanical transducers are used as the primary transducers, meaning the initial input is applied to them, while the output obtained from them can be used directly to measure the quantity or it can be given as input to the secondary transducer, which are mostly of electrical type.

15.3 PASSIVE ELECTRICAL TRANSDUCERS

The three passive elements in an electric circuits are resistor, inductor and capacitor. The transducers that are based on the variation of the parameters due to application of any external stimulus are known as passive transducers.

✦ Resistive Transducers

The dc resistance R can be given by

$$R = \frac{\rho l}{a} \tag{15.1}$$

where l = conductor length in m
a = area of cross section in m^2
ρ = specific resistivity in Ω-m

Change in the value of the resistive element can be brought about by subjecting the element to external stimulus that affects either dimensions of the element or its resistivity. The dimensional changes can be brought about by subjecting the resistive elements, to pressure, force or torque, directly or by means of some primary transducers. Resistive strain gauges enable measurement of strain of mechanical members on to which they are bonded. The resistivity of the material medium constituting the path of the current in the resistor varies with the temperature and composition of the medium. Resistance thermometers are known to be exceptionally good for temperature measurements.

15.3.1 Resistance Thermometers

In resistance thermometers, the temperature changes are measured in terms of resistance changes. The resistive element is usually made of a solid material, a metal, metallic alloy or a semiconductor compound. The resistivity of the metal increases with temperature, but in semiconductors and insulators generally decreases.

Wire-wound elements employ considerable length of wire, and it is free to expand. The length also increases with increase in temperature. Hence, as temperature changes, the change in resistance will be due to changes, in both length and resistivity. The temperature coefficient of resistance, is given by

$$\alpha = \frac{1}{\Delta T} \cdot \frac{\Delta R}{R_0} \tag{15.2}$$

ΔT = change in temperature, °C
$\Delta R / R_0$ = fractional change in resistance
R_0 = resistance at 0°C

The resistance R_T at any other temperature T°C is given by $R_T = R_0(1 + \alpha T)$ (15.3)

The wire resistance thermometers as shown in Fig. 15.2 usually consist of a coil wound on a mica or ceramic former, which also serves as a support mount for the coil. This element is normally enclosed in a protective tube of pyrex glass, porcelain, quartz or nickel. The element is brought to contact with the fluid whose temperature is to be measured. The two terminals of

the Platinum wire is connected with Wheatstone bridge for measuring the change in resistance and so the temperature is measured [Eq. (15.3)].

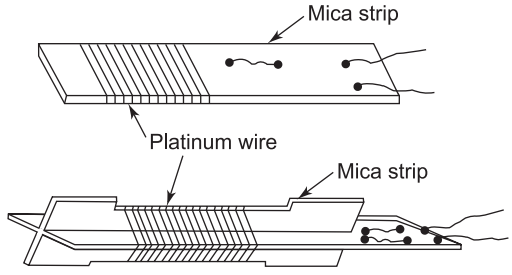


Fig. 15.2 Resistance thermometer

✦ Thermistor

It is a two-terminal semiconductor slab whose resistance decreases with increase in temperature, unlike a metal. Materials employed are oxides of cobalt, nickel, copper, iron, uranium and manganese.

A thermistor has very high negative coefficient of temperature 3–5% per °C, ideal for temperature measurement. Its resistance can be empirically expressed as

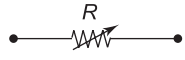
$$R = R_0 \exp \beta\left(\frac{1}{T} - \frac{1}{T_0}\right) \tag{15.4}$$

where R_0 = resistance at T_0 (k)
R = resistance at T(k)
β = constant to be determined experimentally

For large values of T, it can be approximate as

$$R = R_0 \exp\left(\frac{\beta}{T}\right)$$

❑ *Symbol*



✦ Parameters of Interest

- **Time constant**—Time required for resistance to fall to 63% of the final value; Range: 1–50 s
- **Dissipation factor**—Power dissipated in watts/temperature in °C; Range: 1–10 mW/°C
- **Resistance ratio**—R(20°C)/R(125°C): Range: 3–60

✦ Application

Measurement of

1. Temperature
2. Flow and pressure

3. Liquid level
4. Voltage and power
5. Vacuum
6. Thermal conductivity

✦ Resistive Displacement Transducers

The simplest form of converting linear or angular displacement into a change of resistance is the resistive element provided with a movable contactor. The change in resistance is brought out by only a change in length or portion of the resistor from one end to the point of contact. The transducer of resistance *Rp* is usually energised by dc or ac supply, and electrical output signals are obtained by using simple electrical circuits shown in Fig.15.3. For angle measurement, the resistor elements is in circular form and the contactor is rotable as shown Fig. 15.3(a). Various other way of measuring x or θ are shown in Fig. 15.3(b), (c) and (d).

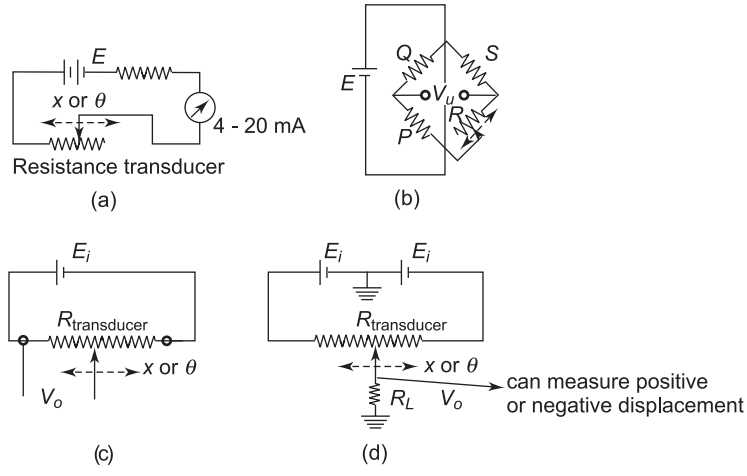


Fig. 15.3 Resistance transducer and its different circuits

✦ Resistive Strain Transducer

Strain-gauge pressure transducers and strain-gauge accelerometers are employing strain gauges as the secondary element along with a suitable primary mechanical transducer for converting the basic quantity under measurement into stress. The resistances wire strain gauges are available in two forms: unbonded-type and bonded-type systems. They employ four sapphire posts and hold four equal lengths of tungsten-platinum wire of 5 μm diameter, which are common as the four wires can form a Wheatstone bridge circuit providing temperature compensation. The four posts as shown in Fig. 15.4 are mounted on a star spring structure. When the centre of the star spring is subjected to the force under measurement, the star spring flexes, with each pair of the strain-gauge elements on opposite sides going into strain of opposite polarity. The resulting unbalance voltage of the bridge circuit is proportional to the force and hence, the pressure on the diaphragm. The whole assembly is encased in housing with provisions for admission of pressure and for electrical connections.

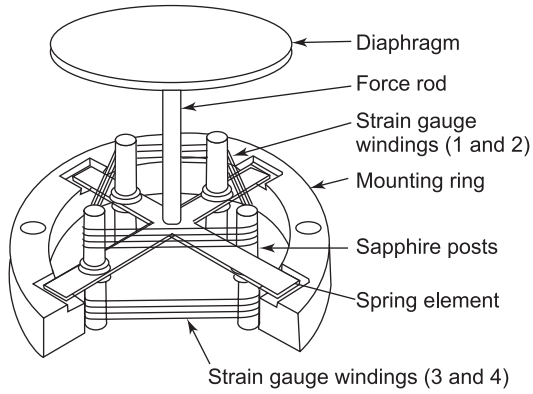


Fig. 15.4 Unbonded-type pressure-measurement transducers

The bounded-strain transducer is used for stress analysis. The pressure transducer uses a variety of sensing devices to provide an electrical output proportional to applied pressure. The sensing device employed in the transducer under discussion is bonded, metal-foil strain gauges.

While designing a strain-gauge pressure transducer, two fundamental considerations are kept in mind—one is the mechanical pressure-sensing element and the other is the electrical strain gauge bridge. The sensing element is typically a diaphragm or tube whose internal volume contains applied pressure. The fluid pressure causes the element to deflect in a predictable manner causing surface strains as well as applied force. The strain gauge is bonded to the non-pressurised face of the sensing element and responds to the surface strains or the strain gauge can be bonded to a separate structure usually a cantilever beam, driven by the force input of the diaphragm.

15.3.2 Inductive Transducers

In this, the self-inductance of a coil or mutual inductance of a pair of coils is altered in value due to a variation in the value of the quantity under measurement. This type of transducers posses considerable sensitivity and scope of application, especially for displacement and thickness measurements.

✦ Linear Variable Differential Transformer (LVDT)

The LVDT is based on mutual inductance type with variable coupling between the primary and the two secondary coils. It consists of a primary coil, uniformly wound over a certain length of the transducer, and two identical secondary coils symmetrically wound on either side of the primary coil and away from the centre as shown in Fig. 15.5(a). The iron core is free to move inside the coil in either direction from the central (null) position. When the primary is excited by ac supply, the secondary emfs are equal to each other with the core lying in the central position. The secondaries are connected in series but in phase opposition so that the net voltage is zero. Displacement of the core in either direction from the centre position results in output voltages proportional to displacement but of opposite polarity. See Figs 15.5(a) and (b).

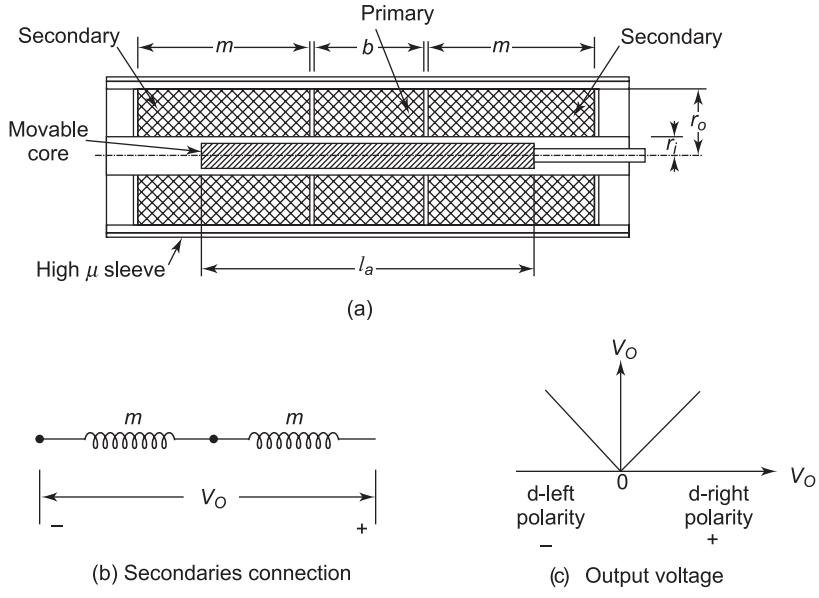


Fig. 15.5 Linear Variable Differential Transformer (LVDT)

15.3.3 Capacitive Transducers

It is a non-loading, non-contact and non-invasive type of transducer for displacement measurements. Capacitive transducers are also known as *proximity transducers*. They measure the nearness of an object without any mechanical coupling between them. The only coupling is through the electrostatic forces of attraction between the object and one plate of the capacitor.

The capacitance is given by

$$C_0 = \varepsilon_0 \, \varepsilon_r \, A/d \tag{15.5}$$

ε_0, ε_r = absolute and relative permittivity, respectively
A = area of the plate in m^2
d = separation between the plates in m

✦ Proximity Transducers

In some applications, the proximity of an object with respect to the fixed plate of the transducers is desired. Electrical circuits that develop output voltages proportional to the separation between the plates are available. This transducer giving an output signal ρ_0 proportional to x_0 is shown in Fig. 15.6. The output signal ρ_0 is given by

$$\rho_0 = \frac{C_f \, x_0}{\varepsilon_0 \, A} E_m \sin \omega t \tag{15.6}$$

C_f = capacitance of the standard capacitor
e_{ex} = $E_m \sin \omega t$ = sinusoidal applied voltage

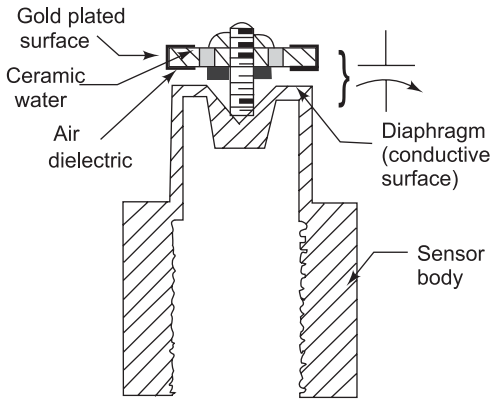


Fig. 15.6 Proximity transducers or capacitance transducers

If the object is vibrating, the amplitude of vibrations can be measured and in such a case, ρ_0 is an amplitude-modulated wave. Capacitor transducers can also be used for displacement, strain, pressure, microphone, level, moisture measurement, etc.

From Fig. 15.6, the proximity transducers are based on the fact that dielectric constants of liquid, solid & gases change under pressure. The capacitance pressure relation is non-linear and is affected by temperature variation. The measurement of this capacitance is done by a resonance circuit.

The distance between the proximity transducer plates is variable, the area of the plates and the dielectric constant remain constant. This is the most commonly used type of variable capacitance transducer.

For measurement of displacement of the object, one plate of the capacitance transducer is kept fixed, while the other is connected to the object. The changed capacitance is measured easily and it calibrated against the input quantity, which is displacement. This principle can also be used to measure pressure, velocity, acceleration etc.

15.4 | ACTIVE ELECTRICAL TRANSDUCERS

Thermocouples and piezoelectric crystals are the popular examples of active transducers.

15.4.1 Thermoelectric Transducers

It converts thermal energy into electric energy. Three phenomena which govern the behaviour of a thermocouple are the Seebeck effect, the Peltier effect and the Thomson effect.

✦ Seebeck Effect

If two wires of different metals are joined together to form two junctions and if the two junctions are at different temperatures, an electric current will flow round the circuit. The current flows across the hot junction from the former to the latter metal of the following series:

Bi – Ni – CO – Pd – Pt –

U – Cu – Mn – Ti – Hg – Pb–
Sn – Cr – MO – Ph – Ir – Au – Ag –
Zn – W – Cd – Fe – As – Sb – Te

If metal *A* is of copper and metal *B* is of iron then the current flows from copper to iron at the hot junction and from iron to copper at the cold (reference) junction as shown in Fig. 15.7(a). The Seebeck emf depends on the difference in the temperature of the two junctions.

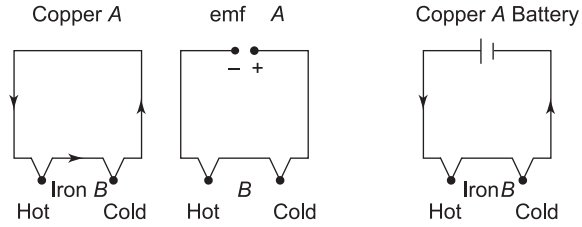


Fig. 15.7 (a) Seebeck effect (b) Peltier effect

✦ Peltier Effect

It is the reverse phenomenon of Seebeck effect. An external emf is connected as shown in Fig.15.7(b) and a current is forced through the junctions. It is observed that heat is absorbed when the flow of current across iron-copper is reversed. The amount of heat liberated or absorbed is proportional to the quantity of electricity that crosses the junction and the amount of heat liberated or absorbed when one ampere passes for a second is called the Peltier coefficient.

✦ Thomson Effect

It is also the reversible heat flow effect and involves the contribution of Seebeck emf in the wire of the same metal, if a temperature difference existed within that particular conductor. When a current flows through a copper conductor having a thermal gradient along its length, heat is liberated at any point where the current is in the same direction as the heat flow, while heat is absorbed at any point where these metals are replaced. In iron, heat is absorbed at any point when the current flows in the direction of heat flow, while heat is liberated, when the current flows in the direction opposite to the flow of heat.

15.4.2 Piezoelectric Transducers

It converts mechanical energy into electrical energy and are based on the direct piezoelectric effect observed in certain nonmetallic and insulating dielectric compounds. Electrical charge is developed on the surface of the crystals, when they are under mechanical strain due to application of stress. Piezoelectric transducers are having high mechanical rigidity so it is used to measure force, pressure, acceleration, torque, strain and amplitudes of vibration. And also it is smaller in size and have high natural frequency, linearity, high sensitivity, wide measuring range and polarity sensitivity. The basic piezoelectric phenomenon is the effect of force applied in longitudinal and transverse directions. The charge sensitivity or piezoelectric *d*-coefficient is the charge developed per unit force. The three modes of operations are thickness expander mode, length expander mode and volume expander mode. These modes are based on the direction of force applied which is to be measured. The materials exhibiting the piezoelectric phenomenon are quartz, Rochelle salt, tourmaline, Ammonium Dihydrogen Phosphate (ADP),

Lithium Sulphate (LS) and DiPotassium Tartrate (DKT). Figure 15.8 shows the piezoelectric crystal wafer used as transducer.

The net piezoelectric effect is represented by the vector of electric polarisation $\overline{P}$ as

$$\overline{P} = \overline{P}_{xx} + \overline{P}_{yy} + \overline{P}_{zz} \tag{15.7}$$

where $\overline{P}_{xx}$, $\overline{P}_{yy}$, $\overline{P}_{zz}$ refer to the effect on the face perpendicular to each axis due to the application of stresses.

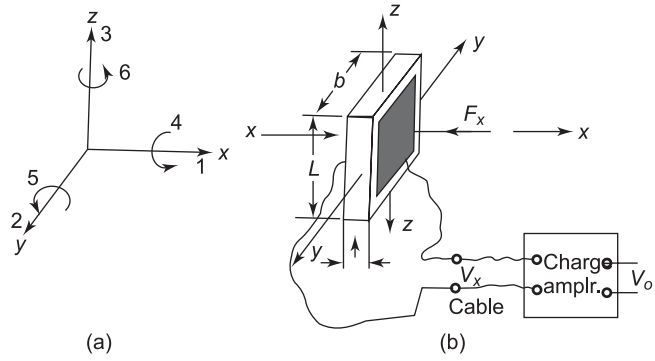


Fig. 15.8 Piezoelectric transducer

15.4.3 Photoelectric Transducers

A photoelectric transducer is based on the effects of physical radiation on matter. If light energy interacts with an electron bound in a metal surface, the entire quantum energy is converted into kinetic energy of the electron. This kinetic energy helps the electrons to move and contributes current in the metal. This is called photoelectric effect, i.e. the effect of visible radiation on the metal (In semiconductor also it is possible). The photo-emissive, photovoltaic, and photoconductive (resistive) are the different forms of photoelectric effects.

Metallic cathode and an anode in an evacuated tube is the main part of photo-emissive transducers. The emitted electrons from cathode are attracted towards the anode, which causes the current flow proportional to the amount of light fallen. The photocurrent depends on the wavelength of the radiation and the material of the surface. The effect of the optical radiation on the semiconductor may be observed as a change in either current, developed voltage or resistance. Photovoltaic cells are self-generating and are favoured for use in exposure meters. Photoresistive cells are passive and the change in the resistance value according to the illumination of light should be measured by suitable circuitry. It is known as Light Dependent Resistor (LDR). Photodiodes and photo transistors are considered to function in both the photo-emissive and photovoltaic modes.

15.4.4 Hall-Effect Transducers

This is one of the galvanomagnetic phenomena in which the interaction between the magnetic field and moving electrical charges results in the development of forces that alter the motion of the charge. This is very much prominent in semiconductor materials. A thin strip of bismuth or *N*-type germanium is subjected to magnetic field *B* normal to its surface as shown in Fig.15.9 while it carries a current *I* along the length of the strip, but normal to *B*. Because of the force from magnetic field the electrons move towards the edges of the strip with a velocity *V*. So the edge surfaces act like charged electrodes and potential difference between two edges is known as Hall potential. E_h is proportional to B and *I*. It is suitable for measurement of magnetic field. The important application is, this transducer has made it possible to measure dc and ac currents in a conductor without interruption of the circuit and without making any electrical contact with the conductor.

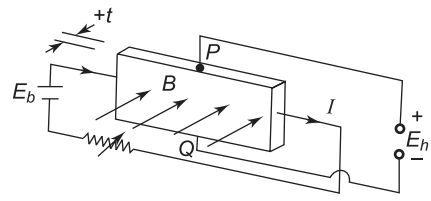


Fig. 15.9 Hall-effect transducer

SUMMARY

- Various types of transducers have been explained in this chapter.
- Their importance, and different types with applications have been discussed.

EXERCISES

➥ Review Questions

1. What is a transducer?
2. Distinguish between active and passive transducers.
3. What is a resistance thermometer?
4. Explain the resistance displacement transducer. Draw the circuit for measuring positive and negative displacement?
5. Explain the working of LVDT.
6. Explain what is Hall effect and its application?
7. Explain the principle of peizoelectric transducer. What are its applications?
8. What is proximity transducer? Can it be used for a vibrating object?
9. Explain Seebeck effect and Peltier effect with diagrams.
10. What is a electric transducer. What are its different forms?
11. Name the transducers for temperature measurement?
12. What are the quality characteristics of a transducer?
13. Distinguish between static and dynamic characteristics of an transducer.
14. Why are the transducers increasingly getting popular?
15. List the problems encountered while measuring small displacement by a capacitive transducer.
16. Explain the transformer type transducers with its working.

17. Discuss the application of piezoelectric effect and materials.
18. Differentiate the photovoltaic, photo-emissive and photoconductive transducers.

➥ Multiple-Choice Questions

1. Which one of the following is an active transducer?
   (a) Strain gauge (b) Thermistor
   (c) Photovoltaic cell (d) Photo-emissive cell
2. A transducer has an output impedance of 1 kΩ and a load resistance of 1 MΩ, the transducer behaves as
   (a) a constant current source (b) a constant voltage source
   (c) a constant power source (d) none of the above
3. In a transducer, the observed output deviates from the correct value by a constant factor, the resulting error is called
   (a) zero error (b) non confirmatory error
   (c) sensitivity error (d) hysteresis error
4. Unbounded strain gauges are
   (a) exclusively used for transducer application
   (b) exclusively used for stress analysis
   (c) commonly used for both transducer application as well as for stress analysis
   (d) none of the above
5. In wire-wound strain gauges, the change resistance on application of strain is mainly due to
   (a) change in length of wire (b) change in diameter of wire
   (c) change in both length and diameter (d) change in resistivity.
6. In an LVDT, the core is made up of a
   (a) non-magnetic material
   (b) a solid ferro-electric material
   (c) high permeability, nickel-iron hydrogen annealed material
   (d) all of the above
7. An LVDT
   (a) exhibits linear characteristics upto a displacement of ±5 mm
   (b) has a linearity of 0.05%
   (c) has an infinite resolution and a high sensitivity which is of the order of 40 V/mm
   (d) all of the above
8. What is the minimum displacement that can be measured with capacitive transducer?
   (a) 1 cm (b) 1 mm (c) 1 μm (d) 1×10^{-12} pm
9. A Hall-effect transducer can be used for measurement of
   (a) power (b) displacement
   (c) current (d) all of the above
10. A photoconductive cell is used for
    (a) high-frequency application (b) low-frequency application
    (c) medium-frequency application (d) all of the above

ANSWERS

✦ Multiple-Choice Questions

1. (c) **2.** (b) **3.** (c) **4.** (a) **5.** (c) **6.** (c) **7.** (d) **8.** (c) **9.** (d) **10.** (b)

CHAPTER 16

Basics of Computers

GOALS AND OBJECTIVES

- ❏ Introduction to basic computers and generation of computers
- ❏ Classification of various types of computers and their hardware and software specifications
- ❏ Discussion of communication packages and development tools
- ❏ Explanation of data communication, computer network and emerging technologies in computing
- ❏ Introduction of next-generation computing paradigm

16.1 | INTRODUCTION

Today computers are influencing every sector and the impact has nowhere been more revolutionary than in electrical engineering. The use of computers has improved our life at work and at home in contrast to our ancestors in the primitive stages of civilisation, who were

using their fingers and pebbles for computing purposes. The need for performing lengthy and complex calculations led to the invention of first variety of calculators, and then finally to computers. In the simplest words, a computer is an electronic machine with capabilities of performing calculations and controlling operations. These operations are performed with the help of instructions (collectively known as a program) to process the given set of data in order to achieve desired results (known as information). Today, computers are an inseparable part of our day-to-day lives—in our workplaces, homes, automobiles, appliances, etc. In fact, the application domain is limited only by human creativity and imagination. The landscape of applications include education, industry, government, medicine, scientific and engineering research, defence, business, commerce, law and even music, arts and sports. The major characteristics that make computers an essential part of our lives include speed, accuracy, reliability, storage capability, versatility and diligence. The Electronic Numeric Integrator and Calculator (ENIAC), developed in 1946 by John Eckert and John Mauchly at the University of Pennsylvania, embodied almost all the components and concepts of today's high-speed, electronic digital computers. In the related development, the Electronic Discrete Variable Automatic Computer (EDVAC) was first developed to use the stored-program concept introduced by John von Neumann. However, the world's first stored-program, electronic digital computer, successfully executed its first program on 21st June 1948. Initially, it was called the "Small Scale Experimental Machine", but was soon nicknamed the "*Baby Computer*". It is also sometimes known as the "Mark 1 prototype". The Baby machine had a 32-bit word length, a memory of 32 words extendable up to 8192 words and a speed of around 1.2 milliseconds per instruction.

The above developments led to the recognition of the computer as an independent discipline. Subsequently, Software Engineering was recognised as an independent discipline in 1968. Both these disciplines remain much younger and are still evolving unlike core engineering. UNIVAC (Universal Automatic Computer) became the first general-purpose and commercially available computer, which was capable of handling numeric and textual information. Since early days Indians have immense contributions to computing disciplines. India commissioned its first digital computer named 'TIFRAC (Tata Institute of Fundamental Research Automatic Calculator) in 1956 just after 4 years commissioning of UNIVAC in the US. Today, India occupies a unique position in computing and information technology with the potential to lead the world in the era of knowledge economy.

16.2 | GENERATION OF COMPUTERS

According to the kind of processor installed in a computer, there are five generations (or stages) of technological development or innovation. The first-generation computers (1940–1956) were built using vacuum tube/thermionic valves. These computers used binary-coded language (also called machine language consisting of 0s and 1s) to perform operations. The notable examples include ENIAC, EDVAC and UNIVAC.

The second generation of computers (1959–1964) was the era of transistor-based computers. The *transistor was invented during 1947* which functions like a vacuum tube in that, it can be used to relay and switch electronic signals. There were obvious differences between the transistor and the vacuum tube. The transistor was faster, more reliable, smaller, and much cheaper to build unlike a vacuum tube. One transistor replaced the equivalent of 40 vacuum tubes. These transistors were made of solid material, some of which is *silicon*, an abundant element. Transistors were found to *conduct electricity faster* and *better* than vacuum tubes. They were also much smaller and gave off virtually *no heat* compared to vacuum tubes. Their use marked a new beginning for the computer. Without this invention, the technological progress in the 1960's would not have been possible.

The third generation of computers (1965–1970) were made up of integrated circuits, which replaced transistors. The *integrated circuit*, or as it is sometimes referred to as *semiconductor chip*, packs a huge number of transistors onto a single *wafer of silicon*. Placing such large numbers of transistors on a single chip vastly increased the power of a single computer and lowered its size and cost considerably. Most electronic devices today use some form of integrated circuits placed on printed *circuit boards* (thin pieces of *Bakelite* or *fibre glass* that have electrical connections etched onto them) sometimes called a *motherboard*. These third-generation computers could carry out instructions in billionth of a second. The size of these machines dropped to the size of small file cabinets.

The fourth generation can be characterised by both the jump to monolithic integrated circuits and the invention of the microprocessor. The monolithic integrated circuits involved putting millions of transistors onto one single chip by which more calculation and faster speeds could be reached by computers. The microprocessor is a single chip that could do all the processing of a full-scale computer. As the size of components and the distances between them became smaller, the speed of computers improved greatly. Initially, the microprocessor was made to be used in calculators, not computers. Subsequently, it led to the invention of *personal computers* (PCs), or *microcomputers*. It wasn't until 1970's that people began buying computers for personal use. One of the earliest personal computers was the *Altair 8800 computer kit*. In 1975, people could purchase this kit and put it together to make their own personal computer. In 1977, the Apple II was sold to the public and in 1981, IBM entered the PC market. Microsoft entered the PC revolution as the co-developer of desktop operating system along with IBM.

The fifth-generation computers are still in the development stage. The computers of the next generation will have millions upon millions of transistors on one chip and will perform over a billion calculations in a single second. The goal of fifth-generation computing is to develop devices that respond to natural language input, and will have the capabilities of learning and self-organising.

16.3 CLASSIFICATION OF COMPUTERS

Due to rapid technological developments and changing information needs, computers are now available in many sizes and types serving diverse purposes. As such, we can classify computers according to the purpose, nature of data handling and functionality (typically

characterised by physical size, performance and application scenarios). This classification is summarised in Fig. 16.1.

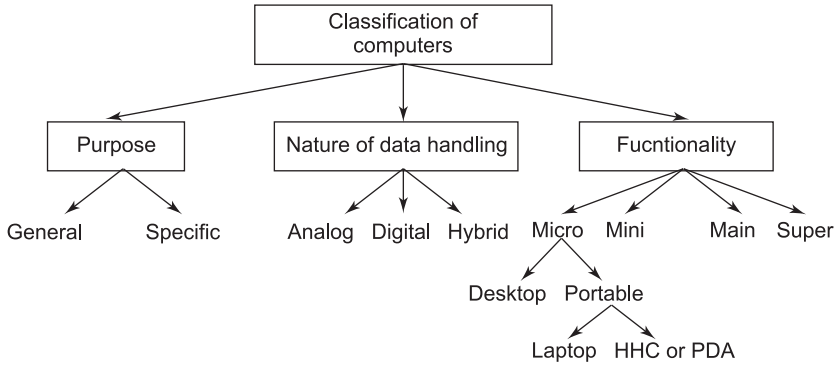


Fig. 16.1 Computer classification

16.4 THE COMPUTER SYSTEM

Typically, a computer (of any generation and/or classification) can be viewed as a system that comprises several units or parts such as computer hardware, peripherals, communication equipment and software. These interrelated units or parts work together for converting data into information. The simplified computer system architecture is depicted in Fig. 16.2.

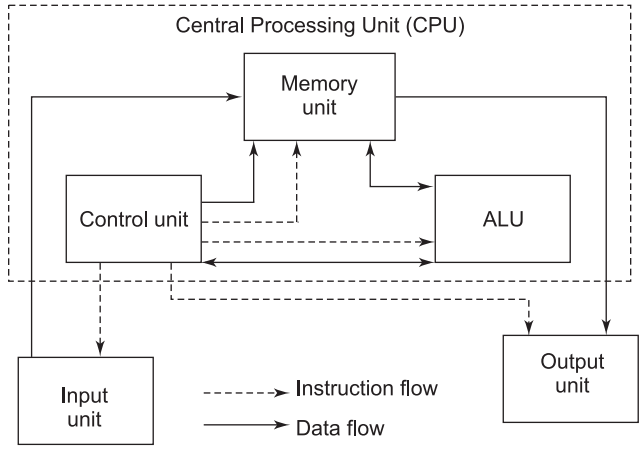


Fig. 16.2 Computer architecture

The CPU or the microprocessor or simply processors is referred as the brain of a computer system. The processors differ from one another by their instruction set (hard wired in the processors) or the instruction set determines the machine language for the process. In general,

the CPU contains the ALU (Arithmetic and Logic Unit) that performs computations, the control unit that can issue control signals, and registers (words of memory inside the CPU). The main job of the CPU is to execute (i.e. interpret) machine-language programs, one instruction at a time. A program is a sequence of machine-language instructions, stored in consecutive memory locations. To execute programs, the CPU uses two special registers: *PC (Program Counter)* contains the address (location) of the instruction being executed at the current time and *IR (Instruction Register)* contains the current instruction being executed. Instructions are executed in a sequence of operations called the *instruction cycle*. One instruction cycle may have multiple machine cycles. One machine cycle consisting of the following operations:

- **Fetch**—Feeding the instructions from memory to ALU
- **Decode**—Converting instruction to control signals that initiate the required operation
- **Execute**—completing the task and generating output

The instruction cycle is repeated indefinitely, as long as the machine is on. Figure 16.3 gives details of the *datapath* within the CPU of a typical von Neumann machine.

In general, one machine cycle corresponds to one cycle of the data path. The control unit (not shown in the data-path diagram), generates signals to direct the operations of the data path such as choosing registers from the register file to be loaded into registers *A* and *B* (or generally working registers), choosing the ALU operation for this cycle (add, subtract, and, or, etc.), choosing the destination register for storing the result, and handling the instruction sequence.

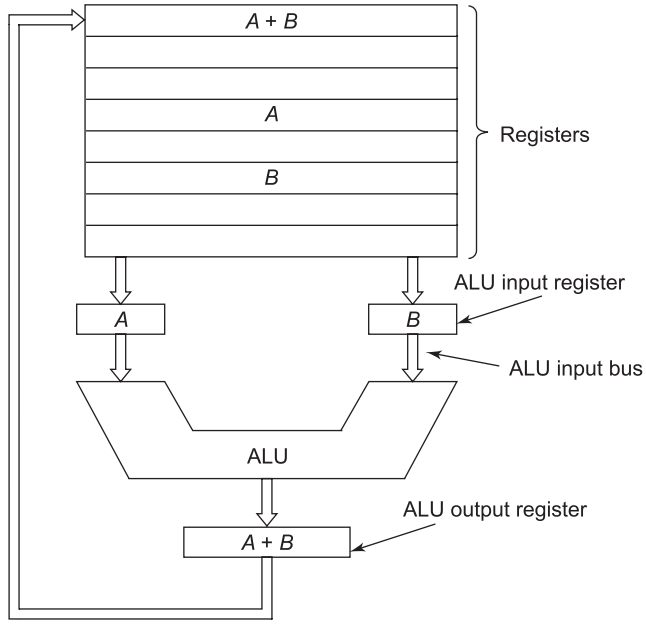


Fig. 16.3 The datapath of a typical von Neumann machine

Further, a specially designed buffer storage called *cache* is used to improve computer performance by reducing access time. It holds instructions and data that are likely to be needed for the next operation by the processor. The cache copies frequently accessed data and instructions from primary storage (main memory) or secondary storage (disks).

❑ ***Disk Cache*** A disk cache is in a reserved segment of primary memory or in an extra memory on the disk controller card. It contains a large block of frequently accessed data copied from a disk. The data in a disk cache can be used to fulfill the data requests from a processor at high-speed. The disk cache lets the processor avoid a slow disk access.

❑ ***Memory Cache*** A memory cache is a high-speed memory storage between the memory and the CPU. It is smaller and much faster than the main memory (primary storage). The memory cache copies blocks of instructions and data from the main memory so that execution and data updating are performed in the higher-speed memory bank.

16.5 | COMPUTER HARDWARE

The major constituents of computer hardware are the processor, memory and input/output devices. As mentioned previously, the size and cost of hardware has been falling due to rapid technological advancements. Further, many hardware components may be integrated on a single board, known as the motherboard. Today's motherboards may have a built-in hard-disk controller, graphics, multi-media and network interface cards. The following sections provide brief descriptions of major hardware constituents.

16.5.1 Processors

A processor is a microprocessor whose elements are miniaturised into one or a few integrated circuits contained in a single silicon microchip, which can execute instructions. Intel and Motorola are the major companies that produce important microprocessors. In a microcomputer, the Central Processing Unit (CPU) is held on a single microprocessor and requires a system clock, primary storage, and power supply to perform its processing functions. The capacity of a microprocessor chip is represented in word sizes. A word size is the number of bits (e.g. 8, 16 or 32 bits) that a computer (CPU) can process at a time. If a word has more bits, the processing is faster and more powerful. For example, a 16-bit-word CPU can access and manipulate 2 bytes (1 byte = 8 bits) at a time, while a 32-bit-word CPU can access and manipulate 4 bytes at a time. Therefore, the 32-bit CPU is faster than the 16-bit CPU. Based upon the instruction sets, there are two types of computing architectures, CISC and RISC.

❑ ***CISC Chips*** CISC stands for *Complex Instruction Set Computer*. It is pronounced as "sisk." CISC is a computer architecture that has large sets of instructions in the order of several hundred instructions. Intel's Pentium chip uses CISC design.

❑ ***RISC Chip*** RISC stands for *Reduced Instruction Set Computer*. In RISC, most programs generally use only a few instructions. This architecture has a small number of instructions built into the circuits and if those basic instructions are made to execute faster than RISC

computers increase performance. Although RISC machines are only around 30% faster than CISC machines, RISC chips are less expensive to produce.

16.5.2 Primary Memory (Internal Storage)

Primary memory (*internal storage*, *main memory* or *memory*) is the computer's working storage space that holds data, instructions for processing, and processed data (information) waiting to be sent to secondary storage. Physically, primary storage is a collection of RAM chips. The contents are held in the primary storage only *temporarily*. Its *capacity* varies with different computers. Data or instructions are stored in primary storage locations called *addresses*.

16.5.3 Secondary Storage (Hard Disk Drives—HDD)

As the primary memory is small and volatile in nature, there is need to use secondary storage. The secondary storage is typically of large and permanent capacity. The most commonly used and available secondary storage is the hard disk drive or HDD. A *hard disk* is a magnetic disk made of metal and covered with a magnetic recording surface. It is a nonvolatile storage device which stores digitally encoded data on rapidly rotating platters with magnetic surfaces. Hard disks come in removable and fixed varieties that hold from several hundreds of megabytes to several gigabytes. Older HDDs quoted their smaller capacities in megabytes, some of the first drives for PCs being just 5 or 10 MB, but recent HDDs are of few gigabytes to few hundred gigabytes. The capacity of an HDD can be calculated by multiplying the number of cylinders by the number of heads by the number of sectors by the number of bytes/sector (most commonly 512 bytes/sector).

16.5.4 Tertiary or Auxiliary Storage

Tertiary, or auxiliary storage, is typically an external storage meant for data transfer and back-up purposes. This storage includes floppy diskettes, magnetic tape, optical disk, CD ROMs and USB drives. Floppy diskettes are mostly used for boot-up programs and rarely as auxiliary storage. Currently, CD-ROMs and USB Drives are predominantly used as auxiliary storage.

CD-ROM (an abbreviation of "Compact Disc read-only memory") is a pre-pressed Compact Disc that contains data accessible but not writable by a computer. The Compact Disc format was originally designed for music storage and playback. CD-ROMs are popularly used to distribute computer software, including games and multimedia applications, though any data can be stored (up to the capacity limit of a disc). Some CDs hold both computer data and audio with the latter capable of being played on a CD player, whilst data (such as software or digital video) is only usable on a computer (such as PC CD-ROMs). Discs are made from a 1.2 mm thick disc of polycarbonate plastic, with a thin layer of aluminum to make a reflective surface. The most common size of a CD-ROM disc is 120 mm in diameter, though the smaller Mini CD standard with an 80 mm diameter, as well as numerous nonstandard sizes and shapes (e.g. business-card-sized media) are also available. Data are stored on the disc as a series of microscopic indentations. A laser is shown onto the reflective surface of the disc to read the pattern of pits and lands ("pits", with the gaps between them referred to as "lands"). Because the depth of the pits is approximately one-quarter to one-sixth of the wavelength of the laser

light used to read the disc, the reflected beam's phase is shifted in relation to the incoming beam, causing destructive interference and reducing the reflected beam's intensity. This pattern of changing intensity of the reflected beam is converted into binary data.

A **USB flash drive** is a NAND-type, flash memory, data storage device integrated with a USB (Universal Serial Bus) interface. USB flash drives are typically removable and rewritable, much shorter than a floppy disk and lightweight. Storage capacities typically range from 64 MB to 64 GB with steady improvements in size and price per gigabyte. Some allow 1 million write or erase cycles and have 10-year data retention, connected by USB 1.1 or USB 2.0. USB Memory card readers are also available, whereby rather than being built-in, the memory is a removable flash memory card housed in what is otherwise a regular USB flash drive, as described below.

USB flash drives offer potential advantages over other portable storage devices, particularly the floppy disk. They are more compact, faster, hold much more data, have a more durable design, and are more reliable for lack of moving parts. Additionally, it has become increasingly common for computers to ship without floppy disk drives. USB ports, on the other hand, appear on almost every current mainstream PC and laptop. These types of drives use the USB mass storage standard, supported natively by modern operating systems such as Windows, Mac OS X, Linux, and other Unix-like systems. USB Memory card readers are also available, rather than being built-in, the memory is a removable flash memory card housed otherwise a regular USB flash drive, as described below.

16.5.5 Input Devices

The keyboard is a main input device used to feed data and instructions to the processor. There are many specialised input devices, viz. mouse, touch screen, light pen, graphics tablet, scanner, bar code readers and smart cards. Storage devices such as disk drives and even floppy diskettes can also serve as input devices. Voice-based input devices are the latest developments providing next-generation input technology for human-computer interaction. The human computer interaction will become more sophisticated and complex in the era of fast approaching, disappearing, computing scenarios or smart environments. Computer-to-computer interaction will also assume great importance in the networked world.

16.5.6 Output Devices

The Visual Display Unit (VDU), also commonly called monitor, is the mainly used output device. The other output devices include printers and plotters. Storage devices such as disk drives and even floppy diskettes can also serve as output devices.

A monitor or display is a piece of electrical equipment, which displays images generated from the video output of devices, such as computers, without producing a permanent record. The monitor comprises the display device, simple circuitry to generate and format a picture from video, sent by the signals source, and usually an enclosure. Within the signal source, either as an integral section or a modular component, there is a display adapter to generate video in a format compatible with the monitor. There are two types of monitors: Cathode-Ray Tube (CRT) monitor and Liquid Crystal Display (LCD) monitor. The CRT monitors are big and heavy looking like a television, but have a better display resolution and often a higher

frequency. The liquid crystal display (LCD) monitors, are thin, flat, and lightweight. It is a newer technology than CRTs. The quality can be the same or even better than a CRT, but these types of monitors usually cost more than CRT monitors. A newer LCD monitor typically consists of a Thin Film Transistor (TFT).

A printer is a peripheral device, which produces a hard copy (permanent human-readable text and/or graphics) of documents stored in electronic form, usually on physical print media such as paper or transparencies. Printers are mainly classified into line, daisy-wheel, dot matrix, laser and inkjet printers. The data received by a printer may be a string of characters, a bitmapped image, or a vector image. Some printers can process all three types of data, while some cannot. The speed of early printers was typically measured in units of characters per second. More modern printers are measured in pages per minute.

16.6 | COMPUTER SOFTWARE

A computer system is a useless box unless otherwise it is provided with appropriate data and programming instructions. Programming has been a fascinating world since the advent of the computer. Sometimes, programming may be equated with computing itself. However, we need to take a holistic view as there have been several developments since 1948, the year the *Baby Computer* was introduced. Today, we have an independent software engineering discipline which deals with development and management of large and complex programmes working as a unified system. Similarly, we have to deal with the information beyond data, which dominated the early computing domain. Another branch of study, *informatics*, deals with the science of information, the practice of information processing and the engineering of information systems. India has been a pioneer in these fields since the last two decades. Due to dominance of software and services, IBM in association with academia and industry have proposed a new discipline—SSME. The Service Science, Management and Engineering (SSME) is a growing multi-disciplinary effort that integrates aspects of computing, operations research, engineering, management science, business strategy, social and cognitive science and legal science disciplines. However, we limit our discussion here in the following sections dealing with basic software and its applications.

16.6.1 System Software

System software is a program that manages computer resources and operations of a computer system while it executes various tasks such as processing data and information, controlling hardware components and allowing users to use application software. That is, systems software functions as a *bridge* between computer system hardware and the application software. System software is made up of many language translators, control programs, including the operating system, communications software and database management system.

16.6.2 Operating Systems

An operating system is a collection of integrated computer programs that provide recurring services to other programs or to the user of a computer. These services consist of disk and file

management, memory management and device management. In other words, it manages CPU operations, input/output activities, storage resources, diverse support services, and controls various devices. An operating system executes many functions to operate a computer system efficiently. Among them, four essential functions are the following:

❑ ***Resource Management*** An operating system manages a collection of computer hardware resources by using a variety of programs. It manages computer-system resources, including its CPU, primary memory, virtual memory, secondary storage devices, input/output peripherals, and other devices.

❑ ***Task Management*** This function of the operating system controls the running of many tasks. It manages one program or many programs within a computer system simultaneously. That is, this function of operating system manages the completion of users tasks. A task-management program in an operating system provides each task and interrupts the CPU operations to manage tasks efficiently. Task management may involve a multitasking capability.

❑ ***File Management*** This is a function that manages data files. An operating system contains file management programs that provide the ability to create, delete, enter, change, ask and access files of data. They also produce reports on a file.

❑ ***User Interface*** It is a function of an operating system that allows users to interact with a computer. A user-interface program may include a combination of menus, screen design, keyboard commands, etc. A well-designed user interface is essential for an operating system to be popular. Because of the function, users can load programs, access files and accomplish other tasks.

❑ ***Virtual Memory*** This is a technique for an operating system to manage memory. An operating system simulates significantly larger memory capability than the real memory capacity of its actual primary storage unit. It allows computers to process larger programs than the physical memory circuit would allow.

❑ ***Multitasking*** This refers to the capability of operating systems that run several computing tasks in one computer at the same time. This is controlled by the task-management program in an operating system. It is also called *multiprogramming* and *multithreading*.

The operating system along with language translators typically constitutes a basic platform for computer programming and software development. The language translators can be categorised as assembler, compiler, and interpreter. An assembler translates the assembly-language programme into machine-level-language program. A compiler is used for translating an entire piece of source code written in a high-level language to machine-level language whereas an interpreter translates statement by statement. A modern assembler creates an object code by translating assembly instruction mnemonics into opcodes, and by resolving symbolic names for memory locations and other entities. A cross compiler is a compiler capable of creating executable code for a platform other than the one on which the compiler is run.

16.6.3 Programming Languages

A programming language is used to write programs involving a computer to perform computations. The complex computations typically involve algorithms. A computer program may possibly control external devices such as printers, robots, and so on. It may contain constructs for defining and manipulating data structures or controlling the flow of execution, typically following a prescribed syntax. Most programming languages are purely textual; they use sequences of text including words, numbers and punctuation, much like written natural languages. On the other hand, there are some programming languages, which are more graphical in nature, using spatial relationships between symbols to specify a program. The syntax of a language describes the possible combinations of symbols that form a syntactically correct program. Programming language syntax is usually defined using a combination of regular expressions. There are five generations (levels) of programming languages. The first three generations are procedural and the next two are nonprocedural languages. A *procedure* is effectively a list of computations to be carried out for completing the desired task. So the procedural languages are simply integration of procedures (also known as functions, subroutines, or methods), small sections of code that perform a particular function to be carried out to reach the final state. In nonprocedural languages (fourth-and fifth-generation languages), the user only has to tell what to do, the system will take care of steps to be followed and the logics.

The first-generation language is machine language (involves only 0's and 1's) which directly controls the hardware resources. The second-generation language is assembly programming language in the form of mnemonics, which is an advancement over the machine language. The third-generation languages include high-level languages like BASIC, COBOL, FORTRAN, Pascal and C. The example of fourth-generation language is Structured Query Languages (SQL), which has become a native language for database management systems. SQL is basically a command-based language without low-level interaction with computer hardware. Many computer scientists have advocated in favour of PROLOG (Programming in Logics) as the fifth-generation language though the fifth generation hardware is still being debated. The fifth-generation languages will enable development of intelligent systems capable of making decisions with or without users' involvement.

The other classification of programming languages is structural and object-oriented, programming languages. *Structured programming* is a special type of procedural programming. It requires that programmers break the program structure into small pieces of code that are easily understood. It is often associated with a "top-down" approach to design. The most popular structured programming languages include C, Ada and Pascal. In Object-Oriented Programming (OOP), the designer specifies both the data structures and the types of operations that can be applied to those data structures. This pairing of a piece of data with the operations that can be performed on it is known as an *object*. A program thus becomes a collection of cooperating objects, rather than a list of instructions. Objects can store state information and interact with other objects, but generally each object has a distinct, limited role. Some of the key characteristics of OOP are inheritance, encapsulation and polymorphism. The most popular object-oriented programming languages include C++, Java, C# and Python.

16.6.4 Application Software

Application software consists of programs that direct computers to perform specific information-processing activities for end users. These programs are called *application packages* because they direct the processing required for a particular use, or *application*, which users want to accomplish. Thousands of application packages are available because there are thousands of different jobs end users want computers to do. Application software includes a variety of programs that can be subdivided into general-purpose and application-specific categories.

✦ General-Purpose Application Programs

General-purpose application packages are programs that perform common information processing jobs for end users. For example, word-processing programs, electronic spreadsheet programs, database management programs, graphics programs, communications programs and integrated packages are popular with microcomputer users for home, education, business, scientific and many other general purposes. They are also known as *productivity* packages, because they significantly increase the productivity of end users. This packaged software is also called *commercially off-the-shelf software* (*COTS*) *package*, because these products are packaged and available for sale. Many features are common to most packaged programs. Some examples are word-processing packages, spreadsheets, etc. which are used to create, manipulate, print the documents and they can even use some formulas and calculate different outcomes. Microsoft office tools and Linux open sources are the most common application programs.

✦ Application-Specific Software

Many application programs are available to support specific applications of end users. These can be classified as *Scientific and Business Application Programs*.

❑ ***Scientific Application Programs*** Programs that perform information processing tasks for the natural, physical, social and behavioral sciences, engineering and all other areas involved in scientific research, experimentation and development. There are so many other application areas such as education, music, art, medicine, etc. The trend in computer application software is towards multipurpose, expert-assisted packages with natural language and graphical user interfaces.

❑ ***Business Application Programs*** These are programs that accomplish the information processing tasks of important business functions or industry requirements.

✦ Graphics Packages

A graphics program can typically display numeric data in a visual format for analytical or presentation purposes. Any other types of presentation graphics displays are also possible. Most of the graphics packages support freehand drawing, while desktop publishing programs provide predrawn clip art graphics for insertion into documents. Popular business graphics packages are Harvard Graphics, Freelance, Corel Draw, etc. There are two types of graphics

programs. *Analytical graphics programs*, which are used to analyse data and *presentation graphics programs*, which are used to create attractive finished graphs for presentations or reports.

✦ Database Management Packages

A *database* is a large collection of data entered into a computer system and stored for future use. The computerised information in the database is organised so that the parts that have something in common can be retrieved easily. A database management package or *Database Management System* (*DBMS*) is a software package used to set up, or structure, a database. The four types of DBMS are hierarchical, network, relational DBMS, and object-oriented DBMS. The most popular one is Relational DBMS. The structured query language (SQL) serves as a native language for most DBMS. It is also used to retrieve and manage information from a database. Most DBMS packages can perform four primary tasks:

❑ ***Database Development*** Define and organise the content, relationships, and structure of the data needed to build a database

❑ ***Database Interrogation*** Access the data in a database for information retrieval and report generation. A user can selectively retrieve and display information and produce printed reports and documents

❑ ***Database Maintenance*** Add, delete, update, correct, and protect the data in a database

❑ ***Application Development*** Develop prototypes of data entry screens, queries, forms, reports, and labels for a proposed application

✦ Communications Packages

Communications software packages for microcomputers are also viewed as general-purpose application packages. These packages can connect a microcomputer equipped with a modem to a public and private network. Communications software enables a microcomputer to send and receive data over a telephone or other communications line. Communications programs are used by all kinds of people inside and outside business. Examples are students doing research papers, travellers making plane reservations, consumers buying products, investors getting stock quotations and economists getting government statistical data. Communications programs give microcomputers a powerful feature, which is connectivity. Connections with microcomputers open a world of services. Before the advent of the Internet, the popular communications software used were ProComm, SmartCom and Crosstalk. Some common usage of computer communications programs are as follows:

❑ ***Data Banks*** With this communications program, users can access enormous computerised databases—data banks of information. Some of these, such as Dialog, resemble huge electronic encyclopedias.

❑ ***Message Exchanges*** Communications programs enable users to leave and receive messages on electronic bulletin boards or to use electronic mail services. Electronic bulletin boards exist for people interested in swapping all kinds of software or information. Many organisations now have electronic mailboxes.

❑ ***Financial Services*** With communications programs, users can look up airline reservations and stock quotations. Users can order discount merchandise and even do home banking and bill paying.

✦ Development Tools and Special Purpose Software

Software is omnipresent in most of the scientific, engineering, industry and other fields. As such, several domain-specific or special-purpose software tools have been developed to increase productivity, quality and functionality. Further, we have the Computer-Aided Software Engineering (CASE) Tool—an interesting set of tools, which even automates the development of software itself. A typical CASE tool helps in requirements engineering, modelling, architecture, design, verification and validation of software systems. The following sections discuss tools used in other branches of science and engineering.

❑ ***MATLAB*** is a high-performance language for technical computing. It integrates computation, visualisation, and programming in an easy-to-use environment, where problems and solutions are expressed in familiar mathematical notation. Typical uses are computation, algorithm development, modelling, simulation, prototyping, data analysis, exploration, visualisation, scientific and engineering graphics, application development, including Graphical User Interface building, etc.

❑ ***Mathematica*** is a versatile, powerful application package for doing mathematics and publishing mathematical results. It runs on most popular workstation operating systems, including Microsoft Windows, Apple Macintosh OS, Linux and other Unix-based systems. Mathematica is used by scientists and engineers in disciplines ranging from astronomy to zoology; typical applications include computational number theory, ecosystem modelling, financial derivatives pricing, quantum computation, statistical analysis and hundreds more.

❑ ***Electronic Design Automation (EDA)*** is the category of tools for designing and producing electronic systems ranging from Printed Circuit Boards (PCBs) to integrated circuits. EDA for electronics has rapidly increased in importance with the continuous scaling of semiconductor technology. Some users are foundry operators, who operate the semiconductor fabrication facilities, (or "fabs") and design-service companies who use EDA software to evaluate an incoming design for manufacturing readiness. EDA tools are also used for programming design functionality into FPGAs. Cadence provides a wide range of EDA tools.

❑ ***AutoCAD*** is a Computer Aided Design (CAD) program used by just about every engineering and design office in the world. Although there are alternative CAD packages, AutoCAD is by far the most widely used system. Autodesk's AutoCAD is the industry leader in CAD packages used by civil engineers, architects, mechanical and electrical engineers, aeronautical engineers plus many other disciplines. The drawings with accurate scale can be created and published using AutoCAD's powerful features. 3D models can also be created giving the designer absolute control over the design from start to finish. The computerised model can be viewed through a 360° angle, and even 'rendered' with a texture on screen to give an idea of the finished product.

16.7 DATA COMMUNICATION AND COMPUTER NETWORK

The term *communication* in simple words means exchanging (i.e. receiving or sending) information. In ancient times, the usage of beating drums, smoke signals, mirrors reflecting sunlight, and so on were the common means of communication. With the advancement in science and technology, various communication devices were developed. Voice communication became common after the invention of the telephone by Alexander Graham Bell in 1870. The major breakthroughs in 1980 led to setting up of data communication and networked computers, which revolutionised the field of information and communication technology. The advent of the Internet in 1960s and subsequent technological developments have transformed the world into a global village. The following discussion summarises some of the basic concepts of data communication.

A data-communication system consists of different components that are used to work together to transfer data from one place to another. These are central computer, control programs, terminals, modem, data communication media, data transmission modes and networking. The computers that manage transmission of data between different computers is called a *central computer*. Control programs control the flow of data from one computer to the other. These are operating systems that control the working of the computers. A terminal is the basic communication unit. Many persons can work on a single computer by connecting many terminals to the large computer. Through telephone-line data transfer from one place to other in the form of analog signals. Thus, for transferring data from one computer to another computer through telephone line, digital signals must be converted to analog signals. For receiving digital signals by another computer, these analog signal must be converted into digital signals. Different data-transmission modes are used such as twisted cable, fibre optical cable, microwave, etc.

There are two basic categories of transmission media: guided and unguided. *Guided transmission media* uses a "cabling" system that guides the data signals along a specific path. The data signals are bound by the "cabling" system. Guided media is also known as *bound media*. Cabling is meant in a generic sense in the previous sentences and is not meant to be interpreted as copper wire cabling only. *Unguided transmission media* consists of a means for the data signals to travel but nothing to guide them along a specific path. The data signals are not bound to a cabling media and as such are often called *unbound media*. There are 4 basic types of guided media: open wire, twisted pair, coaxial cable, and optical fibre. Unguided Transmission Media is data signals in the form of Radio Frequencies (RF) that flow through the air. They are not guided or bound to a channel to follow. The RF wave propagations are classified by ground wave, ionospheric, and Line of Sight (LOS) Propagation.

Based on the permissible direction of the data flow, the communications systems can be categorised as simplex, half-duplex and full-duplex. In *simplex* systems, the data flows in only one direction on the data communication line (medium). Examples are radio and television broadcasts. In *half-duplex* systems, the data flows in both directions but only one direction at a time on the data communication line, for example, conversation on walkie-talkies. In *full-duplex* systems, the data flows in both directions simultaneously. A common example is

Modulator-Demodulator (Modem) based communication, which can be configured for data flow in both directions.

Further, the data transmission may be synchronous or asynchronous. *Asynchronous transmission* sends only 1 character at a time; a character being a letter of the alphabet or number or control character. Preceding each character is a start bit and ending each character is 1 or more stop bits. *Synchronous transmission* sends packets of characters at a time. Each packet is preceded by a *start frame*, which is used to tell the receiving station that a new packet of characters is arriving and to synchronise the receiving station's internal clock. The packets also have end frames to indicate the end of the packet.

16.7.1 Computer Communication Protocols

The International Standards Organization (ISO) developed a set of standard protocols called the Open Systems Interconnection (OSI). The OSI model separates each network's functions into seven layers of protocols or communication rules. This model identifies functions that should be offered by any network system. The OSI Model clearly defines the interfaces between each layer. This allows different network operating systems and protocols to work together by having each manufacturer adhere to the standard interfaces.

❑ ***Physical Layer*** The physical layer sends data from the user computer to a host computer (and vice versa). This layer is concerned essentially with computer hardware, whereas the upper layers are interested in communications software. This layer, for example, manages voltage of electricity, timing factors or connector standards.

❑ ***Data-Link Layer*** The data-link layer formats the received data into a record called a frame. This layer also is in charge of error detection.

❑ ***Network Layer*** The network layer provides the physical layer with the ability to transfer records from one computer to another. This layer provides for the functions of internal network operations such as addressing and routing.

❑ ***Transport Layer*** The transport layer allows communication to take place between the user and host computers. That is, this layer takes care of end to end validity and integrity of the transmission. OSI transport services takes place through layer 1 to layer 4, which are collectively responsible for acquiring data and information from the sender to the receiver.

❑ ***Session Layer*** The session layer is in charge of starting, maintaining, and ending each logical session between the interacting end-user computers.

❑ ***Presentation Layer*** The presentation layer formats incoming data so that it can be presented on the receiving terminal for end users. In other words, it is in charge of displaying, formatting and editing user inputs and outputs.

❑ ***Application Layer*** The application layer controls the input from the user computer and allows the application program to be run on the host computer. This layer is the end user's access to the network.

US Department of Defence project ARPAnet officially adopted the Transmission Control Protocol/ Internet Protocol (*TCP/IP*) *developed in the 1970s by pioneering network engineers Vinton Cerfand Bob Kahn*. TCP/IP has two parts. TCP protocol controls data transfer that is the function of the transport layer in the OSI model. IP protocol provides the routing and addressing mechanism that are the roles of the network layer in the OSI model. This protocol is supported by many hardware vendors from microcomputers to mainframes. It is used by most universities, federal governments, and many corporations. TCP/IP is also the network protocol used on the Internet thus making it the *de facto* standard while the OSI remained as the reference model.

16.7.2 Types of Networks

Communications networks vary in geographical size. Networks may be constructed within a building or across several buildings. Networks may also be citywide and even international, using both cable and air connections. There are three major network types: Local Area Network (LAN), Metropolitan Area Networks (MAN) and Wide Area Network (WAN).

A computer communications network contained in a small area such as a commercial building is known as a local area network (LAN). A LAN usually is accomplished with either telephone, coaxial, or fibre optic cables. The major benefit of a LAN is that it can help reduce costs by allowing people and microcomputers to share expensive resources. Also, a LAN enables users to participate in office automation systems. Network designers can choose the method in which messages are controlled in a LAN. Two basic methodologies are the token-passing approach and the contention-based approach. The *token-passing approach* allows the designers of a network to achieve a degree of centralised control. A group of data bits, or "token", is passed from one network node to another. A node can only send a message when it is in possession of the token. In the *contention-based approach*, a node that wishes to send a message first listens to determine if another node is currently sending a message. If not, the node attempts to send its message. However, the lack of centralised control can result in a collision—two nodes attempting to send messages simultaneously. Consequently, contention-based approach is usually not suitable for networks with a large amount of communications activity. The Ethernet is the most popular example using the contention-based approach.

The next larger network than LAN may be a Metropolitan Area Network (MAN) usually spanning a city or county area. It interconnects various buildings or other facilities within this citywide area. For example, linkages can be established between two commercial buildings. A more recent use of MAN technology has been the rapid development of cellular phone systems.

A wide area network (WAN) is one that operates over vast distances spanning cities, states or national boundaries. This network interconnects computers, LANs, MANs and other data-transmission facilities. Typically, WAN will employ communications circuits such as long-distance telephone wires, microwaves and satellites.

16.7.3 Network Topology

Networks can be classified by their topology. The topology is the basic geometric arrangement of the network. Communications channels can be connected in different arrangements using several different topologies.

Four basic types of network configurations are star, bus, ring and mesh. The LANs typically use bus or star and sometime ring topologies. Star and mesh topologies are commonly used in MANs and WANs. The networks are usually built using a combination of several different topologies.

A *star topology* is one in which a central unit provides a link through which a group of smaller computers and devices are connected. The central computer is commonly called a *host computer*. In the star network, all interactions between different computers in the network travel through the host computer. The central unit will poll each to decide whether a unit has a message to send. If so, the central computer will carry the message to the receiving computer.

In a *bus configuration*, each computer in the network is responsible for carrying out its own communications without the aid of a central unit. A common communications cable (the bus) connects all of the computers in the network. As data travels along the path of the cable, each unit performs a query to determine if it is the intended recipient of the message. As the information passes along the bus, it is examined by each terminal to see if the data is for it.

A *ring configuration* features a network in which each computer is connected to the next two other computers in a closed loop. Like the bus network, no single central computer exists in the ring configuration. Messages are simply transferred from one computer to the next until they arrive at their intended destinations. Each computer on the ring topology has a particular address. As the messages pass around the ring, the computers validate the address. If the message is not addressed to it, the node transmits the message to the next computer on the ring.

The *mesh* is a netlike configuration in which there are at least two pathways to each node. In a *mesh topology*, computers are connected to each other by point-to-point circuits. In the topology, one or more computers usually become switching centres, interlinking computers with others. Although a computer or cable is lost, if there are other possible routes through the network, the damage of one or several cables or computers may not have vital impact except the involved computers. However, if there are only few cables in the network, the loss of even one cable or device may damage the network seriously.

16.7.4 Types of Internetworks

An internetwork is composed of two or more networks or network segments connected. Any interconnection among or between public, private, commercial, industrial or governmental networks may also be defined as an internetwork or simply an internet. In modern practice, the interconnected networks use the Internet Protocol (IP). There are at least three variants of internetwork, depending on who administers and who participates in them: Intranet, Extranet and the Internet. An *intranet* is an IP network within an organisation or a corporate. An intranet may or may not have connections to the Internet. An intranet provides IP-based tools such as Web browsers and file transfer applications that are under the control of a single administrative entity. That administrative entity closes the intranet to all but specific, authorised users. A large intranet will typically have at least one Web server to provide users with organisational information. An intranet is called an extranet when it provides access to

itself from an authorised external organisation or an employee on the move (e.g. marketing or field engineer).

An *extranet* is a network or internetwork that is limited in scope to a single organisation or entity. It may also have limited connections to the networks of one or more usually, but not necessarily, other trusted organisations or entities (e.g. a company's customers may be given access to some part of its intranet creating in this way an extranet, while at the same time the customers may not be considered 'trusted' from a security standpoint). Technically, an "extranet may also be categorised as a CAN, MAN, WAN, or other type of network, although, by definition, an extranet cannot consist of a single LAN; it must have at least one connection with an external network.

The *Internet* is a specific internetwork. It consists of a worldwide interconnection of governmental, academic, public, and private networks using the Internet Protocol Suite. The Internet has become a backbone underlying the World Wide Web (WWW). Participants in the Internet use an addressing system (IP Addresses) administered by the Internet Assigned Numbers Authority (IANA). The Internet Engineering Task Force (IETF) coordinates protocols development. Service providers and large enterprise exchange information about the reachability of their address spaces through the Border Gateway Protocol (BGP), forming a redundant worldwide mesh of transmission paths.

16.7.5 Application of the Internet

The major applications of the internet are e-mail, remote login, discussion groups and information resources.

✦ E-mail on the Internet

Electronic mail (e-mail) is one of the most rapidly growing developments in networked communications. Users of e-mail have their own file stored on a computer system. This file can be called a "mailbox." Access to a person's mailbox is protected by means of a password. Once logged on to an e-mail account, an end-user may send messages and files to other mailboxes. An individual electronic mail transmission may be sent to one or many recipient accounts. This person may also read messages that have "arrived" in her or his mailbox from other e-mail accounts. Popular uses of e-mail have been to set up meetings within business organisations and to distribute memoranda throughout an organisation. Anyone with access to the Internet can send e-mail to anyone else on the Internet. An Internet e-mail address has two parts, the individual user's account address and the address of the computer. The computer's address, in turn has two parts, the computer name, and its domain. The general format is therefore, user@computer.domain. Note that the "at" symbol (@) separates the user's account from the computer address, and a period (.) separates the name of the computer from its domain.

✦ Telnet (Remote Login)

The service that allows users to connect to a remote Internet host is called *Telnet*. Users on one computer in the Internet can login into other computers on the Internet by a special program, called Telnet, on your computer. This program uses the Internet to connect to the computer

users specify. The users should know the account name and password of the remote computer. In Telnet, a user's computer is called the *local computer*. The other computer that the Telnet program connects is called the *remote computer*. An example of using Telnet is that users can read and send e-mail while travelling.

✦ Discussion Groups

Discussion groups are the Internet users who have joined together to discuss some topic. Two groups are commonly used for business include Usenet and Listserv. *Usenet* is a large collection of discussion groups involving millions of people from all over the world. This is the most formally organised among the discussion groups. To read Usenet articles, users use a program called a newsreader. A *Listserv* is simply a mailing list developed on the large Bitnet network (not on the Internet). The listserv processor processes listserv commands such as requests to subscribe and unsubscribe, while the listserv mailer mails any message it receives to everyone on the mailing list.

✦ Information Resources

The major use of the Internet is to find information. There are six major ways to find and achieve information:

❑ ***FTP*** File Transfer Protocol (FTP) is the underlying set of specifications that support Internet file transfer. In other words, FTP is a service that allows us to copy a file from any Internet host to any other Internet host.

❑ ***Archie*** Throughout the Internet, there are a number of computers, called *Archie servers*, which provide a service to help users find the name of anonymous FTP hosts that carry a particular file. Archie is a tool that allows users to search most of the publicly available anonymous FTP sites worldwide for specific files of interest.

❑ ***Gopher*** The Gopher is a powerful system that allows users to access many resources of the Internet in a simple, consistent manner. To use the Gopher, all users need to do are making selections from a menu.

❑ ***Veronica*** Veronica is a Gopher-based resource that users can use to search gopher space for all the menu items that contain specified words. Veronica is to Gopher what Archie is to FTP. It enables users to search all publicly available Gopher sites by specifying key-words.

❑ ***World Wide Web (WWW)*** WWW or simply Web is one type of information resource that is growing even faster than Internet itself. It is an attempt to organise all the content on the Internet as a set of hypermedia documents. Besides that, it allows users to access all kinds of online resources using a Web browser and a Web server. A browser is a software package for accessing a Web server that stores files using HTML. There are many Web browsers available, for example, Microsoft Internet Explorer (IE), Mozilla FireFox, etc. A Web server stores information in a series of text files called *pages*. These text files or pages use a structured language called HTML (Hypertext Markup Language) to store their information.

❑ ***WAIS*** The WAIS stands for Wide Area Information Service. The original idea behind WAIS was to develop a generalised system of information retrieval that could access collections of data all around the world.

16.7.6 Basic Network Components

All networks are made up of basic hardware building blocks to interconnect network nodes, such as Network Interface Cards (NICs), repeaters, bridges, hubs, switches, routers and gateways. The repeaters and bridges are now not in use. A NIC is a piece of computer hardware designed to allow computers to communicate over a computer network. It provides physical access to a networking medium and often provides a low-level addressing system through the use of Media Access Control (MAC) addresses.

A *hub* is a multiple-port device. When a packet arrives at one port, it is copied to all the ports of the hub for transmission. When the packets are copied, the destination address in the frame does not change to a broadcast address. It does this in a rudimentary way, it simply copies the data to all the nodes connected to the hub. This may lead to a collision on the network. A *switch* is a device that performs switching which reduces broadcast domain. Specifically, it forwards and filters chunk of data communication between ports (connected cables) based on the MAC addresses in the packets. This is distinct from a hub in that it only forwards the datagram to the ports involved in the communications rather than all ports connected. Strictly speaking, a switch is not capable of routing traffic based on IP address, which is necessary for communicating between network segments or within a large or complex LAN. Some switches are capable of routing based on IP addresses but are still called switches as a marketing term. A switch normally has numerous ports with the intention that most or all of the network be connected directly to a switch, or another switch that is in turn connected to a switch.

A *router* is a networking device that forwards data packets between networks using headers and forwarding tables to determine the best path to forward the packets. Routers also provide interconnectivity between like and unlike media. This is accomplished by examining the *header* of a data packet and making a decision on the next hop to which it should be sent. They use preconfigured static routes, status of their hardware interfaces, and routing protocols to select the best route between any two subnets. A router is connected to at least two networks, commonly two LANs or WANs or a LAN and its ISP's network.

A *gateway* is a combination of hardware and software that translates between two different protocols and acts as the connection point to the Internet. Gateways are the indispensable components in order to achieve multimedia communications between terminals connected to heterogeneous networks that use different protocols and have different network characteristics.

As the Internet is also used for commercial and business transactions, the issue of information security is very important. Typically, the network firewalls are used to safeguard any network connected to the Internet. Firewalls are configured and managed to realise an important security policy for the particular needs of a given company or entity. A network gateway typically act as a firewall in that it filters packets and separates a proprietary corporate network, such as an Intranet, from a public network, such as the Internet.

16.8 CONVERGENCE OF COMPUTING AND COMMUNICATION

16.8.1 Computer Telephony Integration

Computer Telephony Integration (CTI) is a set of technologies for integrating computers and telephone systems. CTI functionality falls into two general categories: enabling computers to control the telephone system and enabling the telephone system to display information on computers. A user with a CTI-enabled computer will be able to dial the telephone, answer the telephone, and hang up the telephone, all from a computer. CTI enables users to dial the phone from address books stored on their computer. Most CTI systems facilitate users to interact with teleconferencing systems. A CTI-enabled computer will also display information from the telephone system, such as Caller-ID. Sophisticated CTI systems also display information from touch-tone and Interactive Voice Response (IVR) systems.

16.8.2 Network Convergence

Network convergence is an efficient coexistence of telephone, video and data communication within a single network. The use of multiple communication modes in a single network offers convenience and flexibility not possible with separate infrastructures. Network convergence is also called *media convergence*. Nowadays, a plethora of innovative applications are available, for example, Web surfing, VoIP (Voice over IP), FoIP (Fax over IP), streaming media, videoconferencing, e-business and e-learning.

16.8.3 Device Convergence

The device convergence means bringing the three worlds of communications, computing and consumer electronics in a single device. This convergence is ushering in a new epoch of multimedia, in which voice, data and images are combined to render services to the users. Typical examples include Web TV, e-mail and World Wide Web access via digital TV decoders and mobile phones, webcasting of radio and TV programming on the Internet.

16.8.4 Voice-over-Internet Protocol (VoIP)

VoIP is a protocol optimised for the transmission of voice through the Internet or other packet-switched networks. VoIP is often used abstractly to refer to the actual transmission of voice (rather than the protocol implementing it). This latter concept is also referred to as IP telephony, Internet telephony, voice-over broadband, broadband telephony and broadband phone. VoIP-to-VoIP phone calls are sometimes free, while VoIP calls connecting to public switched telephone networks (VoIP-to-PSTN) may have a cost that is borne by the VoIP user. Voice-over-IP systems carry telephony signals as digital audio, typically reduced in data rate using speech data compression techniques, encapsulated in a data packet stream over IP.

16.8.5 Videoconferencing

A videoconferencing (or *videoteleconferencing*) is a set of interactive telecommunication technologies, which allow two or more locations to interact via two-way video and audio transmissions simultaneously. The core technology used is digital compression of audio and

video streams in real-time. The hardware or software that performs compression is called a codec (coder/decoder). The compressed digital stream is transmitted through a digital network of some kind (usually ISDN or IP network).

16.9 EMERGING TECHNOLOGIES IN COMPUTING

16.9.1 Mobile Computing

Mobile computing enables users to work from a nonfixed location using portable computing and communications devices such as laptops, notebooks, palmtops, smart cell phones and Personal Digital Assistants (PDAs). Mobile voice communication has been widely established throughout the world impacting our life at work and home. An extension of the above development forms the basis of mobile computing. A PDA is such a device, which have both color screens and audio capabilities, enabling them to be used as mobile phones, smartphones, Web browsers, or portable media players. Many PDAs can access the Internet, intranets or extranets via Wi-Fi. The wireless-fidelity (Wi-Fi) is used to connect to the Internet when within range of a wireless network. The main aims of Wi-Fi are the following: make access to information easier, ensure compatibility and co-existence of devices, eliminate cabling and wiring, and eliminate switches, adapters, plugs, pins and connectors. The Bluetooth is a wireless protocol utilising short-range communications technology facilitating data transmission over short distances from fixed and/or mobile devices, creating wireless Personal Area Networks (PANs). Bluetooth provides a way to connect and exchange information between devices such as mobile phones, telephones, laptops, personal computers, printers, GPS receivers, digital cameras and video game consoles.

16.9.2 Grid Computing

Grid computing is a form of distributed computing whereby a "super and virtual computer" is composed of a cluster of networked, loosely coupled computers, acting in concert to perform very large tasks. This technology has been applied to computationally intensive scientific, mathematical and academic problems through volunteer computing. It is also used in commercial enterprises for diverse applications such as drug discovery, economic forecasting, seismic analysis, and back-office data processing in support of e-commerce and Web services. Grid computing facilitates more cost-effective use of a given amount of computer resources as well as provides as a way to solve problems that can't be approached without an enormous amount of computing power. In some grid computing systems, the computers may collaborate rather than being directed by one managing computer. One likely area for the use of grid computing will be pervasive computing applications—those in which computers pervade our environment without our necessary awareness.

16.9.3 Cloud Computing

Cloud computing means Internet ('Cloud') based development and use of computing. It is a paradigm in which information is permanently stored in servers on the Internet and

cached temporarily on clients that include desktops, entertainment centres, table computers, notebooks, wall computers, handhelds, etc. The majority of cloud-computing infrastructure currently consists of reliable services delivered through next-generation data centres that are built on compute and storage virtualisation technologies. The services are accessible anywhere in the world, with the cloud appearing as a single point of access for all the computing needs of consumers.

16.9.4 Green Computing

Green computing is the study and practice of using computing resources efficiently. The goals are similar to green chemistry, i.e. reduce the use of hazardous materials, maximise energy efficiency during the product's lifetime, and promote recyclability or biodegradability of defunct products and factory waste, etc. Modern IT systems rely upon a complicated mix of people, networks and hardware. Green computing is the environmentally responsible use of computers and related resources. Such practices include the implementation of energy-efficient central processing units (CPUs), servers and peripherals as well as reduced resource consumption and proper disposal of electronic waste (e-waste). The work habits of computer users and businesses can be modified to minimise adverse impact on the global environment.

16.10 | NEXT-GENERATION COMPUTING PARADIGMS

Silicon-based technologies have dominated the computing domain since its inception more than six decades ago. Researchers have been pursuing a number of alternatives to silicon-based computing. Now, the technology pundits say that silicon-based technologies are fast reaching towards the saturation point and would have to accommodate a swell of new computing paradigms, which are decidedly silicon-free. These nontraditional computing paradigms include DNA, quantum, molecular, biological, optical and nano-fluidic computing. All these are in their infancy but have the potential to complement the current silicon-based technologies if not to replace them. Most probably, these new paradigms would be applied to special-purpose computing. The following paragraphs provide a few examples of new computing paradigms.

DNA computing is a form of computing which uses DNA, biochemistry and molecular biology, instead of the traditional silicon-based computer technologies. It is fundamentally similar to parallel computing in that it takes advantage of the many different molecules of DNA to try many different possibilities at once. For certain specialised problems, DNA computers are faster and smaller than any other computer built so far. However, DNA computing does not provide any new capabilities from the standpoint of computability theory, the study of which problems are computationally solvable using different models of computation.

Quantum computing is direct use of distinctively quantum mechanical phenomena, such as superposition and entanglement, to perform operations on data. In a classical (or conventional) computer, information is stored as bits; in a quantum computer, it is stored as qubits (quantum binary digits). The basic principle of quantum computation is that the quantum properties can

be used to represent and structure data, and that quantum mechanisms can be devised and built to perform operations with these data. If large-scale quantum computers can be built, they will be able to solve certain problems much faster than any of current computers.

Molecular computing are massively parallel computers taking advantage of the computational power of molecules (specifically biological). Moletronics specifically refers to the sub-field of physics, which addresses the computational potential of atomic arrangements. *Molecule cascade computing* is the newest area in the development of alternatives to traditional computing. This technique is based on forming circuits by creating a precise pattern of carbon monoxide molecules on a copper surface. By nudging a single molecule, it has been possible to cause a cascade of molecules, much like toppling dominoes. Different molecules can represent the 1's and 0's of binary information, making possible calculations. While this technique may make possible circuits hundreds of thousands of times smaller than those used today, it shares with the other alternatives the fact that a number of problems must be solved for it to be ever suitable for practical applications.

Biological computing is the use of living organisms or their component parts to perform computing operations or operations associated with computing, e.g. storage. The various forms of biological computing take a different route than those used by quantum or optical computing to overcome the limitations to performance that silicon-based computers face. Rather than focusing on increasing the speed of individual computing operations, biological computing focuses on the use of massive parallelism, or the allocation of tiny portions of a computing task to many different processing elements. Each element in and of itself cannot perform its task quickly, but the fact that there is an incredibly huge number of such elements, each performing a small task, means that the processing operation can be performed far more quickly.

Optical computing uses light instead of electricity (i.e. photons rather than electrons) to manipulate, store and transmit data. Photons have fundamentally different physical properties than electrons, and researchers have attempted to make use of these properties, mostly using the basic principles of optics, to produce computers with performance and/or capabilities greater than those of electronic computers. Most research projects focus on replacing current computer components with optical equivalents, resulting in an optical digital computer system processing binary data. This approach appears to offer the best short-term prospects for commercial optical computing, since optical components could be integrated into traditional computers to produce an optical/electronic hybrid.

SUMMARY

- Basic computers, classification and different types of computers are introduced with their hardware and software specifications.
- Future computers and next generation computing paradigm are also discussed.

EXERCISES

➥ Review Questions

1. What are the main objectives of fifth-generation computers?
2. What are the main functions of the control unit in microprocessors?
3. What components are provided for the computers with on chip memory (storage)?
4. Compare the features of CISC and RISC.
5. What are the main functions of an operating system?
6. What are the main functions of gateway in the network components?
7. Discuss the merits of convergence in computing.
8. What is green computing? Also discuss its importance.
9. In what way is the nontraditional computing is superior to silicon-based computing? What are the main properties of nontraditional computing?

➥ Multiple-Choice Questions

1. How many generation of computers are there?
   (a) Four (b) Five (c) Six (d) Two
2. The computer system is used to
   (a) store data (b) process data (c) retrieve data (d) all of these
3. Cache memory is used
   (a) to speed up data processing (b) to slow down data processing
   (c) to start processing (d) none of these
4. Which one is not an input device?
   (a) Light Pen (b) Graphics Tablet (c) Storage devices (d) Printers
5. Virtual memory helps process
   (a) small programs (b) larger programs
   (c) the programs requiring more memory after RAM (d) none of these
6. Multitasking means
   (a) to run the same program on different machines
   (b) to run more than one program on a single machine
   (c) to run a single program on the single machine
   (d) all of the above
7. Which one is not a network component?
   (a) Huts (b) Switch (c) Router (d) Hard disk
8. Which one helps route packets in the network?
   (a) Hub (b) Switch (c) Router (d) Bridges
9. Which is not an operating system?
   (a) Win XP (b) Ms Word (c) Unix (d) Linux
10. Which network topology is the best?
    (a) Bus (b) Star
    (c) Ring (d) It depends on the geographic conditions

ANSWERS

✦ Multiple-Choice Questions

1. (b) **2.** (d) **3.** (a) **4.** (d) **5.** (c) **6.** (b) **7.** (d) **8.** (c) **9.** (b) **10.** (d)

CHAPTER 17

Clock and Timing Circuits

GOALS AND OBJECTIVES

- Introduction of clock and timing circuits
- Analysis and functional operation of IC 555 timer—astable and monostable operation with block diagrams
- Time constant changes in different operations

17.1 INTRODUCTION

The heart of a computer is the clock which produces clock pulses with the *precise cycle time*. The clock pulses advance the various circuits of the computer one step at a time. The clock pulses are normally rectangular; positive pulses followed by negative pulses as shown in Fig. 17.1. The duty cycle could be other than 50% but the cycle time must be precise. All logic elements must complete their transition in one cycle.

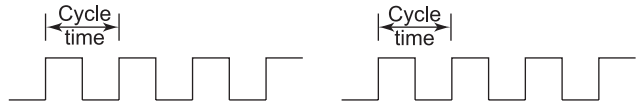


Fig. 17.1 Ideal clock waveform

The clock pulses can be generated by an IC-555 timer or by a crystal oscillator. We will study the IC-555 timer.

17.2 THE IC-555 TIMER

A popular and versatile analog-digital integrated circuit is the 555 timer. The entire circuit is housed in an 8-pin package as shown in Fig. 17.2, whose detailed circuitry is drawn in Fig. 17.3. It comprises

- Two op-amp comparators set at $(2/3\ V_{CC})$ and $(1/3\ V_{CC})$ respectively
- A three-resistor circuit, to obtain voltage $2/3\ V_{CC}$ and $1/3\ V_{CC}$, to set the comparator

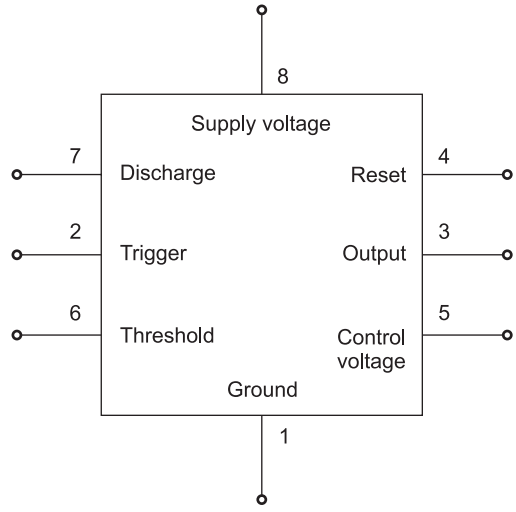


Fig. 17.2 Pin diagram of 555 timer

- The comparator's output sets or resets the *flip-flop (F/F)* which feeds the output stage. The F/F operates as

  $R = 1, S = 0$, output = 0

  $R = 0, S = 1$, output = 1

The F/F also operates a transistor which when driven low, discharges a timing capacitor (external).

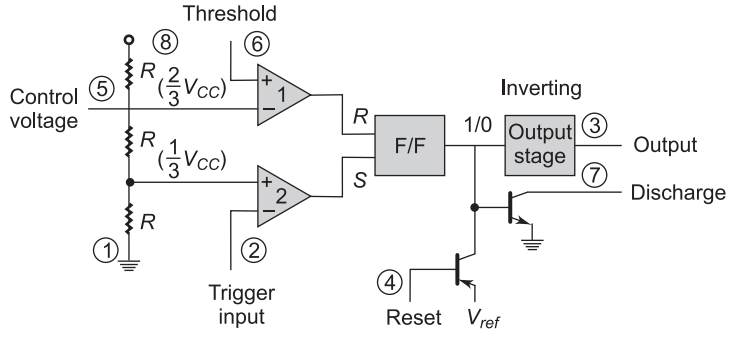


Fig. 17.3 Internal details of IC-555 timer

17.3 ASTABLE OPERATION

A common application of the 555 timer is as astable multivibrator or *clock circuit*. Figure 17.4 shows a stable circuit built using two external resistors and a capacitor, which sets the timing in intervals of the output.

The capacitor begins to charge from the dc source V_{CC}. When the voltage of the threshold pin 6 tends to increase beyond 2/3 V_{CC}, the comparator 1 saturates and its output triggers the flip-flop and so the output at pin 3 goes low. At the same time, the transistor becomes 'on' causing the output at pin 7 to discharge the capacitor through R_2 at time constant $\tau_2 = R_2C$.

As the capacitor voltage which is the trigger input at pin 2 falls below (1/3) V_{CC}, the comparator 2 output causes the flip-flop to reset, the output at pin 3 becomes high and the transistor goes 'off'. The capacitor now begins to charge through R_1 and R_2 at the time constant $\tau_1 = (R_1 + R_2)C$. The process then repeats continuously.

The output waveform and capacitor charging and discharging are shown in Fig. 17.5.

By writing the exponential expression for R_2C discharging and $(R_1 + R_2)$ charging from V_{CC}, we can find T_{high} and T_{low} periods. These results are,

$T_{high} \approx 0.7\ (R_1 + R_2)C$; exact 0.7 is 0.639

$T_{low} \approx 0.7\ R_2C$

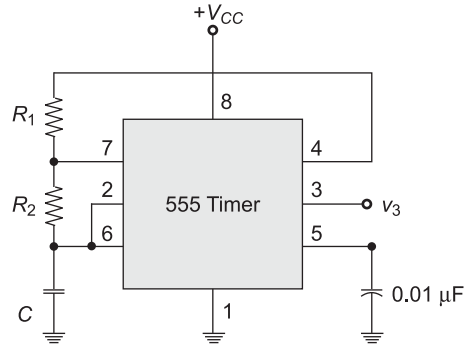


Fig. 17.4 Astable operation of IC-555 timer

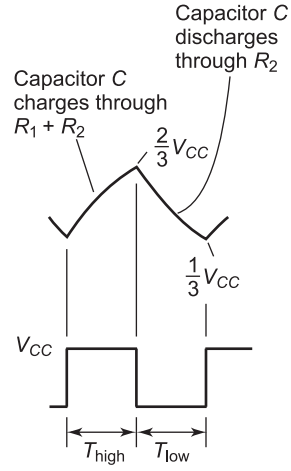


Fig. 17.5 Output waveform and capacitor charging and discharging cycle

The oscillation period

$$T = T_{high} + T_{low}$$

The oscillation frequency

$$f = \frac{1}{T} = \frac{1.44}{(R_1 + 2R_2)C} \tag{17.1}$$

$$\text{Duty cycle} = \frac{T_{high}}{T_{high} + T_{low}} = \frac{R_1 + R_2}{R_1 + 2R_2}$$

$$= \frac{R_1 + R_2}{R_1 + 2R_2} \text{ ; less than 50\%} \tag{17.2}$$

✦ Derivation of Charging and Discharging Time

❑ ***Discharging*** $v_c(t_1) = \frac{2}{3}V_{CC}\, e^{-t_1/\tau_1} = \frac{1}{3}V_{CC}$

$$2e^{-t_1/\tau_1} = 1, \quad \ln 2 - \frac{t_1}{\tau_1} = 0 \text{ or } t_1 = \tau_1 \ln 2 \tag{17.3}$$

❑ ***Charging*** $v_c(t) = V_{CC} + Ae^{-t/\tau_2}$

At t_2, $v_c(t_2) = V_{CC} + Ae^{-t_2/\tau_2} = \frac{1}{3}V_{CC} \quad \Rightarrow A = -\frac{2}{3}V_{CC}$

$$v_C(t_2) = V_{CC} - \frac{2}{3}V_{CC}\, e^{-t_2/\tau_2} = \frac{2}{3}V_{CC}$$

$$1 - \frac{2}{3}e^{-t_2/\tau_2} = \frac{2}{3}$$

$$2e^{-t_2/\tau_2} = 1$$

$$t_2 = \tau_2 \ln 2 \tag{17.4}$$

17.4 | MONOSTABLE OPERATION

In monostable operation, a trigger (negative edge) produces a high output, which lasts for a time interval depending on *RC*. The 555 timer connected for monostable operation is shown in Fig. 17.6(a).

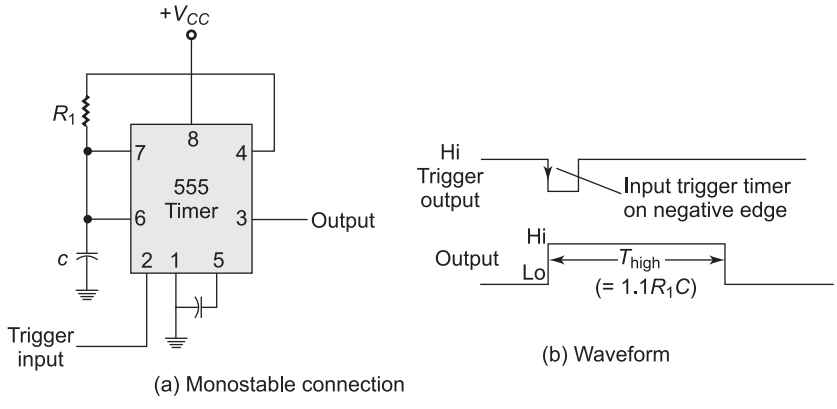


Fig. 17.6 Monostable operation of IC-555 timer

A negative edge trigger is applied at pin 2. Comparator 2 and F/F cause the output at pin 3 to go high, the transistor is 'off', and the capacitor begins to charge through R_1 at $\tau_1 = R_1C$. During this period, the output stays high. So the voltage at pin 6 resistor reaches (2/3) V_{CC}, the output goes low, and the transistor discharges the capacitor. The output stays low afterwards.

The output stays high till the capacitor charges to $(2/3)V_{CC}$. This time is

$$T_{\text{high}} = 1.1\ R_1C \tag{17.5}$$

✦ Derivation

During charging of the capacitor,

$$v_c(t) = V_{CC}\ (1 - e^{-t/R_1C}) \tag{17.6}$$

At T_h $$v_c(t) = \frac{2}{3}V_C \tag{17.7}$$

Substituting in Eq. (17.6), we find

$$T_h = CR_1\ln 3 = 1.1\ R_1C$$

The standard logic symbol for a monostable operator is shown in Fig. 17.7(a). The input is labelled *TRIGGER*, and the output is Q. The complement of the Q output may also be available at $\overline{Q}$. The input trigger circuit may be sensitive to either a positive or negative trigger. In this case, it is negative-edge-triggered. Usually, the output at Q is low when the circuit is in its stable state.

A typical set of waveforms showing the proper operation of a monostable circuit is shown in Fig. 17.7(b). In this case, the circuit is sensitive to a negative trigger at the trigger input, and the output is low when the circuit rests in its stable state. Once triggered, Q goes high, remains high for a predetermined time t, and then switches back to its stable state until another negative trigger appears at the trigger input.

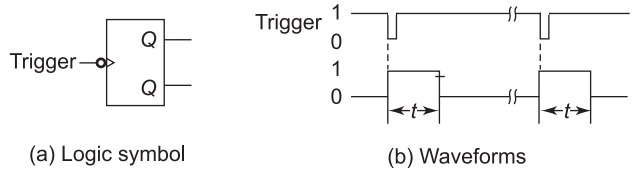


Fig. 17.7 Monostable circuit

EXAMPLE 17.1 For a 555 timer connected as astable multivibrator,

$R_1 = R_2 = 7.5\ \text{k}\Omega$, $C = 0.1\ \mu\text{F}$, $V_{CC} = 5\ \text{V}$

Calculate the frequency of operation and sketch the output waveform.

Solution

$$T_h = 0.7\,(R_1 + R_2)C = 0.7 \times 15 \times 10^3 \times 0.1 \times 10^{-6} = 1.05\ \text{ms}$$

$$T_l = 0.7\,R_2C = 0.525\ \text{ms}$$

$$T = 1.05 + 0.525 = 1.575\ \text{ms}$$

$$f = \frac{1}{T} = \frac{10^3}{1.575} = 635\ \text{Hz}$$

$$V_{oh} = 5\ \text{V}$$

$$V_{ol} = 0.2\ \text{V}$$

The output waveform is drawn in Fig. 17.8.

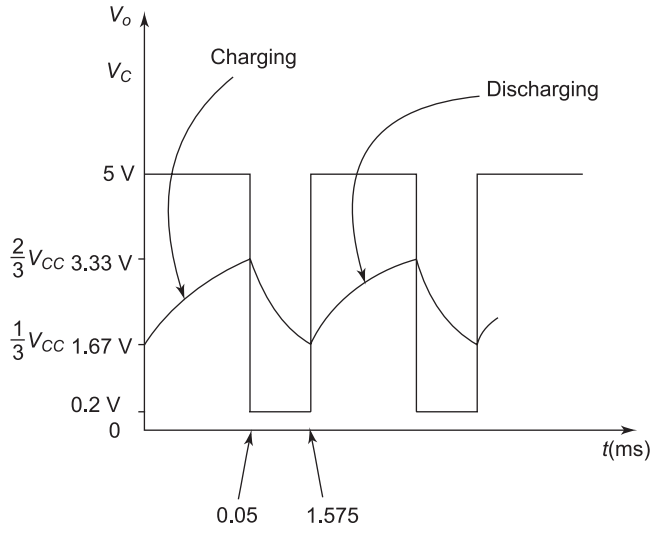


Fig. 17.8

EXAMPLE 17.2 In Example 17.1, what should be the value of *C* for the astable frequency of 300 kHz?

Solution

$$T = 0.7\,(R_1 + R_2)C$$

$$f = \frac{1.44}{(R_1 + 2R_2)C} = \frac{1.44}{22.5 \times 10^3 C} = 300 \times 10^3$$

$$C = \frac{1.44}{22.5 \times 300 \times 10^6} = 0.213 \text{ nF}$$

EXAMPLE 17.3 For the monostable multivibrator of Fig. 17.6, R_1 = 75 kΩ; calculate *C* for one shot high period of 25 μs.

Solution

$$T_h = 1.1\,R_1C$$

$$25 \times 10^{-6} = 1.1 \times 7.5 \times 10^3 C$$

or
$$C = \frac{25 \times 10^{-6}}{1.1 \times 7.5 \times 10^3} = 3 \text{ nF}$$

SUMMARY

- Clock and timing circuits are introduced.
- Analysis and operation of IC 555 timer with block diagrams have been explained.

EXERCISES

Review Questions

1. What is a timer? (in general)
2. In brief, introduce the 555 work as an oscillator.
3. What is the internal structure of 555 timer according to its operation?
4. What are time-constant changes in different operations?
5. Explain the importance of threshold and discharge pins of the 555 timer.

Problems

1. A 555 timer is connected for astable operation at 400 kHz. Calculate the value of *C* if $R_1 = R_2 = 8$ kΩ.
2. Draw the circuit of a monostable 555 timer to produce a single pulse of time period 25 μs. If $R_1 = 8$ kΩ, what value of *C* is needed?

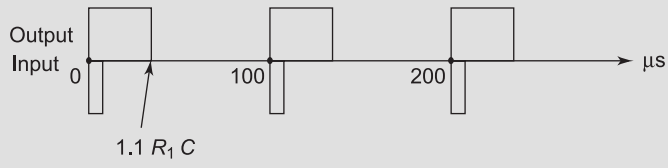
3. Sketch the input and output waveforms for monostable (one shot) 555 timer triggered by 12 kHz clock for R = 5.45 kΩ and C = 5 nF.

➥ Multiple-Choice Questions

1. The supply voltage range of 555 timer is
   (a) 5 to 22 V (b) 10 to 30 V (c) 5 to 18 V (d) none of these
2. The 555 timer output is compatible with
   (a) N-MOS (b) P-MOS (c) C-MOS (d) none of these
3. Which flip-flop is used in the 555 timer?
   (a) SR F/F (b) JK F/F (c) T F/F (d) D F/F
4. In the 555 timer, monostable operation of pulse width of the output is shown in the equation
   (a) $t_p = 2.1\ R_1C_1$ (b) $t_p = 3.2\ R_1C_1$ (c) $t_p = 4\ R_1C_1$ (d) $t_p = 1.1\ R_1C_1$
5. The frequency oscillation of a 555 timer in a stable mode is given by
   (a) $f = \dfrac{1.44}{(R_1 + 2R_2)C_1}$ (b) $f = \dfrac{2.33}{(R_1 + 2R_2)C_1}$ (c) $f = \dfrac{3.12}{(R_1 + 2R_2)C_1}$ (d) $f = \dfrac{4.34}{(R_1 + 2R_2)C_1}$
6. In the bi-stable multivibrator, the stable state is
   (a) one stable state (b) two stable states (c) three stable states (d) none of these
7. A Bi-stable multivibrator is commonly used in a
   (a) square waveform (b) sinusoidal waveform
   (c) triangular waveform (d) sawtooth waveform
8. In the monostable multivibrator, frequency of oscillation is
   (a) $f = 1\ 1.50\ RC$ (b) $f = 1\ 1.38\ RC$ (c) $f = 1\ 2.34\ RC$ (d) none of these
9. In the 555 timer, the upper comparator is connected with
   (a) trigger terminal (b) discharge terminal (c) threshold terminal (d) none of these
10. A bistable multivibrator is normally called
    (a) oscillator (b) voltage regulator (c) voltage controller (d) flip-flop

ANSWERS

✦ Problems

1. 15 μF
2. 2.8 nF
3. 

✦ Multiple-Choice Questions

1. (c) **2.** (c) **3.** (a) **4.** (d) **5.** (a) **6.** (b) **7.** (a) **8.** (b) **9.** (c) **10.** (a)

CHAPTER 18

Communication Engineering

GOALS AND OBJECTIVES

- Introduction of electrical technology and communication systems
- Explanation of frequency spectrum
- Discussion of various types of modulations—amplitude, AM detection, frequency and phase modulation
- Explanation of digital modulation and digital communication
- Types of data transmission—asynchronous and synchronous transmission
- Introduction of different types communications—transmission lines, radio waves, antennas, television and satellite communication
- Telephone network and principle operation of mobile phone
- Discussion of various types of communications—microwave communications and optical fibre communication and its applications

18.1 INTRODUCTION

In electrical technology, one of the greatest applications is communication systems. It enhances the people's lives to a great extent. Today, we enjoy satellite television, fax machine and cellular phones, etc. Communication systems include the generation, storage and transmission of information. The basic elements of communication systems are transmitter, receiver and communication channels. We begin with analog communication, because many digital communication techniques require analog technology to function.

18.2 ELEMENTS OF COMMUNICATION SYSTEMS

Communication systems are used to transfer information from a generation point to the place where it is needed/processed. The information at the generation point is not in the form that can travel long distance through the channel. So a device called a *modulator cum transmitter* is needed. In the receiving end, the information must undergo the reverse process such as decoding/demodulation. The source is mostly analog and sometimes it may be digital. Fig. 18.1 shows the very basic elements of a communication system.

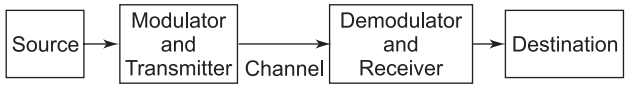


Fig. 18.1 Communication system

Common examples of sources are analog audio, video, signals from measuring devices in the field, and of course some digital data. The source can be described by its signal frequency, like an audio signal is in the range at 20 Hz to 20 kHz and a video signal in the ranges from dc to 4.2 MHz. Digital signals have different bandwidths depending on transmission rate and the method used to convert binary ones and zeroes to electrical signals.

Communication channels can be a pair of conductors, optical fiber or just free space. The signal can be directly sent through a twisted-pair telephone cable. But the radio link through free space cannot be used directly for audio signals; instead an antenna of great height would be required. In this case it is a carrier signal whose frequency is such that it will travel for long distance through a channel. A carrier wave will be altered/modulated by the information signal in such a way that the information can be recovered at the destination. Here, the information signal is called the modulated signal. The carrier signal frequency is much higher than information signal frequency.

18.3 MODULATION

The carrier is a sine wave, and the information signal may be any type of wave form. The instantaneous amplitude of the information signal is used to vary some parameters of the carrier.

The sine-wave carrier signal is represented as

$$e(t) = E_c \sin(\omega_c t + \theta) \tag{18.1}$$

$e(t)$ = instantaneous voltage as a function of time
E_c = peak magnitude
ω_c = frequency, rad/s, and $\omega_c = 2\pi f_c$
t = time
θ = phase angle in radians

In modulation, the amplitude E_c, frequency ω_c and phase θ can be varied in accordance with the instantaneous values of the modulating signal. The modulated signal is not a single frequency signal and it occupies a wide bandwidth. In addition, the bandwidth of the modulated signal depends on the modulating-signal frequency range and modulating scheme in use.

The wavelength of the signal is inversely proportional to frequency of the signal, i.e.,

$$v = f\lambda$$

v = velocity of the wave in meters per second
f = frequency of the wave in hertz
λ = wavelength in meters.

✦ Electromagnetic Spectrum

The term electromagnetic spectrum refers to a range of frequencies and wavelengths associated with existing electromagnetic waves. The different bands specified in the spectrum have dedicated applications in communication and technology. Different modes of communication, such as radio communication, mobile communication, and optical communication, have specific spectrum range for their operation and by no means can afford to extend their application into frequency spectrum dedicated for another application. The tuning of radio is the best example to understand how this spectrum works. When one tunes into 104.7 FM, the station that is broadcasting at frequency 104.7 MHz is being tuned in. In order to listen to different station, we have to turn the dial to another frequency, say 93.9 FM. No two broadcasting stations transmit signal using identical frequencies, otherwise there may exist intersignal interference resulting in unidentified signal. The different governing bodies responsible for coordinating the usage of particular band of spectrum include ITU (International telecommunication Unit), IEEE, and EU. The electromagnetic spectrum as standardised by IEEE for implementation in communication systems is shown in Fig. 18.2.

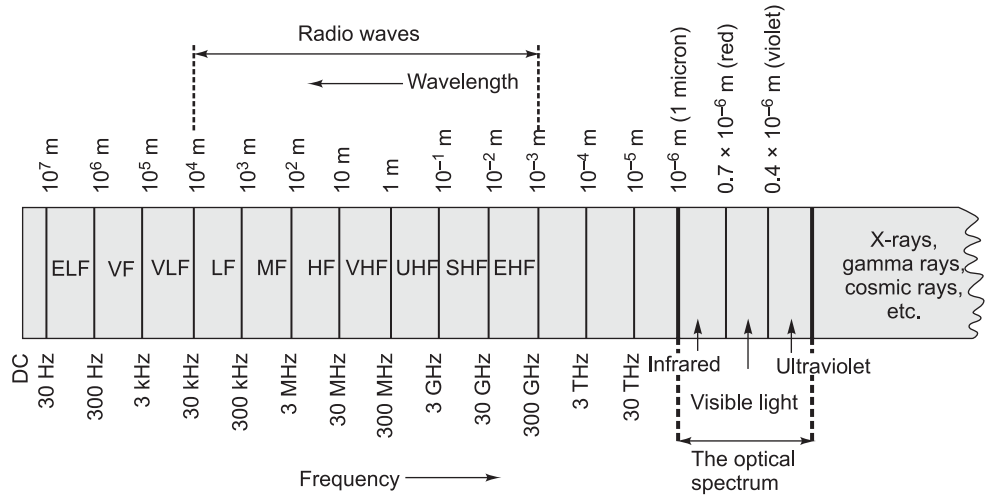


Fig. 18.2 Electromagnetic spectrum

The different bands specified and their applications are summarised in Table 18.1 given below.

Table 18.1 *Different frequency bands in electromagnetic spectrum*

S. No.	Frequency	Wavelength	Description	Applications
1.	Up to 300 Hz	Up to 10,000 km	ELF (Extremely Low Frequency)	
2.	300 Hz-3 kHz	100-1000 km	Voice Frequency	
3.	3 kHz-30 kHz	10-100 km	VLF (Very Low Frequency)	Maritime navigation signals
4.	30 kHz-300 kHz	1-10 km	LF (Low Frequency)	Navigation, Time standards
5.	300 kHz-3 MHz	0.1-1 km	MF (Medium Frequency)	Marine navigation, Aircraft navigation, AM broadcast
6.	3 MHz-30 MHz	10-100 m	HF (High Frequency)	Broadcasting, Mobile radio, Shortwave radio
7.	30 MHz-300 MHz	1-10 m	VHF (Very high Frequency	FM broadcasting, TV signal
8.	300 MHz-3 GHz	10-100 cm	UHF (Ultra High Frequency)	Cell phones, WLAN, Personal Area Network, GPS
9.	3 GHz-30 GHz	1-10 cm	SHF (Super High Frequency)	Satellite, Radar, Backhaul network, Microwave communication
10.	30 GHz-300 GHz	0.1-1 cm	EHF (Extremely High Frequency)	Satellite, Radar, Experimental

The different radiations that have been presented in the electromagnetic spectrum includes Gamma radiation (wavelength less than 10^{-11} m), X-ray radiation (wavelength range of 0.01-10 nm), Ultraviolet radiation (wavelength range of 10-400 nm), Visible radiation (wavelength

range of 390-750 nm), Infrared radiation (wavelength range of 750 nm–1 mm), Microwave radiation (wavelength range of 1 mm-1 m), Radio waves (wavelength range of 1 mm-100 km). The different applications of these radiations include:

1. The radio waves are being utilised in radio communication, transmission of TV signals, wireless networking, mobile communication, radar and other navigation systems, and communication satellites.
2. Microwaves are extremely useful in communication systems including point to point links, radar signals, and spacecraft communication. The heating effect of microwave makes it convenient to be used in medical applications like cancer treatment, cooking food through microwaves and ovens, industrial heating, etc. Some of the other applications include particle accelerators and spectroscopy.
3. Infrared radiations are widely used in research, industrial and medical fields. Typical applications include night vision devices, telecopy, thermal imaging, spectroscopy, surveillance, and short range wireless communication.
4. Visible range of electromagnetic spectrum contains the wavelengths of light that can be detected by human eye.
5. The ultraviolet rays has the property of initiating certain chemical reactions and fluorescence making it effectively implemented for lamps, LEDs, photomultipliers, photography, lithography, spectroscopy, forensic analysis, and optical sensors.
6. X-rays are often treated as ionisation radiations and their exposure to human beings for short duration can be utilised for medical applications like treatment of cancer, detection/ examination of various diseases, and X-ray spectrometry. The property of in-depth penetration is employed in making CT scanners, industrial scanners, and security scanners. Other applications include microscopy, crystallography, fluoroscopy, etc.
7. Gamma rays are implemented in high altitude balloons and satellite missions, non-contact sensors, detectors, irradiation, and medical applications like diagnostics, surgery, cancer treatment, etc.

18.3.1 Amplitude Modulation (AM)

An AM signal can be produced by using the instantaneous amplitude of the information signal (modulating signal) to vary the peak amplitude of higher frequency carrier. Figure 18.3 shows a modulating signal (sine wave for simplicity), the carrier signal of much higher frequency and the modulated signal. It is seen that envelopes of the peaks of the modulated signal resemble the modulating signal.

Normally, the ratio between carrier frequency and modulating signal frequency is a few thousands. The AM is not a simple linear addition of the two signal instead AM is a nonlinear process.

As the amplitude of a sine wave is a positive quantity, the amplitude variation cannot be sinusoidal. To overcome this problem, a constant quantity is added to the modulating signal which then has the form

$$E_c + E_m \sin \omega_m t,\ E_c \geq E_m \tag{18.2}$$

This signal has positive value only. It is multiplied with the carrier wave sin $\omega_c t$ to get the modulating signal.

$$e(t) = (E_c + E_m \sin \omega_m t) \sin \omega_c t \tag{18.3}$$

or
$$e(t) = E_c \sin \omega_c t + E_m \sin \omega_m t \sin \omega_c t \tag{18.4}$$

where E_c = peak amplitude of carrier in volts
E_m = peak amplitude of modulating signal in volts
ω_c = carrier frequency in rad/s
ω_m = modulating frequency rad/s
$e(t)$ = instantaneous value of the modulated signal in volts

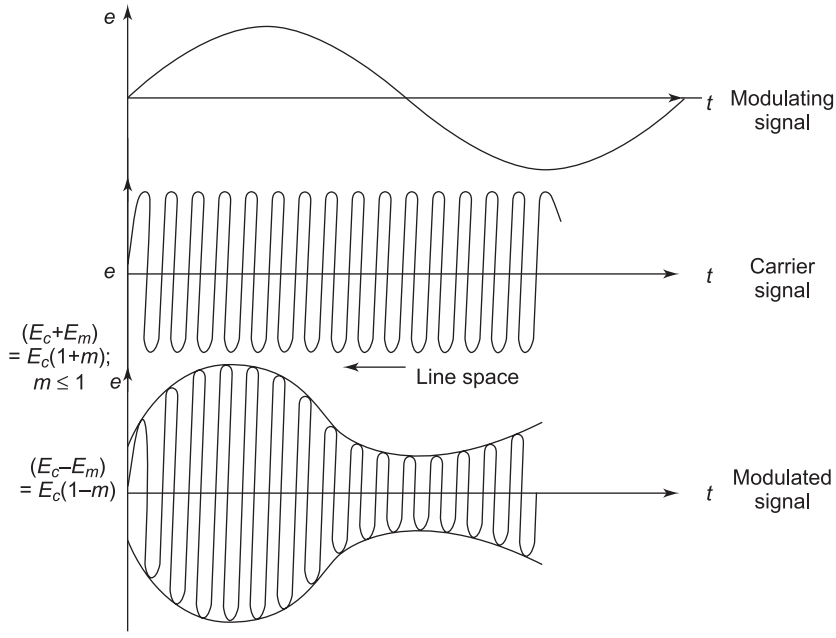


Fig. 18.3 Amplitude modulation

We can express Eq. (18.4) as

$$e(t) = E_c \sin \omega_c t + \frac{1}{2} E_m \cos (\omega_c - \omega_m)t - \frac{1}{2} E_m \cos (\omega_c + \omega_m)t \tag{18.5}$$

Carrier Difference frequency Sum frequency

The amount by which the carrier amplitude gets modified by the modulating signal depends on the *modulation index* defined as

$$m = \frac{E_m}{E_c} \tag{18.6}$$

where $m \leq 1$.

The modulated signal (Eq. 18.4) can be expressed in terms of the modulation index as

$$e(t) = E_c(1 + m \sin \omega_m t) \sin \omega_c t \tag{18.7}$$

As shown in Fig. 18.2, upper envelop in the modulated signal is

$$E_c(1 + m \sin \omega_m t) \tag{18.8}$$

and the lower envelop is

$$-E_c(1 + m \sin \omega_m t) \tag{18.9}$$

It is easily seen that the maximum magnitude of the envelop is

$$E_{\max} = E_c(1 + m) \tag{18.10}$$

And minimum magnitude is

$$E_{\min} = E_c(1 - m) \tag{18.11}$$

as shown in Fig. 18.1. It follows from these equations that

$$m = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}} \tag{18.12}$$

In terms of modulation index, the modulated signal of Eq. (18.5) can be written as

$$e(t) = \underset{\text{Carrier}}{E_c} + \underset{\text{Upper side band}}{\frac{1}{2}\, mE_c \cos(\omega_c + \omega_m)t} - \underset{\text{Lower side band}}{\frac{1}{2}\, mE_c \cos(\omega_c - \omega_m)t} \tag{18.13}$$

For a single modulating frequency, the frequency component magnitude of the modulated signal are shown in Fig. 18.4. This representation is known as *frequency spectrum* or *just spectrum*.

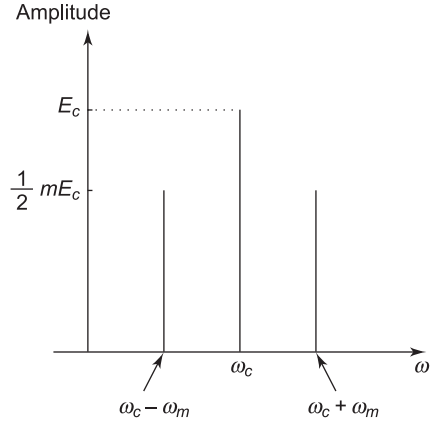


Fig. 18.4 AM spectrum for single ω_m

✦ Spectrum Power

We will consider a single modulating frequency as in Fig. 18.4.

$$\text{Power, } P = \frac{V^2}{R}$$

It is convenient to take $R = 1\ \Omega$ for comparison. Actual power can be found by dividing power by R.

From Fig. 18.4,

$$P_{t(\text{total})} = P_{C(\text{carrier})} + P_{\text{LSB(lower side band)}} + P_{\text{USB (upper side band)}} \tag{18.14}$$

$$P_C = \left(\frac{E_c}{\sqrt{2}}\right)^2 = \frac{1}{2} E_c^2$$

$$P_{LSB} = P_{USB} = \left(\frac{\frac{1}{2}mE_c}{\sqrt{2}}\right)^2 = \frac{1}{8}m^2E_c^2 = \frac{1}{4}m^2P_c$$

$$P_t = P_c + 2 \times \frac{1}{4}m^2 P_c = \left(1 + \frac{1}{2}m^2\right)P_c \quad (18.15)$$

or $$\frac{P_t}{P_c} = \left(1 + \frac{1}{2}m^2\right) \quad (18.16)$$

For $m = 1$ (maximum permissible)

$$P_t = 1.5\, P_c \text{ or } 50\% \text{ more than } P_c \quad (18.17)$$

The modulating signal (like audio and video) is a band of frequencies with varying amplitude of $\Delta\omega$ components. The spectrum of the modulating signal called the base band as shown in Fig. 18.5(a). The corresponding spectrum of the modulated signal is presented in Fig. 18.5(b). It comprises the carrier and lower and upper sidebands which are mirror image of each other. The transmission bandwidth needed is 2 ω_u times the base bandwidth.

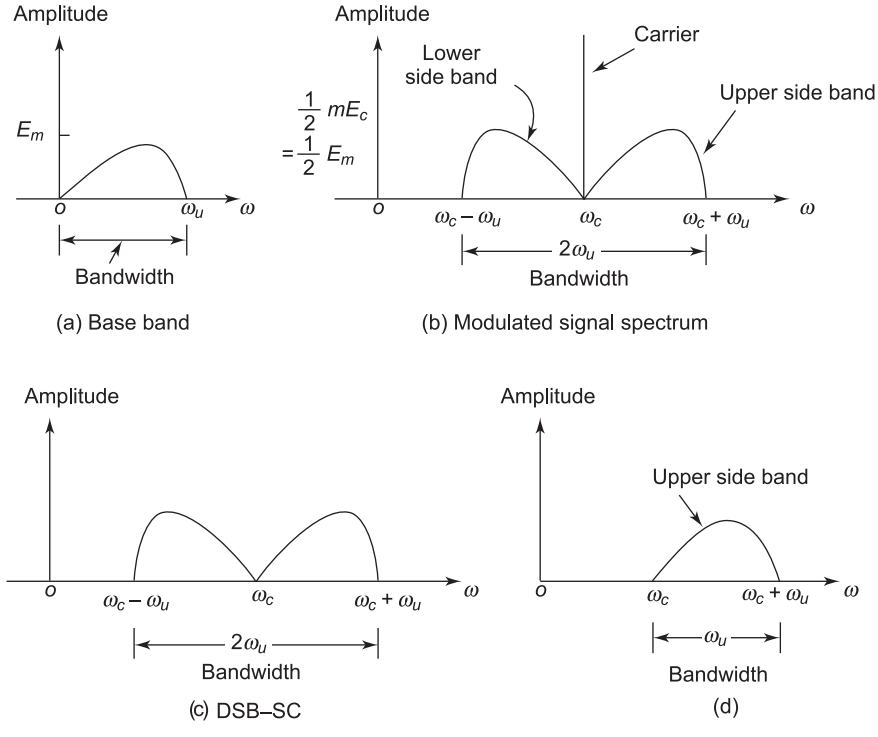


Fig. 18.5 Modulating signal

The carrier need not be transmitted, only two side bands are transmitted. This is called double side-band suppressed carrier, DSB–SC as shown in Fig. 18.5. DSB-SC requires less

transmission power but the carrier has to be generated at the receiving end by a high-frequency oscillator. As a result DSB–SC detector is therefore more complicated than normal AM detector. This type of modulation is used in conjunction with FM (frequency modulation) stereo broadcasting.

As upper and lower side bands are mirror images of each other, it is sufficient to transmit only the upper side band as shown in Fig. 18.5(d). This is called *Single Side Band (SSB)* signal. The detector at the receiving end becomes quite complicated generally not used for commercial broadcast.

✦ Vestigial Side Band (VSB) Signal

This type of AM comprises carrier, upper side band and the rising part of the lower side band (called vestige). VSB modulation is used in television video transmission. The base band of 4 MHz (megahertz) is shown in Fig. 18.6. The television stations are attached to a bandwidth of 6 MHz. A particular station is allotted the band 60–66 MHz. It employs a carrier of f_c = 61.25 Hz above which the upper side band occupies 4 MHz. The carrier is transmitted along with the side band. The vestige component is accommodated in 60–61.25 as shown in Fig. 18.6. The remaining part of the allotted spectrum is used for audio signal which is FM with carrier frequency (central) of 65.75. The difference between the two carrier frequencies is 4.5 MHz. There is a narrow guard band between video and audio, not shown in the figure.

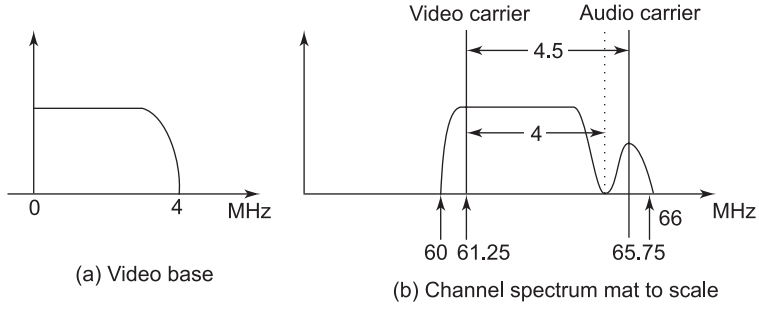


Fig. 18.6 VSB signal

✦ Amplitude Modulation Circuit

Amplitude modulation is achieved by feeding the carrier to an amplifier whose gain varies in accordance with the modulating frequency. A BJT AM circuit is drawn in Fig. 18.7 in which the carrier signal e_c is fed to the base and modulating signal e_m to the emitter. The large modulating signal causes the emitter small signal resistance r_e to vary accordingly as

$$r_e = \frac{v_{be}}{i_e} = \frac{r_\pi}{(1+\beta)} \tag{18.18}$$

As we have seen, the amplifier gain depends on r_e. Change in r_e requires a large e_m signal but the carrier amplitude must meet the condition $m \leq 1$, m = modulation index.

To avoid the problem caused by coupling and bypass capacitor, a dc amplifier modulator is employed: beyond the scope of this book.

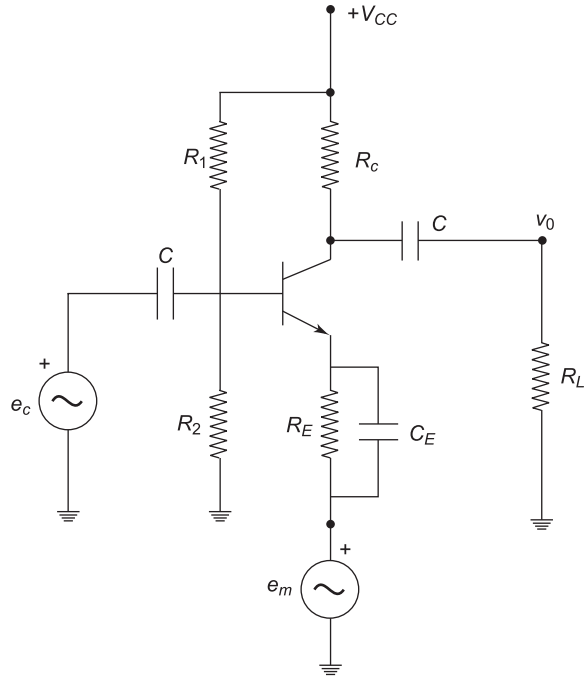


Fig. 18.7 AM modulator

For *broadcasting* the audio modulated signal, a large power has to be fed to the antenna. At the output stage therefore, a class-C power amplifier is employed. In fact the modulation is carried out at the power-amplification stage. The carrier is fed to the base and the modulating signal to the emitter, and the coupling must be though transformers for high power efficiency.

As the conduction period in a class-C amplifier is only about 90°, its output is positive modulated pulses of carrier frequency as shown in Fig. 18.8. In order to get the complete carrier modulated signal, the truncated signal is fed to the *tank circuit* (*LC* parallel circuit) as shown in the figure. At each pulse, the capacitor gets charged and in the off period it discharges through *L*, generating the negative half pulse of the same height as input pulse, thereby creating the modulated signal as shown in the Fig. 18.8.

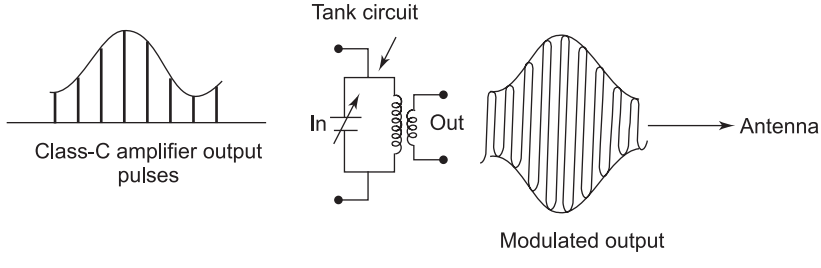


Fig. 18.8 Stages in high power AM transmission

Depending upon the station size and channel length (distance) the power to be fed to the antenna is in the range of hundred of watts, even above 100 kW. For such powers, transistors can not be used as their power handling capacity is much smaller. Therefore, a transmitting station tube (*triode*) version of the class-C amplifier is used. The RF carrier is fed to the grid of the tube and the modulating signal in series with the grid bias. The circuitry will not be described here.

18.3.2 AM Detection (Demodulation)

To recover the information from AM signal at the receiving end, the signal envelop has to be detected. The process of detection or demodulation is simpler than modulation. The signal is passed through a diode to cut-off the lower half and the peaks detected and smoothed out by a parallel *RC* circuit as shown in Fig. 18.9(a). The diode is assumed to be ideal. It is seen that the voltage v_d always equals the capacitor voltage.

As $v(t)$ rises to first peak, so does the capacitor voltage and the capacitor charges to the peak value. As $v(t)$ falls below peak, the diode stops conducting and capacitor begins to discharge through the resistor *R*. The voltage v_a falls at peak with the time constant $t = RC$ as shown in Fig. 18.9(b). As the next wave of $v(t)$ rises above v_d, the diode begins to conduct till the peak. The process then repeats.

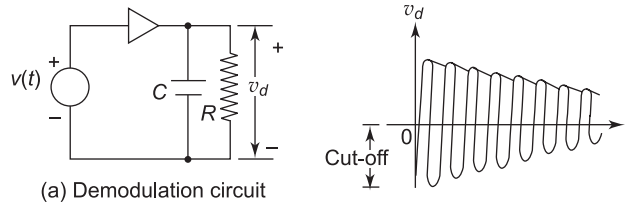


Fig. 18.9 Demodulation process

It is seen from the figure the v_d follows the modulated wave envelop except for small variation at carrier frequency which can easily be filtered out. In order that the magnitude of the variation is small, the time constant *RC* must meet the condition

$$RC >> T_C = \frac{1}{f_c} = \frac{2\pi}{\omega_c};\ T_C = \text{carrier time period} \tag{18.19}$$

However, if the time constant is very large, in the nonconducting time v_d drops very slowly and may miss the next peak. Thus, the modulating envelop may not be detected. This situation is avoided by the condition.

$$RC << T_m = \frac{1}{f_m} = \frac{2\pi}{\omega_m},\ T_m = \text{time period of modulating signal} \tag{18.20}$$

These two conditions can be combined as

$$\frac{2\pi}{\omega_c} \ll RC \ll \frac{2\pi}{\omega_m} \tag{18.21}$$

The detected envelop is indeed

$$v_m(t) = E_c + m\, E_c \sin \omega_m t \tag{18.22}$$

The dc content E_c can be filtered out by a simple *RC* low-pass filter.

18.3.3 Frequency (FM) and Phase (PM) Modulation

The frequency and phase of the carrier signals are closely related, since frequency is the rate of change of phase angle. If either frequency or phase is changed in a modulation system, the other will change as well. FM is extensively used for radio broadcasting. The most important advantage of FM or PM over AM is the possibility of a greatly improved signal-to-noise ratio. But FM signal may occupy much larger bandwidth than that required for an AM signal.

In FM, the frequency of the modulated signal varies with the amplitude of the modulating signal. In PM the phase varies directly with the modulating signal amplitude. In contrast to AM, the amplitude and the power in FM or PM signal do not change with modulation. The FM and PM modulations can be accomplished at low power levels. The FM signal can be expressed as

$$e(t) = A \sin [\omega_c + k_f e_m(t)]\, t \tag{18.23}$$

The sine-wave modulating signal $k_f e_m(t)$, the carrier and the frequency-modulated signal are exhibited in Fig. 18.10.

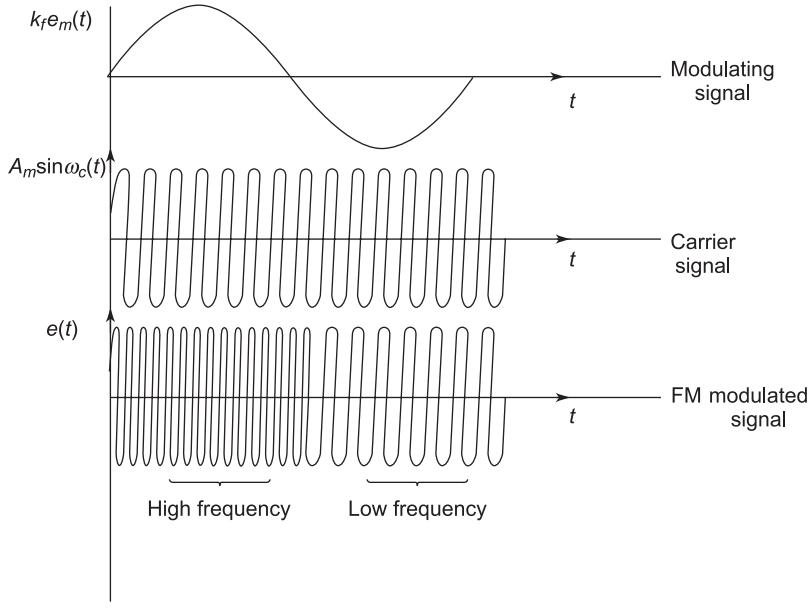


Fig. 18.10 Frequency modulation

The PM signal can be expressed as

$$v(t) = A \sin [\omega_c t + \phi(t)] = A [\sin \omega_c t + k_p e_m(t)] \tag{18.24}$$

FM and PM are generally known as *angle modulation*.

The frequency of the modulated signal is varied in accordance with instantaneous values of modulating signal. If $k_f e_m > 0$, the modulated signal frequency is greater than carrier frequency and if $k_f e_m < 0$, the modulated frequency is less than carrier frequency. And the phase is advanced if $k_f e_m > 0$ and it is lagging if $kfe_m < 0$. The FM can be generated by simply with a

voltage controlled oscillator (VCO). The frequency of a VCO is controlled by the input voltage. It increases and decreases with input voltage. The VCO with zero input voltage is set to produce carrier frequency. If the modulating frequency signal is applied at the input, VCO output is the frequency-modulated signal.

✦ Modulation Index

From Eq. (18.23) , the modulated frequency is

$$2\pi f(\text{mod}) = 2\pi f_c + k_f e_m(t) \tag{18.25}$$

For sinusoidal modulation,

$$2\pi f(\text{mod}) = 2\pi f_c + k_f E_m \sin 2\pi f_m t$$

or
$$f(\text{mod}) = f_c + k'_f E_m \sin 2\pi f_m t;\ k'_f = \frac{k_f}{2\pi} \tag{18.26}$$

The frequency deviation is

$$f_{\text{div}} = k'_f E_m \sin 2\pi f_m t \tag{18.27}$$

The maximum frequency deviation is

$$\zeta = f_{\text{div}}(\text{peak}) = k'_f E_m \tag{18.28}$$

The modulation index is defined as

$$m_f = \frac{\zeta}{f_m}$$

or
$$m_f = \frac{\text{Peak frequency deviation}}{\text{Modulating frequency}} \tag{18.29}$$

The maximum permissible value of frequency deviation is 75 kHz for commercial FM broadcasting.

✦ FM Bandwidth

The precise expression for bandwidth of FM is quite involved. An approximate relationship is

$$\text{BW}_{\text{FM}} \approx 2(\zeta + f_u) \tag{18.30}$$

where f_u = upper frequency of modulating signal bandwidth

ζ = peak frequency deviation, Eq. (18.28)

If $\zeta > f_u$, we refer it as *wideband* FM and if $\zeta < f_u$ it is referred as *narrowband* FM.

18.3.4 FM Demodulation

FM demodulation is accomplished by a Phase-Locked Loop (which is an IC chip). The block diagram of PLL is drawn in Fig. 18.11. It comprises three main parts.

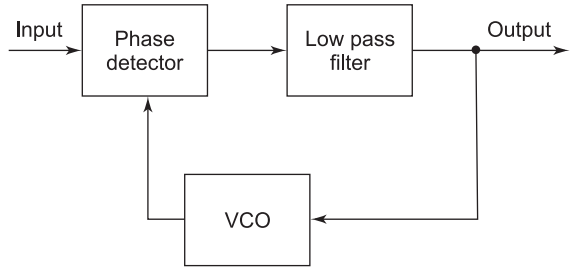


Fig. 18.11 Phase locked loop

- *VCO* Voltage Controlled Oscillator whose frequency depends on the input voltage
- *Phase Detector* It produces a voltage proportional to the frequency (or phase) difference between input and VCO frequency
- *Low-pass Filter* It cuts off the high frequencies.

At zero input voltage VCO is set at the carrier frequency. FM signal is applied at PLL input. If the modulating signal is not present, the VCO frequency matches the input frequency, the PLL output is essentially zero. With the modulating signal and the input frequency changes, it is detected by the phase detector and the PLL output voltage changes to cause the VCO frequency to change so as to match the FM signal frequency. The PLL output voltage thus tracks the modulating signal. PLL detects (or demodulates) the FM signal.

✦ Electrical Noise

Noise is a random disturbance that adversely affects the normal working of electronics devices and systems. This is generated due to the fact that the electrical charge is carried in discrete amounts equal to the charge of an electron. The study of noise is important because it gives a lower limit to the size of the electrical signal that can be amplified by a system without significant deterioration in signal quality.

In most cases noise is an undesirable signal. However, noise has some useful applications such as its use as a broadband test signal, or as a cheap source for microwave heating. Besides electrical noise, there are many other types of noises as well. These are:

1. *Acoustic noise* (noise in sound systems in auditoriums)
2. *Hydrodynamic noise* (fluctuations in smooth fluid flow as studied in civil engineering)
3. *Scintillation noise* (in radio receivers due to atmospheric effects)
4. *Round-off noise* (in computation)
5. *Quantisation noise* (in analog to digital conversion process)

Noise in any system can be either from external sources or internal sources. *Internal sources* are within the semiconductor devices, vacuum devices and in the resistors. These noises are thermal, shot, flicker ($1/f$), burst, avalanche noise and dc noise such as offset error in an operational amplifier. Thermal noise is also known as *Johnson noise*. *External noises* are produced by the lightning discharge in the sky, as galactic noise (around 1.42 GHz) produced by hydrogen clouds lying in the intergalactic space, by sun spots on the sun, by human-made

objects such as electric mixers, etc. that have universal motors in them and due to coupling from devices such as fluorescent tubes. Some of these noises have their amplitude distribution as Gaussian (thermal and shot) whereas others are non-Gaussian (flicker). Their spectral distribution may also be white (thermal) or nonwhite (flicker).

From the viewpoint of spectrum, all the electrical noise can be divided into two categories: white and coloured noise. *White noise* has equal power spectral density from dc to infinite frequency. This is a theoretical model, which is nonphysical in nature because it means that any white noise process has infinite noise power. This is not possible for a physical system. In practice, the power density of white-noise sources starts falling beyond about 1013 Hz. For the ease of modelling, thermal noise is taken as white in nature. The *coloured noise* has non-uniform power spectral density. In a communication system noise may get added to the signal passing through the channel. Such a noise is called *Additive White Gaussian Noise (AWGN).* The noise may also get multiplied to the signal; then it is called as *multiplicative noise* which happens in a fading channel.

The measure of noise in communication is the *signal-to-noise ratio (S/N).* It is the ratio of signal to noise power in decibels It is one of the most important specifications of any communication equipment.

The effect of noise can be reduced by reducing the signal bandwidth increasing the transmitting power and low-noise amplifiers for weak signals.

The noise irrespective of the origin is amplitude additive in nature. It, therefore, directly affects the AM signal but the FM signal is immune to noise. Further, in FM, the greater the frequency deviation (and so the bandwidth), more effective is the signal's noise immunisation. Thus wideband FM is superior to narrowband in relation to noise immunity.

Table 18.2 *Comparison between AM and FM*

S. No.	Characteristics	AM	FM
1.	Channel bandwidth	AM has smaller bandwidth, i.e. 2 fm	FM has larger bandwidth because it produces a larger number of side bands
2.	Operating carrier frequency	AM comparatively utilises lower carrier frequency	FM utilises higher carrier frequency (above 30 MHz) because of its higher bandwidth
3.	Transmission efficiency	AM has less transmission efficiency	FM has better efficiency but not better than AM-SC
4.	Noise performance	AM has poor noise performance	FM has better noise performance
5.	Common Channel Interference (CCI)	Due to CCI, distortion occurs in AM	FM is better due to capture effect
6.	Externally generated noise pulses	In AM such tuning is not essential	FM receiver responds slightly to noise pulses generated by external sources but if it is slightly mistuned then its ability to suppress noise pulses is highly reduced
7.	Area of reception	AM covers more distance than FM	FM is limited to distance, that is it can strongly transmit its signal in smaller area; as distance increases signal quality becomes poorer

❑ ***Additional Information***

- A small guard band is provided between FM stations to avoid interference
- FM stations operate in super VHF and UHF frequency ranges, which limits the signal to Line of Sight (LOS). Individual stations out of LOS can operate at the same frequency.

18.4 TRANSMITTER

A transmitter has to generate a signal with the right type of modulation, with enough power at the proper carrier frequency, and with good efficiency. The output of the transmitter is connected to an antenna. An ideal communication system allows the original information signal to be received exactly, except for a time delay and the transmitter would be capable of modulating any information signal frequency on the carrier, at any modulation level. Most widely used transmitters are

1. **Full carrier AM transmitters** where the carrier is generated by a frequency synthesiser and amplified to its full output power before modulation takes place
2. **Low level modulation of synthesiser (FM)**
3. **Heterodyne system (SSBSC AM)**

✦ Super-heterodyne Receiver

It is the most widely used form of receiver whose block diagram is presented in Fig. 18.12.

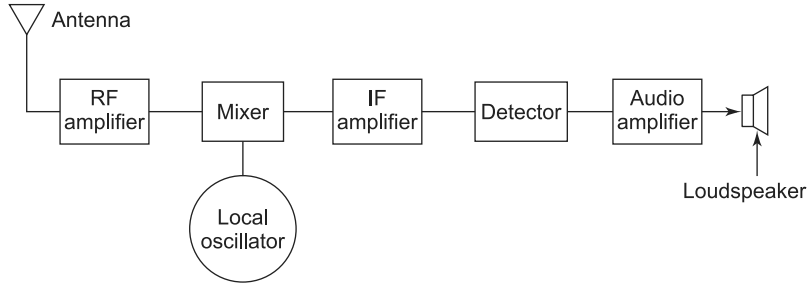


Fig. 18.12 Schematic block diagram of super-heterodyne receiver

A receiver has to perform three functions:

(i) Carrier frequency tuning (select desired channel and tune it),
(ii) Filtering (separate desired signal from other signals that entered the receiver), and
(iii) Amplification (to drive loudspeaker or monitor)

A super-heterodyne receiver does all the three functions. Practically, all the AM radio receivers are super-heterodyne receivers. The super-heterodyne receiver has remained in use because its gain, selectivity and sensitivity characteristics are superior to those of other receiver configurations such as Tuned Radio Frequency (TRF) receiver. Heterodyning means mixing two frequencies together in a nonlinear device. Essentially, there are five sections in

a superheterodyne receiver: RF section, mixer, IF section, detector, and audio amplifier as shown in Fig. 18.12.

The primary functions of RF amplifier are selecting, band limiting and amplifying the received RF signals. RF amplifier provides some gain at a point of low noise in the receiver. A cheap receiver may not have RF amplifier section. The difference will show up in the performance in areas where signal strength is weak. An RF amplifier also determines the sensitivity of the receiver. Several advantages of including an RF amplifier in a receiver are better Signal-to-Noise (SNR) ratio, improved image frequency rejection and better selectivity.

The mixer stage performs heterodyning operation, which is translation or shifting on the frequency scale. This is a nonlinear device and its purpose is to convert the received RF frequencies to Intermediate Frequency (IF), which is the frequency that lies somewhere in between the RF and information frequencies, hence the name *intermediate*. Mixer includes a local oscillator. The most common Intermediate Frequency used in an AM Broadcast Band (550 kHz to 1650 kHz) receiver is 455 kHz.

The IF amplifier is a tuned amplifier, which is tuned at 455 kHz. Thus, this amplifier selects only 455 kHz and all other frequencies are rejected (or attenuated severely). A tuned amplifier can give good gain up to 30 dB to 40 dB. The IF amplifier stage gives about two-third or more of total receiver gain. The bandwidth of IF stage is ±5 kHz (10 kHz) for an AM Double Side Band (AM DSB) broadcast receiver. Most of the receiver gain and selectivity is achieved in the IF section. It is important to note that although the carrier frequency changes in this stage but the bandwidth remains unchanged. Thus the original information contained in the envelop remains unchanged.

The purpose of the detector is to extract the original information from the IF signal. Mostly, an envelop detector is preferred for this purpose because it is simple and cheap. The audio amplifier is a power amplifier. An audio amplifier amplifies the recovered information signal such that it can drive the loudspeaker.

Receiver Parameters

Important receiver characteristics are the following:

✦ Selectivity

Selectivity is the ability of a receiver to accept the desired signal frequency while rejecting all adjacent disturbances (undesired signals). The better the receiver's ability to exclude unwanted signals, the better its selectivity. It is usually poor at high frequency. The degree of selectivity is determined by the sharpness of resonance to which the frequency-determining components (band-pass filters) have been tuned. As the frequency to which the receiver is tuned is approached, the input level required to maintain a given output will fall. As the tuned frequency is passed, the input level will rise. Input levels are then plotted against frequency as shown in Fig. 18.13. The steepness of the curve at the tuned frequency indicates the selectivity of the receiver. Typically, this may be expressed as –6 dB at ±2.5 kHz of centre frequency and –60 dB at ±7.5 kHz of centre frequency, and –60 dB at 7.5 kHz of centre frequency.

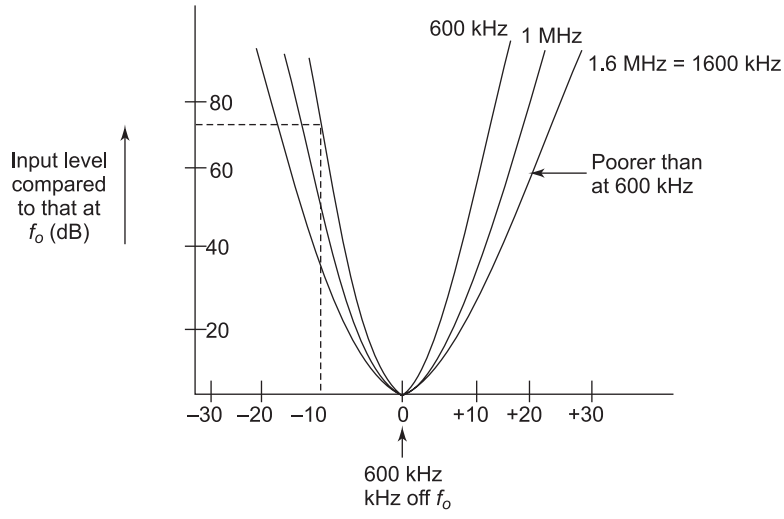


Fig. 18.13 Selectivity of a typical broadcast receiver

✦ Sensitivity

Sensitivity is the ability to amplify weak signals or we can say that it is a measure of the receiver's ability to reproduce very weak signals. The weaker the signals that can be produce a certain signal-to-noise (S/N) ratio, the better that receiver's sensitivity rating. Sensitivity is specified as the signal strength in volt/m. Sensitivity is the input signal at antenna that will generate a signal at detector output with the desired SNR (say 20 dB). That is 20 dB above noise power at detector output.

✦ Fidelity

Fidelity is a receiver's ability to reproduce the input signal accurately. Generally, the broader the band-pass, the greater the fidelity. Good selectivity requires a narrow band-pass. Good fidelity requires a wider band-pass to amplify the outer-most frequencies of the sidebands. Hence, most receivers make a compromise between good selectivity and high fidelity.

18.5 | AUTOMATIC GAIN CONTROL (AGC) CIRCUIT

The dc component of the output signal of the detector is feedback to IF amplifier and also to RF amplifier as negative feedback. The received signal at the receiver can be of the following three types: weak signal, medium signal, strong signal. Weaker signal at the receiver antenna provides lesser dc voltage from the detector output which is fed back to the IF stage. So there is more gain of IF and RF stage. Thus, more output of the receiver. On the other hand, stronger signal at the receiver antenna provides more dc voltage from the detector output, which is feedback to the IF stage. So there is lesser gain of IF and RF stage and thus, lesser output of the receiver. That is, the AGC circuit automatically increases the receiver gain for weak RF

input levels and automatically decreases the receiver gain when a strong RF signal is received. Thus, this circuit smoothens the input signal variations and keeps the output power almost constant.

18.6 DIGITAL COMMUNICATION

Many of the signals involved in today's communication are of digital nature (binary data). Digitising a signal often results in improved transmission quality and great accuracy. Figure 18.14 shows a digital communication system.

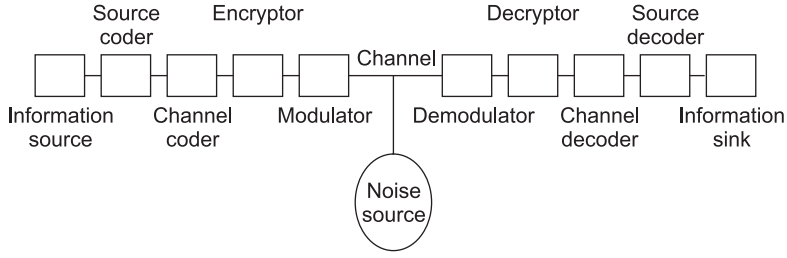


Fig. 18.14 Schematic diagram of a typical digital communication system

Source produces a stream of symbols. The source encoder removes the redundancy from the output of the source and performs data compression or source coding. Channel coder adds some redundancy again so as to achieve reliable communication on a noisy channel. Encryptor encodes message as per the chosen algorithm. The modulator prepares a signal for transmission on the physical channel. The channel adds some noise to the signal passing through it. Decryptor and decoder perform the inverse operations corresponding to the encryptor and modulator. The channel decoder performs error detection and correction. The source decoder decompresses the received data and presents the information to the sink.

Digital systems are not immune to noise and distortion, but it is possible to overcome their effect. The main advantage of digital communication is convenience in multiplexing and switching. Time-division multiplexing is quite easy with digital signals and different types of signals can be multiplexed together on the same channel. The bottleneck of the digital communication systems are the complexity and the larger bandwidth requirement.

18.6.1 Pulse Modulation

For transmitting an analog signal by means of a digital signal, the analog signals need to be sampled at least at twice the maximum frequency signal present in it. Then each sample can be expressed as a binary number for transmission. At the receiver side, the transmitted samples can be reconstructed to form a original signal. Figure 18.15 explains the simplest type of sampling, which is a multiplication of analog signal by train of pulses of having amplitude of 1 V. The output pulses will be equal to the amplitude of the analog signal at sampling times. The sampling rate must meet the condition of the sampling theorem: $f_s > 2 f_u$; f_u being

the upper limit of the signal bandwidth. This condition is easily met by modern electronic sampling hardware.

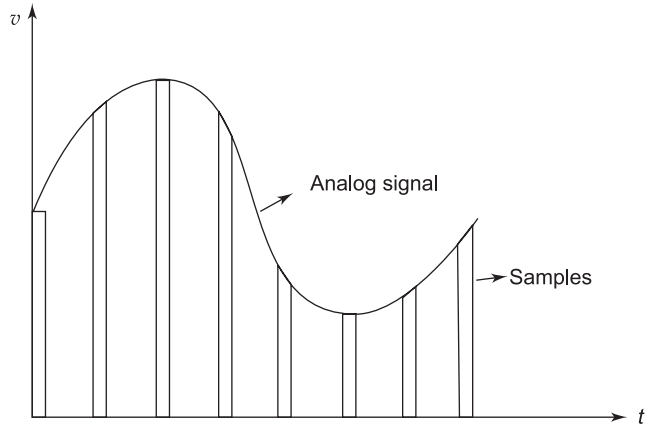


Fig. 18.15 Sampling

18.6.2 Analog Pulse Modulation

Sampling is one step of digital communication. The sampled signal in Fig. 18.15 is an example of *pulse amplitude modulation (PAM)* technique. The amplitude of the sampled signal is proportional to the amplitude of the analog signal at the instant at which it is sampled. The other popular pulse modulation technique is *Pulse Duration Modulation (PDM).* Here, the pulse duration is modified according to the amplitude of analog signal at that instant. PAM used in the high power audio amplifiers is used to modulate AM transmitters and telemetry systems. The PDM is demonstrated in Fig. 18.16.

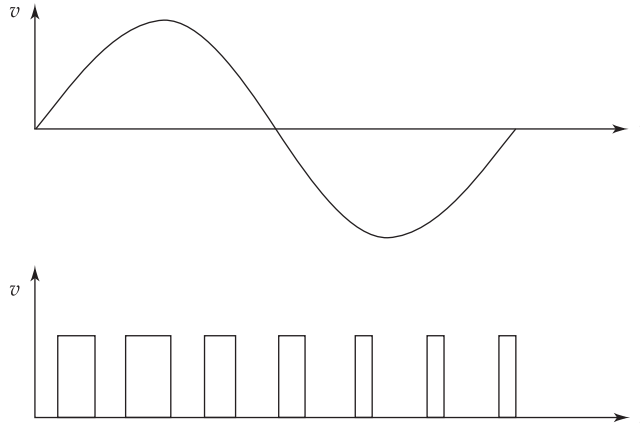


Fig. 18.16 Pulse duration modulation (PDM)

Another type of pulse modulation is *Pulse Position Modulation (PPM).* The amplitude and width of the pulse is same here, but their timings vary with the amplitude of original analog signal. It is also used in telemetry systems.

Pulse Code Modulation is the most commonly used digital modulation scheme. In this, the signal voltages are divided in to levels and each is assigned binary number representing the level close to its amplitude. And this number is transmitted in serial/parallel form. The number of levels available depends on the number of bits used to express the sample value and it is given by $N = 2^m$; N is the number of levels and m is the number of bits per sample. It is also called *quantising*.

18.6.3 Coding and Decoding

The process of converting an analog signal into binary numbers is called coding and the inverse operation, reconstructing the original analog signal from digital signal is called decoding. Both are accomplished in a single integrated circuit device called a *codec*.

18.7 MULTIPLEXING

Multiplexing is a process where multiple analog message signals or digital data streams are combined into one signal and transmitted via transmission medium or simply air. There are two types of multiplexing, *Frequency Division Multiplexing (FDM)* and *Time Division Multiplexing (TDM)*. FDM is in analog form, it means different frequency signals are mixed together and transmitted through a transmitter; in the receiver sides, filter is used to separate desired signal from the group of frequency signals. TDM is used mainly in digital communication. In TDM, each information source is allowed to use all the available bandwidth for a particular period or interval. Many signals can be sent on one channel by sending a sample from each source in rotation. TDM is extensively used in telephony.

18.8 PULSE DEMODULATION

A pulse-modulated signal in radar may be detected by a simple circuit that detects the presence of RF energy. Circuits that are capable of doing this were covered in this chapter in the Sec. 18.3 (detection discussion); therefore, the information will not be repeated here. A radar detector, in its simplest form, must be capable of producing an output when RF energy (reflected from a target) is present at its input.

In communications pulse detectors the modulated waveform must be restored to its original form. There are three basic steps of pulse demodulation: peak detection, low-pass filter and conversion. We need to study only peak detection which is the basic step.

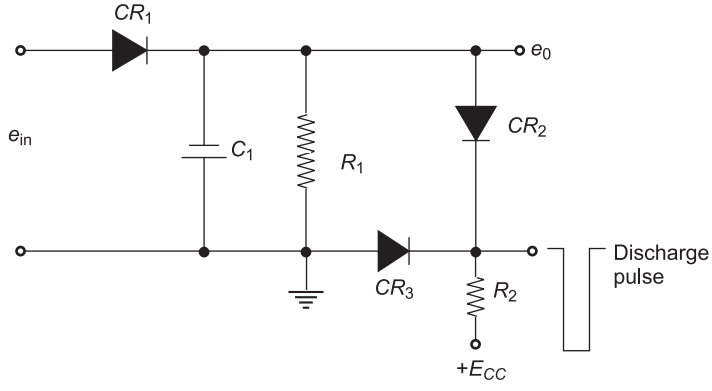
✦ Peak Detection

Peak detection uses the amplitude of a pulse-amplitude modulated (PAM) signal or the duration of a pulse-duration modulated (PDM) signal to charge a holding capacitor and restore

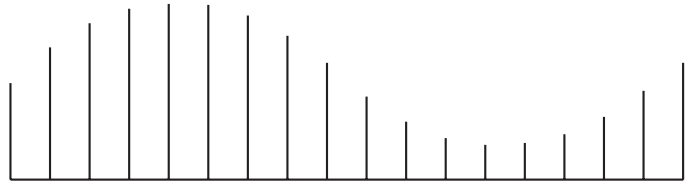
the original waveform. The demodulated waveform will contain some distortion because the output wave is not a pure sine wave. However, this distortion is not serious enough to prevent the use of peak detection.

✦ Pulse-Amplitude Demodulation

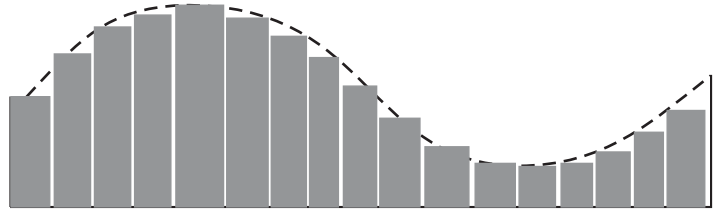
Peak detection is used to detect PAM. Figure 18.17(a) includes a simplified circuit (*a*) for this demodulator and its waveforms (*b* and *c*). CR_1 is the input diode, which allows capacitor C_1 to charge to the peak value of the PAM input pulse. PAM input pulses are shown in Fig. 18.17(b). The CR_1 is reverse biased between input pulses to isolate the detector circuit from the input. CR_2 and CR_3 are biased so that they are normally non-conducting. The discharge path for the capacitor is through the resistor R_1. These components are chosen so that their time constant is at least 10 times the inter-pulse period (time between pulses). This maintains the charge on C_1 between



(a) Circuit to peak detector



(b) Amplitude modulated pulses



(c) Peak detection

Fig. 18.17 Pulse-amplitude detection

pulses by allowing only a small discharge before the next pulse is applied. The capacitor is discharged just prior to each input pulse to allow the output voltage to follow the peak value of the input pulses. This discharge is through CR_2 and CR_3. These diodes are turned on by a negative pulse from a source that is time-synchronous with the timing-pulse train at the transmitter. Diode CR_3 ensures that the output voltage is near 0 during this discharge period. Figure 18.16 shows the output wave shape for this circuit. The peaks of the output signal follow very closely the original modulating wave, as shown by the dotted line. With additional filtering, this stepped waveform closely approximates its original shape.

Pulse-Duration Modulation

The peak detector circuit may also be used for PDM. To detect PDM, you must modify (*a*) of Fig. 18.17 so that the time constant for charging C_1 through CR_1 is at least 10 times the maximum received pulse width. This may be done by adding a resistor in series with the cathode or anode circuit of CR_1. The amplitude of the voltage to which C_1 charges, before being discharged by the negative pulse, will be directly proportional to the input pulse width. A longer pulse width allows C_1 to charge to a higher potential than a short pulse. This charge is held, because of the long time constant of R_1 and C_1, until the discharge pulse is applied to diodes CR_2 and CR_3 just prior to the next incoming pulse. These charges across C_1 result in a wave shape similar to the output shown for PAM detection in (c) of Fig. 18.17.

18.9 | THE TELEPHONE SYSTEMS

The public switched telephone system is the largest used communication system in the world. The telephone system is public in the sense that anyone can connect to it and is possible for anyone to communicate with anyone else. This is the basic difference from broadcasting system (radio and television) to private communication networks. In this, many interesting developments came like fiber optics and digital signal transmission, but it remains, in many ways, consistent with its origins. That is telephone signalling systems and the voltage and current levels that are found in subscriber lines are still unchanged. Also the compatibility is maintained in most areas of the system, so that simple dial-type telephones can connect with modern data-communication equipment with little change (without change also). Now the telephone network has been adapted to data communication, facsimile and even video. Figure 18.18 shows the basic topology of a typical switched telephone system.

Each subscriber is connected to a central office, where connection to other intended subscriber is made. Central office represents one exchange, the subscribers under each central office will have same three digits. Subscribers connected to the same central office can communicate with each other by means of the central office switch which can connect any line to any other line. The central office themselves connected together by trunk line, so that interconnection between the customer from different central offices is possible. Tandom Office is the connection medium when there are no trunk-line connections between two central offices. Normally, each individual subscriber telephone is connected to the central office by a single twisted pair of wires. The wires are twisted to help cancel the interference between circuits known

as crosstalk. Dialing can be accomplished one of the ways. The old fashioned rotary dial functions by breaking the loop circuit at a 10 Hz rate, with the number of interruptions equal to the number dialed. It is called *pulse dialing*. The second is to transmit a combination of two tones for each number. This is known as Dual Tone Multi-Frequency (DTMF) dialing, and it is commonly referred as touch tone dialing.

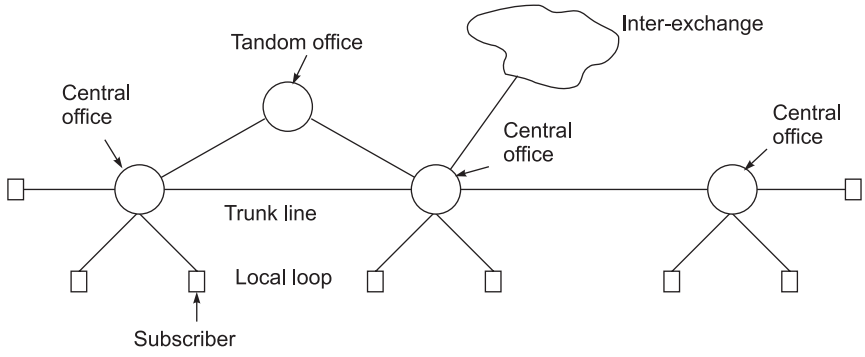


Fig. 18.18 Typical telephone network

Finally, telephone instrument use carbon microphones (transmitters) and magnetic earphones (receivers). Carbon microphones have the advantages of simplicity and the ability to generate a relatively large signal voltage without amplification. In modern telephones, carbon microphones have been replaced by electric condenser microphones for improving audio quality. The signal twisted pair line is required to carry both sides of the conversation simultaneously, thus providing full duplex communication. This could be accomplished by simply connecting both transmitters and both receivers in series.

18.10 | DATA TRANSMISSION

Transmission may be a serial, where single channel is used to transmit bit by bit, or parallel transmission, in which multiple channels are used to transmit several bits simultaneously. A five-bit "Baud of code" has the advantage of faster data transfer for a given bit rate, compared to codes with more bits per character. It is useful for low-data-rate channels like HF radio. The most common code for communication between micro computers is known as ASCII, which stands for American standard code for information interchange. ASCII is a seven-bit code, so it allows 128 possible ($2^7 = 128$) combinations without shifting.

18.10.1 Asynchronous Transmission

There must be a standard clock speed, and the transmitter and receiver clocks must be in phase with each other, so that the receiver checks the level of the line at the correct times. Data communication schemes may be synchronous or asynchronous depending on how the timing and framing information is transmitted. In asynchronous the frame is based on

a single character; while that for synchronous, the frame is much longer block of data. In microcomputers, asynchronous communication is most commonly used. In asynchronous communication, the transmit and receive clocks are free running and set at the same speed. A start bit is transmitted at the beginning of each character, and at least one stop bit is sent at the end of the character. In an asynchronous communication system, the conversion between parallel and serial form is usually performed by an integrated circuit called a Universal Asynchronous Receiver Transmitter (UART). At the transmitting end, a UART converts parallel data from a computer or terminal into serial form, adds start and stop bits, and clocks it out of the correct rate. In receiver side, a UART detect the start bit transition at the beginning of the character, before the receiver and transmitter clocks are in phase. After the start bit has been identified, the signal is sampled once per bit time, approximately in the centre of the bit. Then the UART will assemble the received character in a shift register then transfer it to another register for access by the computer.

18.10.2 Synchronous Communication

Here, the transmitter and receiver are synchronised to the same clock frequency. It is more efficient than asynchronous, because start and stop bits are not necessary. However synchronous communication needs more complex hardware and software. It is mainly used in higher speed communication typical of mainframe computers.

18.10.3 Data Compression

The data consists largely of alphanumeric characters, it is possible that some letters occur more frequently than others. Rather than encoding all letters with the same number of bits as in the ASCII, the more common letters could be encoded with few bits than the less common letters. This technique is called *Huffman coding*. The other data-compression technique is *run-length encoding*. It uses the advantage of the fact that information bits are often repeated.

18.10.4 Encryption

Encryption is an attempt to conceal the information from unintended persons. It means the way in which the data is encoded is not public. So encryption is used when data must be kept secret from unauthorised persons. There are several types of encryptions, the two main types are—private key and public key. *Private-key method* is more convenient. A long binary number is combined with the message data according to an algorithm that is designed, before the data is transmitted. At the receiver side, the same private key is combined with the received data to reveal the message. In public key encryption technical, recipient issues a public key that can be used by anyone to encode messages for that recipient. Messages can only be encrypted, not decrypted, using the public key. The recipient needs a separate, private key to decode the message. That key does not have to be divulged to anyone.

18.11 | DIGITAL MODULATION

Digital signals have become very important in communication. So, to send a digital signals by radio, it is necessary to use a higher frequency carrier wave, just as for analog communication. For the sine-wave high-frequency carrier signals, the same three basic parameters are available for modulation amplitude, frequency and phase. Here, the modulator and demodulator collectively described as a *modem*. We have frequency shift keying (FSK), phase shift keying (PSK) and rarely used amplitude shift keying (ASK). The simplest digital modulation scheme is FSK. In this, two frequencies are transmitted—one corresponding to binary one, the other to zero. In PSK, the phase of the carrier is shifted by 180° if the bits are changed from one to zero or zero to one. If two continuous bits are same, like 11/00 then no phase shift in the carrier frequency signal. In amplitude shift keying 1 is represented by constant amplitude carrier and 0 by absence of carrier (no carrier transmission). Mathematically the pulse signals ($s(t)$) are.

$$\begin{aligned} s(t) &= \text{A sin } \omega_c t, \text{ for pulse period (1)} \\ &= 0, \text{ for pulse period (0)} \end{aligned} \tag{18.31}$$

The digital data has physical form of voltage pulses, HIGH for 1 and LOW for 0. All the pulses have uniform time duration and the pulse rate is quite high. In order to transmit the digital data on a channel, it has to be modulated on a carrier of frequency according to the channel form–telephone line, wireless.

Further details will follow in Section 18.6.

18.12 | MULTIPLEXING AND MULTI-ACCESS

Most of the communication systems require sharing of channels. When all the signals passing through a given channel originate from the same source, it is called multiplexing of signals. Instead, the signals from several different sources combined on a signal channel is called *multiple-access system*. Multiple-access technique often happens in radio communication. The frequency spectrum can be divided up and part of it allocated to different users on a full-time basis. This is *Frequency Division Multiplexing (FDM)* or *Frequency Division Multiple Access (FDMA).*

On the other hand, the whole of the available spectrum can be allocated to each user for part of the time. This is *Time Division Multiplexing (TDM)* or *Time Division Multiple Access (TDMA).* The third form of multiple access is called *Code Division Multiple Access (CDMA).* Spread spectrum communication allows the multiplexing of signals from different sources CDMA. Here, each transmitter to be assigned a different *Pseudo Noise (PN)* sequence. If possible, an orthogonal sequence should be chosen, that is, the transmitter should never be in the same place at the same time. The PN sequence for the transmitter is given only to the receiver that is to operate with that transmitter, so that the receiver will then receive only the correct transmission and all the other receivers will ignore these signals.

18.13 | TRANSMISSION LINES

The signal is sent from the transmitter to receiver by means of a channel. If it is wired, communication transmission lines are the medium to take the signals. The transmission line includes metallic cable and optical fibre. Almost any configuration of two or more conductors can operate as a transmission line. Figure 18.19 shows a coaxial cable, in which the two conductors are connected and are separated by an insulating dielectric, which may be solid or air with a single helical spacer. Coaxial cables are referred to as *unbalanced lines* because they are unsymmetrical with respect to ground. Instead parallel lines are usually operated as balanced lines. Figure 18.18(b) shows parallel line cables.

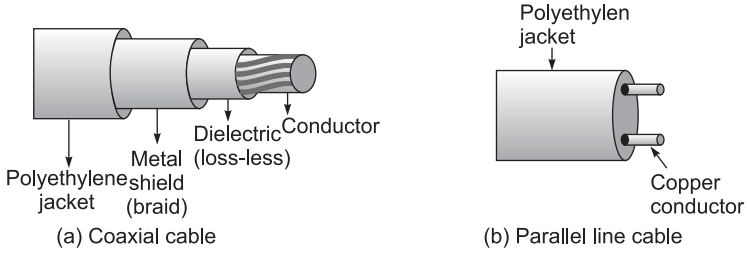


Fig. 18.19 Transmission lines

In parallel line cables, two conductors are separated by a thin ribbon of plastic, but air actually forms a good part of the dielectric. Parallel line cables may be an open-wire-type also. Twisted pairs of wires are often used as transmission lines for relatively low frequencies, because of their low cost. Transmission lines are more complex in their behaviour as frequency increases. Besides resistance in the conductors, inductive and capacitive effects become important with higher frequencies and longer lines. The higher the frequency, the larger the series inductive reactance, and the lower the parallel capacitive reactance. The circuit model of a short line is drawn in Fig. 18.20. For a long line, distributive effects have to be accounted for. Its equivalent circuit can be found in this form.

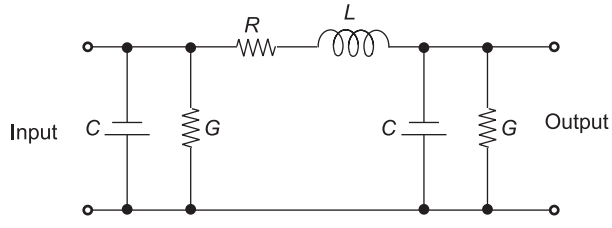


Fig. 18.20 Model of a transmission line for unbalanced case (short line)

Optical Fibre Cables

Optical fibre cable consists of multiple optical fibres that are coated with protective layer of plastic and arranged in a specific pattern. The different arrangements of optical fibres can be

done as per the applications of cable (whether underground or overhead), data transmission required, environmental conditions, distance over which a cable is to be laid, etc. A typical optical fibre cable and the fibre structure are shown in Fig. 18.21(a) and 18.21(b) respectively.

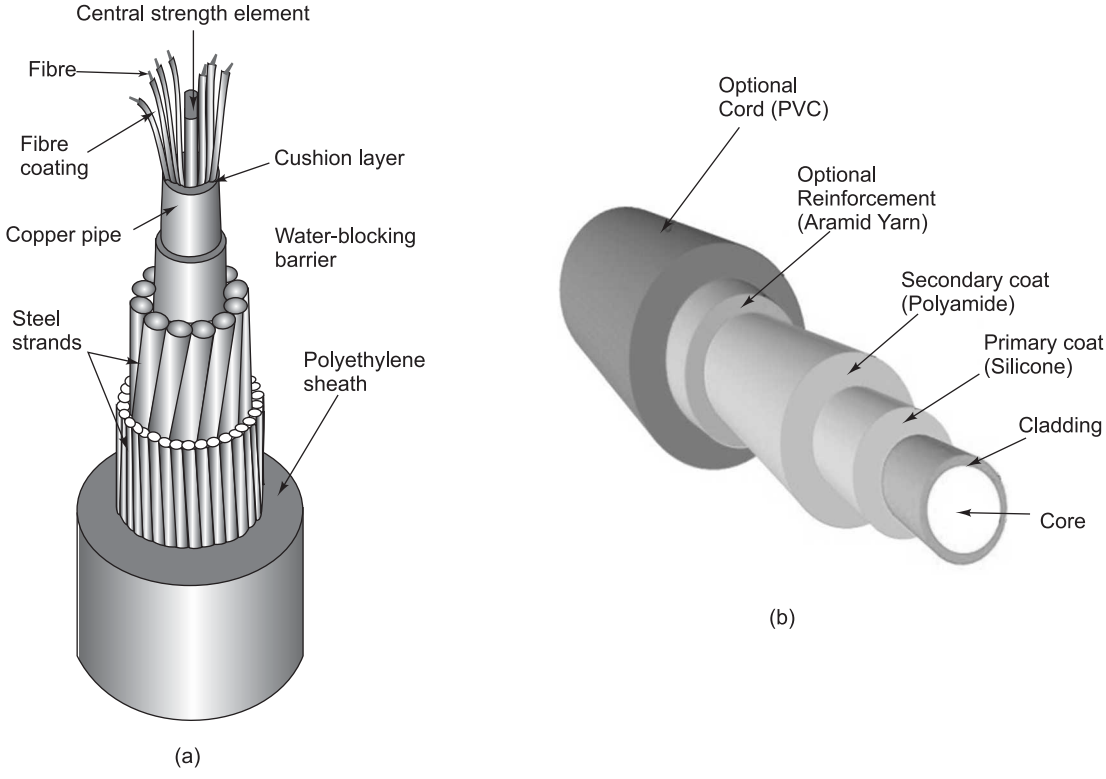


Fig. 18.21 (a) Optical fibre cable, (b) Optical fibre

In an optical fibre, there exists core surrounded by cladding structure. The refractive index of core is bit higher than cladding so as to improve the guiding characteristics of the fibre. Two layer coating of polyamide and silicon is done over cladding structure for its protection and is often placed in tight PVC jacket enclosing a yarn layer. The multiple fibres are placed in the cable, arranged in circular pattern around the central strengthening element. For indoor applications, where the probability of wear and tear is almost negligible, the cable may directly be placed in soft plastic layer. In order to protect the cable from wear and tear when placed in harsh conditions, it may be enclosed in copper pipe surrounded by water blocking barrier and polyethylene sheath. Multiple steel strands may exist to further mechanically strengthen the cable.

18.14 RADIO WAVES

Radio waves are one form of electromagnetic radiation, and the other forms are infrared, visible light, ultraviolet, X-rays and gamma rays. The radio waves that propagate are known as *Transverse ElectroMagnetic (TEM)* waves, which means that the electric field, the magnetic field and the direction of travel of the wave are all mutually perpendicular. Electromagnetic radiation can be generated by different means, but all of them involve the movement of electrical charges. In the case of radio waves, the charges are electrons moving in a conductor or set of conductors called an antenna. Once the electromagnetic waves are launched, they can travel through free space and through many materials. The speed of propagation of EM waves in free space is the same as that of light. Freespace propagation is also of interest in satellite communication.

18.15 ANTENNAS

An antenna is the interface between a transmitter and transmission media and is thus a very important part of the communication path. The antenna is a passive device, the power radiated by a transmitting antenna cannot be greater than the power entering from the transmitter. The antenna gain in one direction results from a concentration of power and is accompanied by a loss in other directions. The term *active antenna* simply describes the combination of a receiving antenna with a low noise preamplifier. The antenna will do both transmission and receive with same gain. The conductors in a transmitting antenna must be sized to handle large currents. The task of the transmitting antenna is to convert the electrical energy travelling along a transmission line into electromagnetic waves in space. The energy in the transmission line is contained in the electric field between the conductors and in the magnetic field surrounding them. At the receiving antenna, the electric and magnetic fields in space cause current to flow in the conductors that wake up the antenna. Some of the energy is thereby transferred from these fields to the transmission line connected to the receiving antenna. The isotropic radiator, five-eight wavelength antenna, discone, helical, monopole, ground-plane antenna, loop antenna and half-wave dipole are the common antennas.

18.16 TELEVISION

The television is a very important part of communication systems. Television video systems form pictures by a scanning process. The image is divided into a number of horizontal lines, which are traced out in synchrony at the camera and receiver. The more the number of lines, the more the resolution in the vertical direction. The North American standard uses 525 lines and 625 is used in Europe. In order to make the image visible all at once, rather than as a series of consecutively drawn lines the process must be completed quickly. The more quickly the images follow one another, the less flicker is present. The images/frames sent at the rate of 30 per second to have good response. Frame rates of 25 or 30 Hz cause noticeable flicker; to

reduce this, interlaced scan is used. It involves transmitting alternate lines of the picture, then returning and filling in the missing lines. As Fig. 18.22 shows, the picture is scanned from left to right and from top to bottom. The electron beam that traces the picture is blanked during the time intervals in which the beam retraces its path, from right to left and from bottom to top.

The ratio of width to height as called *aspect ratio* of the image, and its value is 4:3. For 30 Hz frame rate and 525 lines, the transmission rate is 525 × 30 = 151750 Hz. Good colour reproduction can be achieved by mixing three primary colours: red, green and blue. So, it is necessary to transmit information for three colours: red, green and blue. Thus considerably more information must be transmitted for colour television than for monochrome. Colour television requires three electron guns in the picture tube. The signal which is sent, with picture information along with synchronising pulses, is called composite video signal. There are three colour standards mainly used—NTSC (National Television Systems Committee), PAL (Phase Alternation by Line) and SECAM (Sequential Colour and Memory).

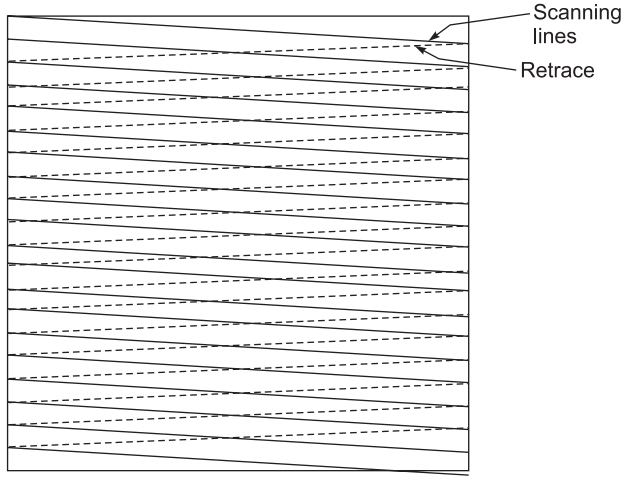


Fig. 18.22 Interlaced video scanning

Monochrome Cathode Ray Tube

The basic structure of a monochrome CRT is shown in Fig. 18.23. The electron gun emits a beam of electrons, the intensity of which is controlled by the video signal. Inside the electron gun a hot cathode emits electrons. The control grid maintains negative potential with respect to the cathode to accelerate and focusing electrons. As the video signal becomes more positive, the intensity of the electron beam is reduced. The G_2, G_3 and G_4 are used to accelerate and focus the electron beam on the screen.

The second anode is often called the *ultov*. Its connection is on the flared part of the tube, called the bell to avoid arcing, which could occur as high voltage is applied to the pins at the end of the tube. The electron beams strike a phosphor coating on the inside of the CRT face plate. In colour the cathode tube will have three electron beams from three electron guns arranged either in a triangle (delta gun) or

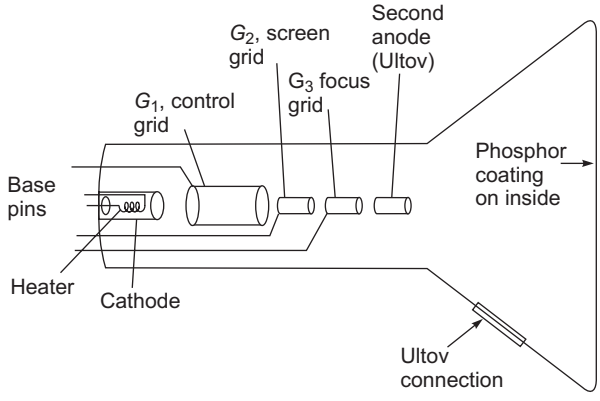


Fig. 18.23 Cathode ray tube

in a straight line (in-line guns). The three video signals are applied at the cathodes, one to each electron gun. The faceplate is covered with a dot pattern using three types of phosphors that glow red, green and blue when bombarded with electrons.

18.17 | SATELLITE COMMUNICATION

The first satellite was Echo, a metallised balloon 30 m in diameter that was launched by the United States in 1960 and orbited at a height of about 1600 km. For getting stronger return signals, a repeater must be put into the satellite. The signal path from the earth station transmitter to the satellite receiver is called the *uplink*, and the path from the satellite to the earth is known as the *downlink*. The choice of orbit is an important consideration in satellite-system design. *Geosynchronous orbit* is such an orbit that occupies a circular orbit above the equator at a distance of 35, 784 km above the earth's surface. At this height, the satellite orbital period is equal to the time taken by the earth to rotate once, that is 24 hours. If the direction of the satellite's motion is the same as that of earth's rotation, the satellite appears to remain almost stationary above one spot on the earth's surface. This is called *geostationary orbit*.

One disadvantage of geostationary satellite is the considerable amount of time it takes for a signal to make the round trip from the earth to the satellite and back. Signals received by the satellite are received at one frequency, amplified and moved to another frequency for transmission ray by an equipment called *transponder*. The transmit and receive frequencies are quite widely separated to avoid interference.

Earth stations vary vastly in design, depending on whether they are used to transmit to a satellite as well as to receive its signals, the type of signals in use (television, data, etc.), and the strength of the received signal. Intelsat satellites have antenna beams that cover nearly half the earth, though at a very low received power density. Domestic communication satellites often use spot beams, which cover only one country or part of a large country. Receiving stations for these satellites do not have to be as sophisticated as those used with Intelsat hemispheric beams. Applications of geostationary satellites, are namely television and radio broadcasting, telephony and data transmission.

Satellites in Low and Medium Earth Orbit (LEO, MEO) also possible to avoid high gain antenna and high powered transmitters in the case of geostationary orbit satellites. The problems of having satellites in LEO and MEO are (i) their position in space is not fixed with respect to a ground station, (ii) the annoying tendency of such satellites to disappear below the horizon, and (iii) Doppler effect. Low Earth Orbit (LEO) satellite are positioned from 300 to 1,500 km above the earth. Medium Earth Orbit (MEO) satellites are positioned about 8,000 to 20,000 km in altitude. The gap between LEO to MEO is to avoid the lower of the two Van Allen radiation belts that surround the earth; this radiation can damage the satellite, equipment.

Ref: *Electronic Communication Systems*, 2nd Edition by Thomson Delmar Blake.

18.18 PRINCIPLE OF OPERATIONS OF MOBILE PHONE

A cellular/mobile system provides standard telephone operation by full-duplex two-way radio at remote locations. It provides a wireless connection to the Public Switched Telephone Network (PSTN) from any user location within the radio range of the system.

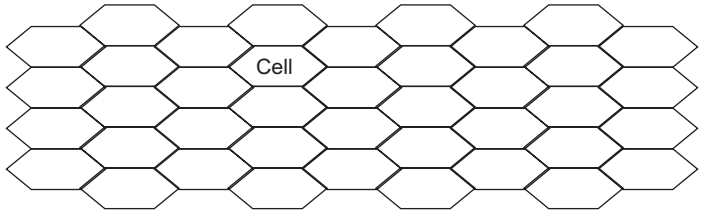


Fig. 18.24 Cell area

The basic concept behind the cellular radio system is that rather than serving a given geographical area within a single transmitter and receiver, the system divides the service area into many small areas known as cells as shown in Fig. 18.24. The typical cell covers only several square kilometres and contains its own receiver and low-power transmitter. The cell area shown in Fig. 18.24 is an ideal hexagon. However, in reality they will have circular or other geometric shapes. These areas may overlap and cells may be of different sizes. A basic cellular system consists of mobile stations, base stations and a *Mobile Switching Centre (MSC).* The MSC is also known as *Mobile Telephone Switching Office (MSTO).*

The MSTO controls the cells and provides the interface between each cell and the main telephone office. Each mobile communicates via radio with one of the base stations and may be handed off to any other base station throughout the duration of the call.

Each mobile station consists of a *transceiver*, an antenna and control circuit. The base station consists of several transmitters and receivers which simultaneously handle full-duplex communication and generally have towers which support several transmitting and receiving antennas. The base station serves as a bridge between all mobile users in the cell and connects the simultaneous mobile calls via telephone lines or microwave link to the MSC. The MSC co-ordinates the activities of all the base stations and connects the entire cellular system to the PSTN. Most cellular systems also provide a service known as *roaming*.

A simple block diagram representing working of mobile networks through GSM is shown in Fig. 18.25.

The cellular system operates in the 800–900 MHz range. The newer digital cellular system has even greater capacity. Some of these systems operate in 1.7–1.8 GHz bands.

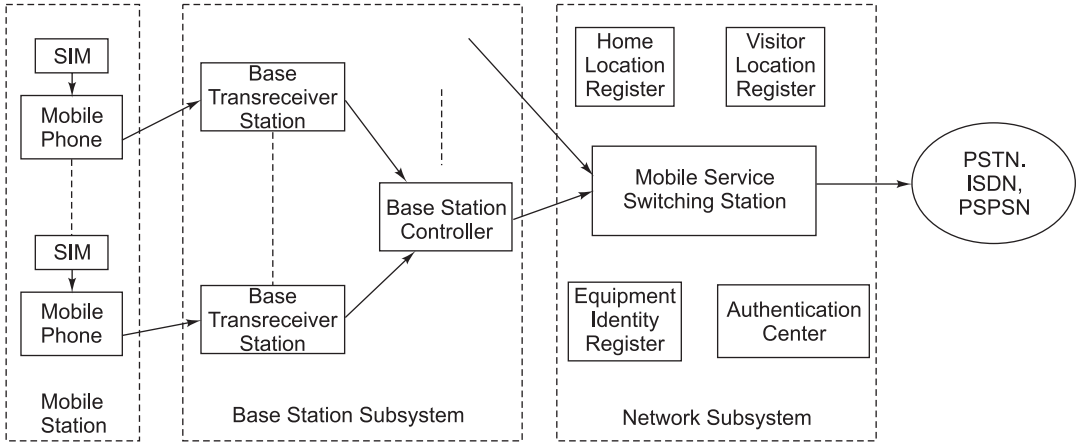


Fig. 18.25 Block diagram of GSM system

Cellular Telephone Unit

Figure 18.26 shows the block diagram of a cellular mobile radio unit. The unit consists of five major parts:

1. Transmitter
2. Receiver
3. Synthesizer
4. Logic unit
5. Control unit

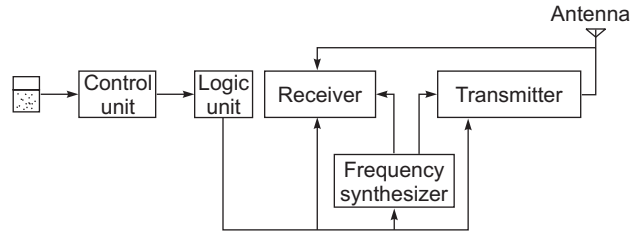


Fig. 18.26 Block diagram of a cellular mobile radio unit

❑ ***Transmitter*** It is the low power FM transmitter operating in a frequency range of 825 to 845 MHz. There is 66630 kHz transmit channel. Transmitter produces a deviation of ± 12 kHz. The modulated output is translated up to final transmitter frequency with the help of mixer, whose second input also comes from frequency synthesizer. The unique feature of high power translator is that output is controllable by the cell site & MTSO (mobile telephone switching office).

❑ ***Receiver*** The cellular receiver consists of RF amplifier, FM demodulator and filters. An RF amplifier boots the level of received cell site signal. Received signal is monitored by MTSO. If the signal is weak in the present cells then mobile unit is shifted to other site where the signal is strong.

❑ ***Frequency Synthesizer*** Frequency Synthesizer is used to generate various signals required for transmitter and receiver.

When a mobile unit initiates a call, MTSO identifies the user and assigns a frequency Channel which is not used by any other mobile in the cell. MTSO sends a unique code for setting channel frequencies.

❑ ***Logic Unit*** Logic unit is micro-processor controlled master control circuit for cellular radio. It basically controls the complete operation of MTSO and mobile unit.

❑ ***Control Unit*** The control unit is a set of speaker, microphone with touch tone dialling facility & it stores the memory like numbers and dialling features.

References: 1. C.W. Lee, *Mobile/Cellular Communication*
2. *www. itbloas. com/community*

18.19 | FAX

FAX is the short form for *facsimile*, also called telecopying, the telephonic transmission of scanned printed material (both text and images) normally to a telephone number connected to a printer or other output device.

✦ Principle of Operation

The original document is scanned with a fax machine, which processes the contents (text or images) as a single fixed graphic image, converting it into a bitmap. The information is then transmitted as electrical signals through the telephone system. The receiving fax machine reconverts the coded image, printing a paper copy.

18.20 | ISDN

The Integrated Services Digital network (ISDN) was developed by ITU-T in 1976. It is a set of protocols that combines digital telephony and data transport services. The whole idea is to digitise the telephone network to permit the transmission of audio, video and text over existing telephone lines.

✦ Services

The purpose of the ISDN is to provide fully integrated digital services to users. These services fall into three categories: bearer services, teleservices (Fig. 18.27a) and supplementary services. (Fig. 18.27b).

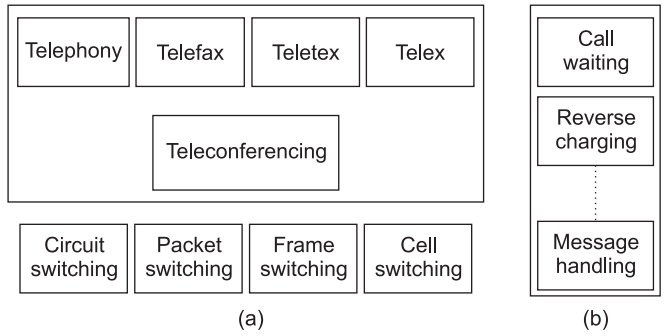


Fig. 18.27 (a) Teleservices and bearer services
(b) Supplementary services

❑ ***Bearer Services*** Bearer services provide the means to transfer information (voice, data and video) between users without the network manipulating the content of that information. The network does not need to process the information and, therefore, does not change the content.

❑ ***Teleservices*** In teleservicing, the network may change or process the content of the data.

Teleservices include telephony, telefax, videofax and teleconferencing. Although the ISDN defines these services by name, they have not yet become standards.

❑ ***Supplementary Services*** Supplementary services are those services that provide additional functionality to the bearer services and teleservices. Examples of these services are reverse charging, call waiting and message handling, all familiar from today's telephone company services.

Integrated Services Digital Network (ISDN) services are much more efficient and flexible than analog services. To receive the maximum benefit from the integrated digital networks, the next step is to replace the analog local loops with digital subscriber loops. Voice transmission can be digitised at the source, thereby removing the final need for analog carriers. In ISDN, all customer services will become digital rather than analog, and the flexibility offered by the new technology will allow customer services to be made available on demand. Most importantly, ISDN will allow all communication connections in a home or building to occur via a single interface.

Figure 18.28 gives a conceptual view of the connections between users and an ISDN central office. Each user is linked to the central office through a digital pipe. These pipes can be of different capacities to allow different rates of transmission and support different subscriber needs.

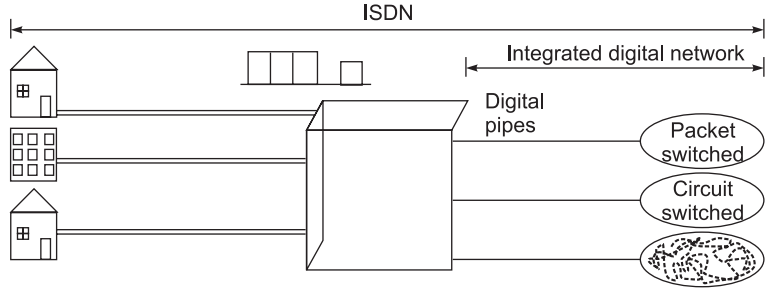


Fig. 18.28 ISDN

There are three types of ISDN channels:

1. Bearer (B Channels)
2. Data (D Channels)
3. Hybrid (H Channels)

❑ ***B Channels*** A bearer channel is defined at a rate of 64 kbps.

❑ ***D Channels*** A data channel can be either of 16 or 64 kbps, depending on the need of the users. Although the name says data, the primary function of a D Channel is to carry control signalling for the B Channels.

❑ ***H Channels*** Hybrid channels are available with data rates of 384 kbps, 1536 kbps, or 1920 kbps. These rates suit H Channels for high data rate applications such as video, teleconferencing and many more digital techniques.

18.21 MICROWAVE COMMUNICATION

Microwave communication is the transmission of signals via radio using a series of microwave towers. Microwave communication is known as a form of "Line of Sight" communication, because there must be nothing obstructing the transmission of data between these towers for signals to be properly sent and received.

✦ Microwave Communication History

Microwave communication technology was developed in the early 1940's by the Western union.

With the development of satellite and cellular technologies, microwave is now less widely used in the telecommunication industry.

✦ Means of Microwave Communication

Microwave communication takes place both by analog and digital formats. While digital is the most advanced form of microwave communication, both analog and digital methods pose certain benefits for the users.

✦ Economical Option

Analog microwave communication may be most economical for users at the tower site simply because it is already paid for and in service. If we are already operating a microwave component, it is most likely analog.

✦ Digital Microwave Options

Digital microwave communication utilises more advanced, more reliable technology. It is much easier to find equipment to support this transmission method because it is the newer form of microwave communication. Because it has a higher bandwidth, it also allows to transmit more data using more verbose protocols. The increased speeds will also decrease the time it takes to select your microwave site equipment.

✦ Frequencies Available for Microwave Communication

It covers the frequency range from UHF to EHF.

Bands UHF (Ultra High Frequency): 300 MHz to 3 GHz

EHF (Extremely High Frequency): 30 GHz to 300 GHz

18.22 OPTICAL FIBRE COMMUNICATION

- Fibre optics deals with the transmission of light through fibres of glass, plastic or other transparent materials and works on the principle of total internal reflection.
- In electronic communication, optical fibres are preferred over copper wires because they are extremely light, small and can be accommodated in a small space. By

using optical fibres, the number of signals that can be transmitted simultaneously is enhanced.

- After the availability of LASER in 1960, there has been fast growth in the field of fibre optics. Optical fibres and LASERs together increase the capacity of communication system by one lakh times the conventional system.
- In 1973, the Airborne Light Optical Fibre Technology (ALOFT) program replaced 302 cables, which weighed 40 kg by the fibre system weighing only 1.7 kg.
- By the late 1970s and early 1980s, every major telephone communication company was rapidly installing new and more efficient fibre system
- In optical region of electromagnetic spectrum, the transmission of information is carried out not by frequency modulation of the carrier, but by varying the intensity of optical power. This is known as *optical communication*.

18.22.1 Block Diagram of Optical Fibre Communication

The basic fibre optic transmission system consists of three basic elements (Fig. 18.29):

1. The optical transmitter
2. The fibre optic cable
3. The optical receiver

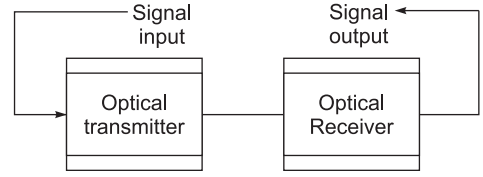


Fig. 18.29 The basic fibre optic communication system

The *transmitter* converts an electrical analog or digital signal into a corresponding optical signal. The source of optical signal can be either a Light Emitting Diode (LEDs) or a laser diode. The block diagram of optical fibre communication is given in Fig. 18.30.

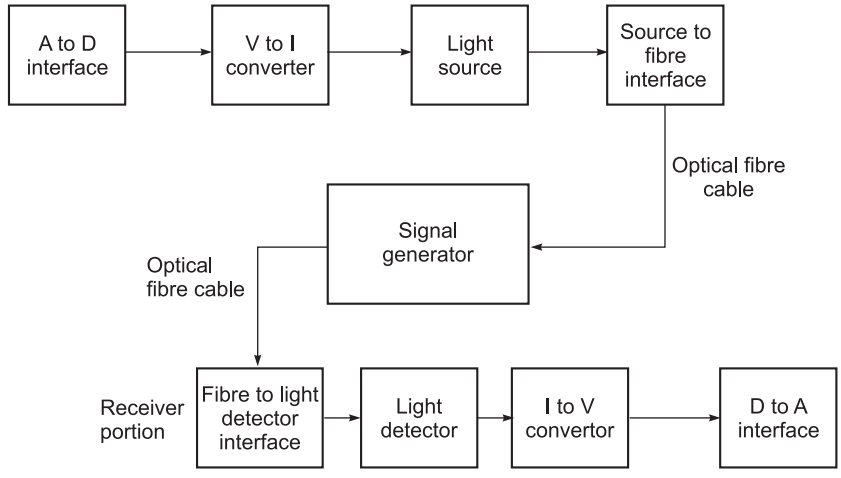


Fig. 18.30 Block diagram of optical fibre communication

The transmission medium in optical fibre communication system is an *optical fibre cable*. It is a very thin and flexible medium that guides light from an optical transmitter to an optical

receiver. The cable consists of one or more glass fibres, which acts as waveguides for the optical signal.

The *receiver* converts optical signals back into a replica of the original electrical signal. The detector of the optical signal is either a PIN-type photodiode or avalanche type photodiode.

18.22.2 Advantages of Optical Fibres

❑ ***Long Distance Transmission*** Optical fibres have lower transmission losses compared to copper wires. Data can be transmitted over longer distances.

❑ ***Large Information Capacity*** Optical fibres have wider bandwidth than copper wires. This property decreases the number of physical lines required for sending a given amount of information.

❑ ***Small Size and Low Weight*** The low weight and the small dimensions of fibre offers a distinct advantage over heavy wire cables in crowded underground city ducts or in ceiling mounted cable trays.

❑ ***Immunity to Electrical Interference*** An especially important feature of an optical fibre relates to the fact that it is a dielectric material which means it does not conduct electricity.

❑ ***Enhanced Safety*** Optical fibres offer a high degree of operational safety, since they do not have the problems of ground loops, sparks and potentially high voltages inherent in copper lines.

❑ ***Increased Signal Security*** The optical signal is well confined within the fibre and an opaque coating around the fibre absorbs any signal emission.

18.22.3 Application of Optical Fibre Communication

1. Optical Fibre Communication (OFC) proves efficient and cost effective in public network application.
2. Fibre optical cables can be used under the sea.
3. Public utility organisations like railways, TV transmissions etc. find tremendous use of OFC.
4. Colleges, universities, offices, industrial plants, etc. employ optical fibres within their LAN (Local Area Network) systems.
5. Fibre optic is widely used in telecommunication such as voice telephones, video phones, telegraph services, message services, etc.

SUMMARY

- ➢ Communication systems have been introduced.
- ➢ Various types of modulations discussed.
- ➢ Digital modulation and digital communication have been explained.
- ➢ Different types of communications have been dealtwith.

EXERCISES

➥ Review Questions

1. What are advantages of digital communication over analog communication?
2. What is the approximate frequency range of operation of cellular telephony?
3. What is an electromagnetic spectrum?
4. Give the different frequency bands and their corresponding wavelengths as defined by electromagnetic spectrum.
5. Give the applications of different radiations presented in electromagnetic spectrum.
6. Which types of modulation is resistant to channel noise—AM or PM? Explain why.
7. What is the difference between sensitivity and selectivity of a receiver?
8. A 2 mV (peak to peak) the signal varying between –1 mV to +1 mV is quantised using 64 levels. What is the number of bits required to represent one sample of the quantised signal?
9. In AM, what is the maximum value of modulation index. Why cannot it be more than that?
10. Explain how frequency modulation is accomplished using CVO.
11. Define modulation index of FM.
12. What is wideband FM?
13. Write the expression of the bandwidth of FM.
14. In TV signal transmission, why is AM is used for video signals and FM for audio signals?
15. Which modulated signal has higher bandwidth—AM or FM?
16. What are DSB–SC signals and SSB–SC signals? Compare the transmission channel bandwidth needed for these kinds of signals.
17. What is meant by base-band?
18. Sketch the spectrum of AM signal.
19. Write a brief account of VCO.
20. Draw the block diagram of PLL. How is FM demodulation achieved by PLL.
21. What conditions must be met by the time constant of the AM demodulation circuit?
22. Distinguish between asynchronous and synchronous data transmission.
23. Describe the function of a modem.
24. What is meant by encryption? What are the two main types of encryption.
25. In satellite communication what is a transponder?
26. Describe interlaced video screening. Relate it to the transmission rate.
27. In digital modulation by FSK, drawn the signal waveform for data 0100110. Also, draw the signal for PSK. Which is easier to implement?
28. Compare and distinguish AM demodulation and pulse demodulation.
29. What is a tank circuit. Describe the circuit operation.
30. Why super-heterodyne method is used in radio receivers?
31. Discuss the construction of optical fibre cable.
32. Why is the refractive index of core made higher than the cladding?

➥ Multiple-Choice Questions

1. The following is not the purpose of modulation:
   (a) Multiplexing
   (b) Effective radiation
   (c) Narrow banding
   (d) Increase in signal power

2. The modulation index of an AM wave is changed from 0 to 1. The transmitted power is
(a) halved (b) increased by 50% (c) quadrupled (d) unchanged

3. Which one is an advantage of AM over FM?
(a) FM has wide bandwidth
(b) Probability of noise spike generation is less in AM.
(c) FM has better fidelity
(d) FM is more immune to noise

4. The following is not an advantage of FM over AM
(a) Sputtering effect (b) Capture effect (c) Fidelity (d) Noise immunity

5. Which of the following modulations is digital in nature?
(a) PPM (b) PAM (c) DM (d) None of these

6. Quantisation noise occurs in
(a) PAM (b) PWM (c) DM (d) None of these

7. In a single-tone FM discriminator, (S/N) is
(a) proportional to square of deviation (b) inversely proportional to deviation
(c) proportional to cube of deviation (d) proportional to deviation.

8. A PAM signal can be detected by using
(a) an integrator (b) a band-pass filter
(c) a high-pass filter (d) low-pass filter

9. Flat-top sampling leads to
(a) an aperture effect (b) aliasing
(c) loss of signal (d) none of these

10. Pulse stuffing is used in
(a) synchronous TDM (b) any TDM
(c) asynchronous TDM (d) none of these

ANSWERS

✦ Multiple-Choice Questions

1. (d) **2.** (b) **3.** (a) **4.** (b) **5.** (c) **6.** (c) **7.** (a) **8.** (d) **9.** (a) **10.** (c)

CHAPTER 19

Measuring Instruments

GOALS AND OBJECTIVES

- ❏ Introduction of electrical and electronic instruments
- ❏ Classification of instruments and types of indicating instruments
- ❏ Functioning mechanism of millimeter or volt-ohm millimeter (VOM) and oscilloscope
- ❏ Frequency and phase measurement
- ❏ Digital instruments—resolution and sensitivity
- ❏ Oscilloscope

19.1 INTRODUCTION

Measurement normally involves an instrument as a physical means of determining a variable or quantity. An instrument is defined as a device for finding out the value or magnitude of a variable or quantity. Measurement is a means to achieve the final goal, i.e. instrumentation. The electronic instrument depends on electrical or electronic principles for its measurement function.

The measurement of a given quantity is nothing but the result of a comparison between the quantity and a predefined standard (direct method). In engineering applications, indirect methods are normally preferred. It consists of a transducer, which converts the quantity to be measured in an analogous form. This analog signal is then processed by some intermediate means and is fed to the final device, which finally gives the measurement result.

19.2 ELECTRICAL AND ELECTRONIC INSTRUMENTS

The elements used in such instruments are

(i) a detector,
(ii) an intermediate transfer device, and
(iii) an indicator, a recorder or a storage device.

Mechanical instruments, due to their high inertia and noise, are hardly used nowadays. Their application is restricted to measurement of a slowly varying pressure.

Electrical instruments depend on the mechanical movement of an indicating device having some inertia and thus have a limited time response (0.5–24 s). Nowadays, electronic instruments are used for fast responses required for most scientific and industrial measurements. They are used for the detection of electromagnetically produced signals such as radio, video and microwaves, space applications and computers.

Some important terms pertaining to 'measurement' are defined now.

❑ ***Instrument*** A device for finding the value or magnitude of a quantity or variable.

❑ ***Accuracy*** It tells us about the nearness of the measured value towards the true value, i.e. the measure of conformity to the true value.

❑ ***Precision*** It refers to the degree of agreement within a group of measurements or instruments, i.e. the measure of reproductivity. Precision has two characteristics: conformity, and the number of significant figures to which measurements may be made.

❑ ***Resolution*** It is defined as the smallest change in input that can be detected by an instrument.

❑ ***Sensitivity*** It is the ratio of output signal or response of the instrument to a change of input or measured variable.

❑ ***True Value*** (A_t) It is the average of the infinite number of measurements, when the average deviation tends to become zero.

❑ ***Error*** An error is a deviation from the true value of the measured variable.

19.2.1 Errors

No measurement can be made with perfect accuracy. There are three types of errors: gross, systematic and random. *Gross errors* are mainly human errors like misreading of instruments, incorrect adjustment, improper application of instruments and computational mistakes. *Systematic errors* have the same magnitude and sign for a given set of conditions. These

errors accumulate at the end of the measurement. *Random errors* are caused due to random variations in the parameter or the system of measurement. These errors result in the deviation of magnitude of the variable measured by the instrument from the true value of the variable. The difference in the measured value and the true measured value gives rise to static errors.

Absolute static error $\delta A = A_m - A_t$ (19.1)

where, A_m is the measured value, and
A_t is the true value.

❑ ***Relative Static Error*** $\varepsilon_r = \delta A/A_t = (A_m - A_t)/A_t$ (19.2)

❑ ***Static Error Correction*** $\delta C = A_t - A_m = \delta A$ (19.3)

✦ **Limiting Error**

In most indicating instruments, accuracy is guaranteed to a certain percentage of full-scale reading. The limits of these deviations from the specified value are known as limiting errors.

$$\delta A = A_a - A_s \quad (19.4)$$

where A_a is the actual measurement, and
A_s is the specified (nominal) value.

❑ ***Guarantee Error (ε_r)***

$$\varepsilon_r = \delta A/A_S \quad (19.5)$$

19.2.2 Instrument Efficiency

1. Ammeter

$$\eta_A = I_{\text{fsd}}/\text{power consumed} = 1/V_{\text{fsd}} \quad (19.6)$$

2. Voltmeter

$$\eta_{\text{v}} = V_{\text{fsd}}/\text{power consumed} = 1/I_{\text{fsd}} \quad (19.7)$$

where I_{fsd} is the full scale current (A), and
V_{fsd} is the full-scale voltage (V).

EXAMPLE 19.1 A 60 mV/120 mV dual-range millivoltmeter, when used to measure the voltage across two points in a dc circuit, gives a reading of 27.5—30 mV when 60 mV and 120 mV ranges, respectively, are employed. Assuming that the meter has been correctly calibrated, estimate the true value of the voltage existing across the two points in the dc circuit. It is known that the millivoltmeter has a sensitivity of 10 kΩ/volt.

Solution Let R_x be the resistance between the two points and I be the current through R_x.

For 60 mV range,

Resistance of millivoltmenter $(R_v) = 60\text{ mV} \times 10\text{ k}\Omega/\text{V} = 600\ \Omega$

$\therefore$ $27.5\text{ mV} = I \times 600\ R_x/(R_x + 600)$ (i)

For the 120 mV range,

Resistance of millivoltmeter $(R_v) = 120\text{ mV} \times 10\text{ k}\Omega/\text{V} = 1200\ \Omega$

$$\therefore \quad 30 \text{ mV} = I \times 1200\, R_x/(R_x + 1200) \qquad \text{(ii)}$$

Dividing Eq. (i) by (ii), we get

$$27.5/30 = (1200 + R_x)/2\,(600 + R_x)$$

Solving $R_x = 120\ \Omega$; $I = 0.275$ mA

$\therefore$ actual voltage $= R_x I = 120 \times 0.275 = 33$ mV

19.3 CLASSIFICATION OF INSTRUMENTS

Instruments are broadly divided into two classes:

✦ Absolute Instruments

These give the quantity to be measured in terms of an instrument constant and its deflection, e.g. tangent galvanometer.

✦ Secondary Instruments

These directly give the magnitude of the electrical quantity to be measured, e.g. ammeter, voltmeter. The principle of working of all electrical measuring instruments depends on the various effects of electric current or voltage. The effects utilised in the manufacturing of electrical instruments are magnetic, heating, chemical and electromagnetic in nature. Indicating instruments consist essentially of a pointer moving over a calibrated scale attached to the moving system pivoted on jewelled bearings.

19.3.1 Basic Requirements for Measurement

For a satisfactory working of indicating instruments, the following three types of torques are required:

✦ Deflection (Operating) Torque

It is necessary to make the moving system of the instrument (pointer) move from its zero position.

$$T_{dc} = N\,Bi\,ld = Gi \qquad (19.8)$$

$\therefore \quad T_{dc} \propto$ measurand

where N = number of turns of coil

B = flux density in the air gap at the coil position

l = length of vertical side of coil

d = length of horizontal side of coil

i = current through the coil

$G = NBld$ displacement constant of the galvanometer

✦ Controlling (Restoring) Torque

Some controlling force is employed either by a spring or gravity to limit the movement.

(a) A spring is generally used for controlling torque (phosphor bronze is the most suitable material).

$$T_c = k\theta \tag{19.9}$$

where θ is the deflection of the pointer and k is the controlling torque constant.

(b) Gravity method as shown in Fig. 19.1 is also sometimes used for controlling torque.

$$T_c \propto \sin\theta$$

✦ Damping Torque

This torque is necessary to avoid oscillation of the moving system about its final deflected position owing to the inertia of the moving parts and to bring the moving system to rest in its deflected position quickly.

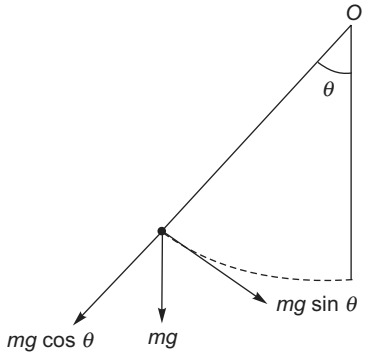


Fig. 19.1 Controlling torque due to gravity

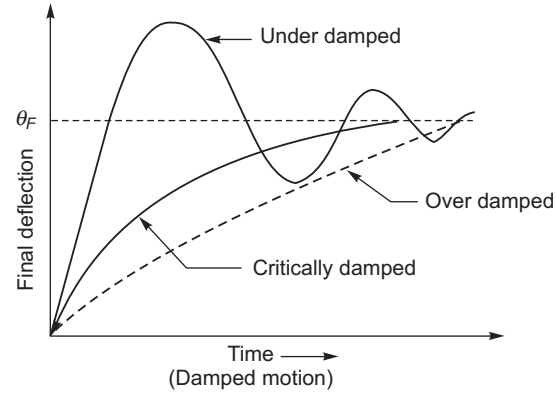


Fig. 19.2 Different cases of damping of an instrument

The various methods of obtaining damping are air friction, fluid friction and eddy current. Figure 19.2 shows the possible cases of damping of an instrument.

Damping torque $T_{dm} = D\,\mathrm{d}\theta/\mathrm{d}t$ (19.10)

where D is called the damping constant.

The restarting torque due to inertia of the moving system is

$$T_i = J(\mathrm{d}^2\theta/\mathrm{d}t^2) \tag{19.11}$$

where J is called the inertia constant.

✦ Equation of Damping Motion

The general differential equation of damping motion is

$$J(\mathrm{d}^2\theta/\mathrm{d}t^2) + D(\mathrm{d}\theta/\mathrm{dt}) + k\theta = Gi \tag{19.12}$$

Hence, the auxiliary is

$$Jm^2 + Dm + k = 0$$

$$\therefore \quad m = [-D \pm j\sqrt{(4kJ - D^2)}]/2J$$

Thus, deflection $\theta = A\, e^{m_1 t} + B\, e^{m_2 t}$

Under steady state,

$$\frac{d^2\theta}{dt^2} = 0,\ \mathrm{d}\theta/\mathrm{d}t = 0,\ \theta = \theta_f$$

$\therefore$ $\theta_f = Gi/k$ = final steady state deflection

Case I: If $D^2 < 4\,kJ$ then the response is underdamped

Case II: If $D^2 = 4\,kJ$ then the response is critically damped

Case III: If $D^2 > 4\,kJ$ then the response is overdamped

The angular frequency of damped oscillation is

$$\omega_d = \frac{\sqrt{4kJ - D^2}}{2J} \tag{19.13}$$

$\therefore$ $$R = \frac{G^2}{2\sqrt{(kJ)}} \tag{19.14}$$

where, R is the series resistance for critical damping.

19.3.2 Electromagnetic Damping

Electromagnetic damping is produced by the induced effects when the coil moves in the magnetic field and a closed path is provided for the currents to flow. Electromagnetic damping is because of

1. eddy currents produced in the metal core, and
2. current circulated in the coil circuit by emf generated in the coil when it rotates.

19.4 TYPES OF INDICATING INSTRUMENTS

Indicating instruments can be divided into different types according to their working principles.

19.4.1 D' Arsonval Movement

The basic Permanent Magnet Moving Coil (PMMC) mechanism is often called the D'Arsonval Movement, named after its inventor. PMMC instruments are accurate and suitable for dc measurements only.

When the current I passes through a coil, a deflecting torque is produced on the coil.

$$T_{dc} = K_1 I$$

where K_1 = constant = NBA
N = number of turns in a coil
B = flux density
A = area of the coil

The deflecting torque causes a restoring torque in the spring attached to the pointer, which is given by

$$T_c = K_2\theta$$

where θ is the angular deflection of the pointer and k_2 is the controlling torque constant.

For final steady deflection,

$$T_{dc} = T_c$$

$$K_1 I = K_2\theta$$

$$\therefore \quad \theta = (K_1/K_2)I \tag{19.15}$$

These instruments require very low power consumption and current for full-scale deflection.

19.4.2 Constructional Details

The permanent magnet (PM) has a semicircular spacing, which is fixed soft iron core to provide low reluctance flux path. The Moving Coil (MC) is placed in the annular space between PM and core. The coil of many turns is wound on a former as shown in Fig. 19.3(a). The coil is connected to a vertical spindle placed in low friction jewelled bearings. Two helical springs are connected to the spindle at one end and at the other end to a fixture. The springs thus counter the MC movement. The pointer *P* moves along a calibrated scale.

Note: Spindle, centre weight, helical springs points moving on a calibrated scale form part of all measuring instruments. The current to be measured is passed through the MC through flexible links.

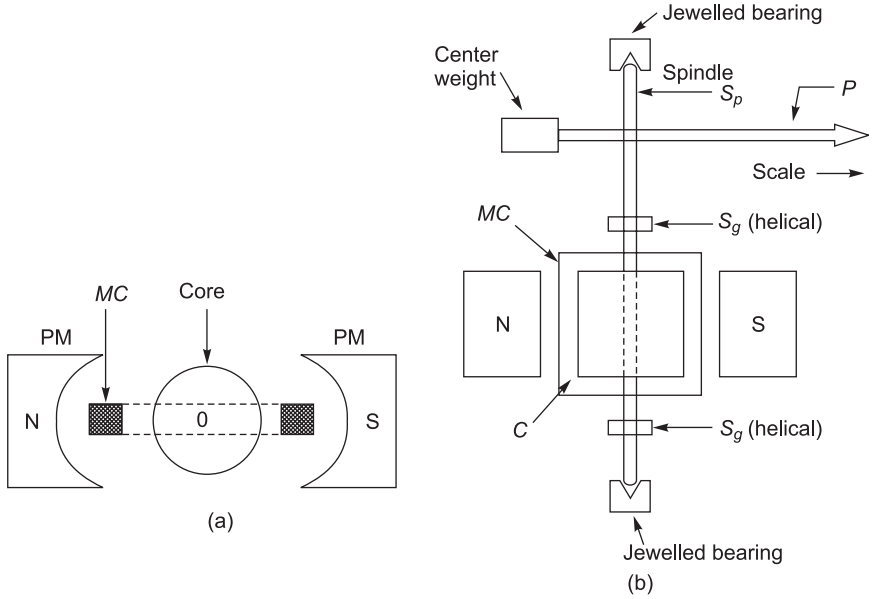


Fig. 19.3 PMMC instrument

✦ Galvanometer Sensitivity

The sensitivity of a galvanometer can be specified in terms of current sensitivity and voltage sensitivity.

Current sensitivity $S_I = d/I$ mm/μA (19.16)

where d is the deflection of the galvanometer scale divisions in mm and I is the galvanometer current in μA.

Voltage sensitivity $S_V = d/V$ mm/mV (19.17)

where V is the voltage applied to the galvanometer in mV.

✦ dc Ammeter

The coil winding of the PMMC movement is small and light and it can carry only small currents. For large current measurement, major part of the current is bypassed through a shunt resistance, R_{sh} (Fig. 19.4).

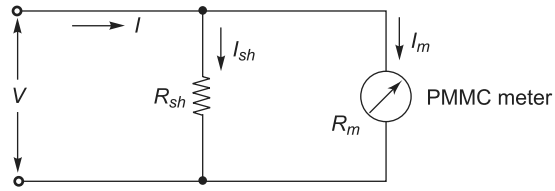


Fig. 19.4 dc ammeter

Clearly, $I_{sh} R_{sh} = I_m R_m$

$\therefore \quad R_{sh} = I_m R_m/(I - I_m)$

But $I_{sh} = I - I_m$

The current to be measured is passed through the MC through flexible links.

$= R_m/(m - 1)$ (19.18)

where $m = I/I_m$ = multiplying factor

R_m = internal resistance of the meter

I_{sh} = shunt current > I_m

I = full-scale current of the ammeter including the shunt

✦ Multirange Ammeter

The current range of the dc ammeter may be further extended by a number of shunts, selected by a range switch. Such a meter is called a multirange ammeter. The meter shown in Fig. 19.5 is called a universal shunt.

✦ dc Voltmeter

The basic D'Arsonval movement can be converted into a dc voltmeter with the addition of a series resistor or multiplier, as shown in Fig. 19.6. The multiplier limits the current through the movement so as not to exceed the value of the full-scale deflection current (I_{fsd}).

A dc voltmeter is connected across a source of emf or a circuit component. The terminals are generally marked positive and negative since polarity must be observed.

The value of a multiplier required to extend the voltage is calculated as follows:

$V = I_m (R_s + R_m)$ or $R_s = V/I_m - R_m$ (19.19)

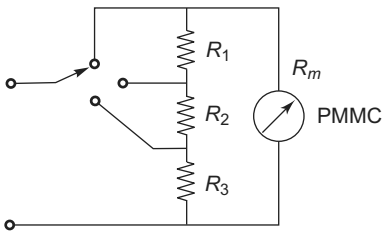


Fig. 19.5 Multirange ammeter

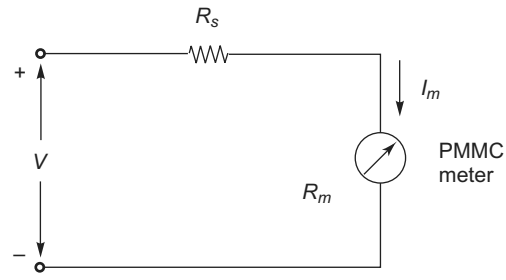


Fig. 19.6 dc voltmeter

Multiplying factor $m = V/V_m = I_m (R_m + R_s)/I_m R_m$

$$m = 1 + R_s/R_m$$

$$\therefore \quad R_s = (m - 1)R_m \tag{19.20}$$

where I_m = deflection current of the meter
R_s = multiplier resistance
V = full-scale voltage of the instrument
V_m = Voltage across movement

✦ Multirange Voltmeter

A multirange voltmeter can be obtained by the addition of a number of multipliers together with a range switch. There are two types of multirange voltmeters.

1. The meter shown in Fig. 19.7(a) has only one resistance connected to the PMMC meter.
2. The meter shown in Fig. 19.7(b) has all multipliers connected in series with the PMMC meter.

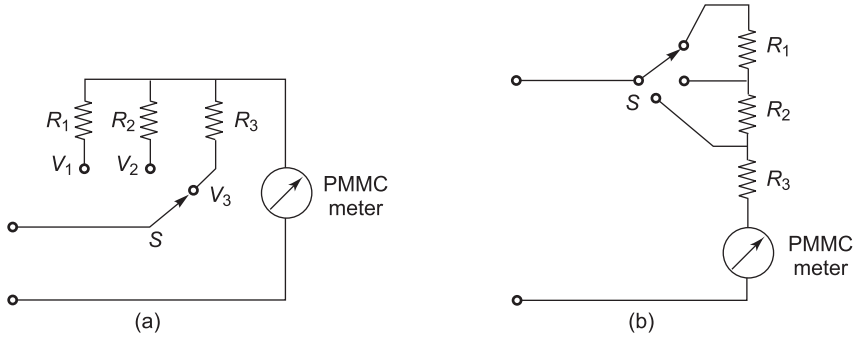


Fig. 19.7 Multirange voltmeters

Voltmeter sensitivity (ohm/V rating)

$$S = 1/I_{\text{fsd}}\ \Omega/\text{V} \tag{19.21}$$

The total resistance of the voltmeter will be

$$R_{\text{total}} = S \times V,\ R_s = (S \times V) - R_m \tag{19.22}$$

❑ ***Loading Effect*** A low sensitivity meter may give a correct reading when measuring voltages in low resistance circuits. A voltmeter when connected across two points in a highly resistive circuit, acts as a shunt for that portion of the circuit and thus reduces the equivalent resistance in that portion of the circuit. The meter will give a lower indication of the voltage drop that actually existed before the meter was connected. This effect is called the *loading effect* of an instrument.

EXAMPLE 19.2 A moving-coil instrument gives full-scale deflection with 25 mA. The resistance of the coil is 5 Ω. It is required to convert this meter into an ammeter to read up to 5 A. Find (a) the resistance of the shunt to be connected in parallel with the meter, and (b) the value of series resistance for the above meter to read up to a voltage of 20 V.

Solution

(a) Full scale $I_{\text{fsd}} = 25$ mA

$\therefore$ current through shunt resistance (R_x) = 4.975 A

$$25 \text{ mA} \times 5\ \Omega = R_x \times 4.975 \text{ A}$$

$$\therefore \quad R_x = 0.025\ \Omega$$

(b) For the meter to read 20 V,

$$20 \text{ V} = (R_x + 5) \times 25 \text{ mA}.$$

$$\therefore \quad R_x = 795\ \Omega$$

EXAMPLE 19.3 If the instrument of Example 19.2 is to be converted into a multirange voltmeter to read up to 40 V and 60 V, find the additional resistance to be connected in series with the instrument.

Solution Let R_1 be the resistance in series (Fig. 19.8)

$$40 = (R_1 + 5) \times 0.025$$

$$\therefore \quad R_1 = 1595\ \Omega$$

Let R_2 be the other resistance in series.

Then, $$60 = (R_1 + R_2 + 5) \times 0.025$$

$$R_2 = 800\ \Omega$$

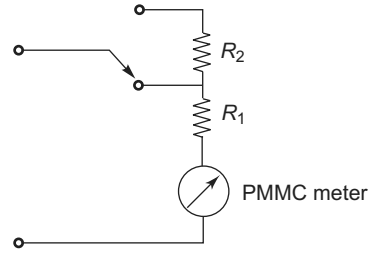


Fig. 19.8

19.4.3 Ohmmeter

This instrument is used to measure resistance. It is mainly of two types.

✦ Series-Type Ohmmeter

It essentially consists of a D'Arsonval movement connected in series with a resistance and a battery to a pair of terminals to which the unknown resistance is connected, as shown in Fig. 19.9. The current through the movement thus depends on the magnitude of the unknown resistor R_x and the meter indication is proportional to the value of R_x.

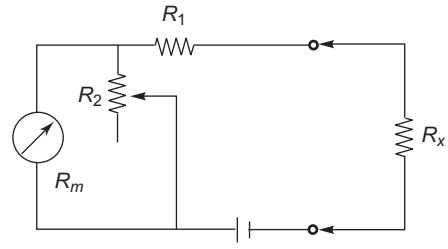


Fig. 19.9 Series type ohmmeter

When R_x is zero, maximum current flows in the circuit and shunt resistance R_2 is adjusted until the movement indicates full-scale current. The full-scale current position of the pointer is marked 0 Ω on the scale. When R_x is infinity, the current drops to zero and movement indicates zero current, which is marked as '∞' on the scale.

A convenient quantity to use in the design of a series type ohmmeter is the value of R_x (= R_h) which causes half-scale deflection of the meter.

$$R_h = R_1 + (R_2 \parallel R_m)$$

Then, total resistance for the battery is 2 R_h.

$$\therefore \quad I_t = 2\, I_h = E/R_h,\ I_2 = I_t - I_{\text{fsd}}$$

where I_t = current through R_x for producing full-scale meter deflection
I_h = current for producing half-scale meter deflection
I_{fsd} = current through movement causing full-scale deflection
I_2 = current through R_2

Also, voltage across shunt (R_2) is equal to voltage across movement, (R_m), i.e. $E_{\text{sh}} = E_m$

$$\text{or} \quad I_2 R_2 = I_{\text{fsd}}\, R_m \tag{19.23}$$

$$\therefore \quad R_2 = I_{\text{fsd}}\, R_m\, R_h/(E - I_{\text{fsd}}\, R_h)$$

$$R_1 = R_h - I_{\text{fsd}}\, R_m\, R_h/E \tag{19.24}$$

✦ Shunt-Type Ohmmeter (Fig. 19.10)

When R_x is zero, the meter current is zero. If R_x is infinity, a current finds a path only through the meter and by appropriate selection of the value of R_x, the pointer can be made to read full scale. This ohmmeter has a zero mark at the right-hand side of the scale. It is particularly used for low resistance.

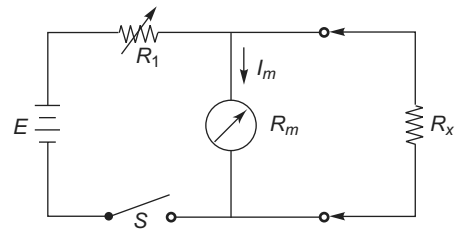


Fig. 19.10 Shunt-type ohmmeter

When R_x is infinite, the full-scale meter current will be

$$I_{\text{fsd}} = E/(R_1 + R_m)$$

For any value of R_x,

$$I_m = ER_x/[R_1 R_m + R_x (R_1 + R_m)]$$

Then

$$S = I_m/I_{\text{fsd}}$$

$$\therefore \quad S = R_x/[R_x + R_1R_m/(R_1 + R_m)]$$

$$= R_x/(R_x + R_p) \tag{19.25}$$

where

$$R_p = R_1 \| R_m$$

19.4.4 ac Indicating Instruments

PMMC meters cannot be used for measuring ac signals because the meter will show the average value of ac, i.e. zero deflection.

✦ Electrodynamometer

It can be used for ac as well as dc measurements. It contains two types of coils, a fixed coil and a moving coil. The field is produced by a fixed coil. This coil is divided into two sections to give a more uniform field.

Deflecting torque $T_d = BANI$ (19.26)

where B = flux density of the magnetic field in which the coil moves
A = area of cross section of coil
I = current through the coil
N = number of turns of coil

But $B \propto I$

$\therefore \quad T_d \propto I^2$

Instantaneous deflecting torque $T_i = i_1 i_2 \dfrac{\mathrm{d}M}{\mathrm{d}\theta}$ (19.27)

where i_1 = instantaneous current in the fixed coil
i_2 = instantaneous current in the moving coil
M = mutual inductance of coils
θ = deflection of pointer

❑ *Operation with dc*

$$T_d = I_1 I_2 \frac{\mathrm{d}M}{\mathrm{d}\theta} \tag{19.28}$$

where I_1 and I_2 are dc currents through fixed and moving coils, respectively.

This deflecting torque deflects the moving coil to such a position where the controlling torque ($k\theta$) of the spring is equal to the deflecting torque.

$$\therefore \quad k\theta = I_1 I_2 \frac{\mathrm{d}M}{\mathrm{d}\theta}$$

$$\therefore \quad \theta = (I_1 I_2/k)\frac{dM}{d\theta} \tag{19.29}$$

❑ ***ac Currents*** The meter will show the average value of the deflecting torque.

$$\therefore \quad T_{av} = \frac{dM}{d\theta}\frac{1}{T}\int i_1 i_2\, dt \tag{19.30}$$

❑ ***Sinusoidal Currents***

Let $i_1 = I_{m1} \sin \omega t$

and $i_2 = I_{m2} \sin (\omega t - \phi)$

The average torque $(T_{av}) = (dM/d\theta)\,(1/T)$

$$\int I_{m1} I_{m2} \sin \omega t \sin (\omega t - \phi) dt = (I_{m1} I_{m2}/2) \cos \phi\, dM/d\theta$$

$$\therefore \quad T_{av} = I_1 I_2 \cos \phi \frac{dM}{d\theta} \tag{19.31}$$

where I_1 and I_2 are the rms values of i_1 and i_2, respectively. At steady state,

$$k\theta = I_1 I_2 \cos \phi \frac{dM}{d\theta}$$

$$\therefore \quad \theta = \frac{I_1 I_2}{k} \cos \phi \frac{dM}{d\theta} \tag{19.32}$$

✦ Electrodynamometer in Current and Voltage Measurement

In an electrodynamometer, current under measurement itself produces a magnetic field flux in which the movable coil rotates.

A Fixed Coil (FC) splits into equal halves and provides the magnetic field in which the movable coil (MC) rotates. The two coil halves are connected in series and are fed by the current under measurement as shown in Fig. 19.11.

Meter deflection $\propto \sqrt{(\text{average } i^2)}$

The meter, therefore, reads the rms or effective value of ac.

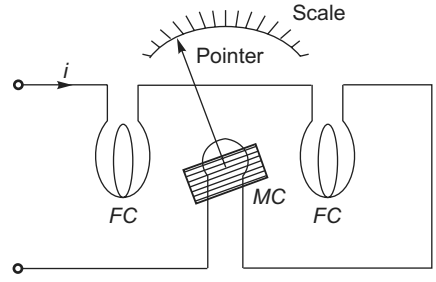


Fig. 19.11 Electrodynamometer type ammeter

In ammeters, fixed and movable coils are connected in series and, therefore, carry the same current. Hence,

$$I_1 = I_2 = I,\ \phi = 0$$

$$\therefore \quad \theta = (I^2/k)\frac{dM}{d\theta} \tag{19.33}$$

In the voltmeter, the fixed and movable coils are connected in series with a high non-inductive resistance.

Hence,

$$I_1 = I_2 = V/Z,\ \phi = 0$$

$$\therefore \quad \theta = (V^2/Z^2k)\ \mathrm{d}M/\mathrm{d}\theta \tag{19.34}$$

The appropriate selection of the shunt value converts the electrodynamometer into the desired range of the ammeter and the addition of the series resistance converts the meter into a voltmeter exactly like the dc ammeter and voltmeter discussed earlier.

These types of ammeters and voltmeters can measure either ac or dc quantities.

✦ Electrodynamometers in Power Measurement

The electrodynamometer movement is used extensively in measuring power. It may be used to indicate both dc and ac power for any waveform of voltage and current and is not restricted to a sinusoidal waveform.

The fixed or field (current) coil, shown in Fig. 19.12 as two separate elements, is connected in series and carry the total line current (i_c). The movable (potential) coil located in the magnetic field of fixed coils is connected in series with a current-limiting resistor (R_p) across the power line and carries the small current (i_p). The instantaneous value of current in the movable coil is

$$i_p = e/R_p$$

where e is the instantaneous voltage across the power line, and R_p is the total resistance of the movable coil and its series resistor.

The deflection of movable coil is proportional to the product of the two currents i_c and i_p.

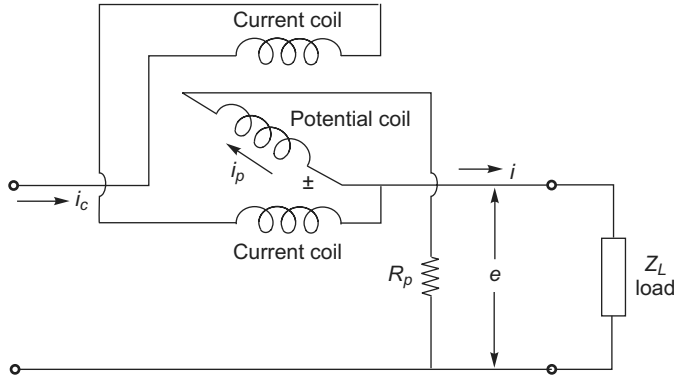


Fig. 19.12 An electrodynamometer in power measurement

Average deflection over one period

$$\theta_{\mathrm{av}} = (k/T) \int i_c\, i_p\, dt$$

Average power in circuit

$$\therefore \qquad P_{av} = 1/T \int ei\, dt \tag{19.35}$$

$$\therefore \qquad P_{av} = \theta_{av}$$

which indicates that the electrodynamometer movement of Fig. 19.12 has a deflection proportional to the average power.

✦ Electrodynamometer Wattmeters

An electrodynamometer wattmeter, shown in Fig. 19.13, has a current coil and a pressure coil.

The instantaneous torque is

$$T_i = i_1 i_2 \frac{dM}{d\theta}$$

where i_1 and i_2 are the instantaneous currents in the two coils and M is the mutual inductance of coils.

Therefore, the average deflecting torque is

$$T_i = (V I_c / R_p)\cos\phi \frac{dM}{d\theta}$$

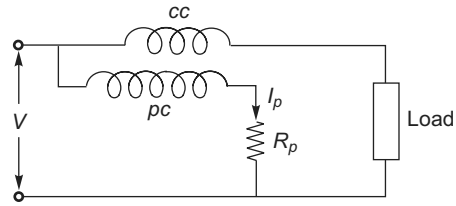


Fig. 19.13 Electrodynamometer wattmeter

where ϕ is the lagging phase angle of current in the current coil, I_c is the current coil current and R_p is the series resistance of pressure coil circuit.

$$\theta = (V I_c / kR_p)\cos\phi \frac{dM}{d\theta} \tag{19.36}$$

If the pressure coil has some inductance then

$$W = (V I_c / kR_p)\cos(\phi - \alpha)\cos\alpha \frac{dM}{d\theta} \tag{19.37}$$

where α is the lagging phase angle between current in pressure coil and voltage supply. But the true value of power is

$$W_{true} = (V I_c / kR_p)\cos\phi \frac{dM}{d\theta}$$

$$\therefore \qquad W/W_{true} = \cos\alpha\,[\cos\alpha + (\sin\alpha)\,(\tan\phi)] \tag{19.38}$$

19.4.5 Three-Phase Power Measurement

There are several ways in which three-phase power can be measured. Most important and common of these is the one called the *two-wattmeter method*, especially used when the load is unbalanced. Two wattmeters are used in a three-wire system (Fig. 19.14) with delta or star-connected load. The total instantaneous power consumed by the load is $v_1 i_1 + v_2 i_2 + v_3 i_3$

Instantaneous reading of $\qquad W_1 = i_1 (v_1 - v_3)$

Instantaneous reading of $\qquad W_2 = i_2 (v_2 - v_3)$

$\therefore$ Total instantaneous power in the load = $v_1 i_1 + v_1 i_2 - v_3 (i_1 + i_2)$

From Kirchhoff's law, $i_3 = -(i_1 + i_2)$

$\therefore$ Total instantaneous power = $v_1 i_1 + v_2 i_2 + v_3 i_3$

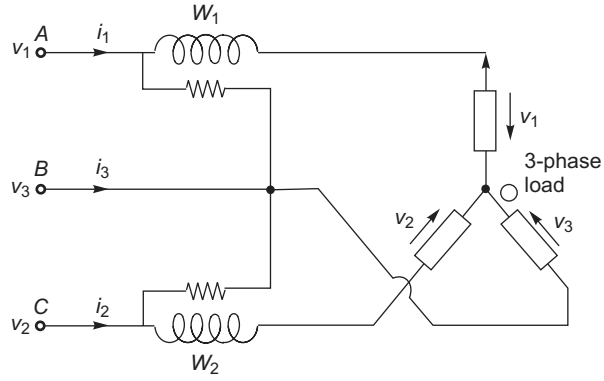


Fig. 19.14 Two-wattmeter method

Thus, the sum of the two wattmeter readings is equal to the power consumed by the load (balanced or unbalanced). Figure 19.15 shows the phasor diagram for a balanced star-connected load of Fig. 19.14.

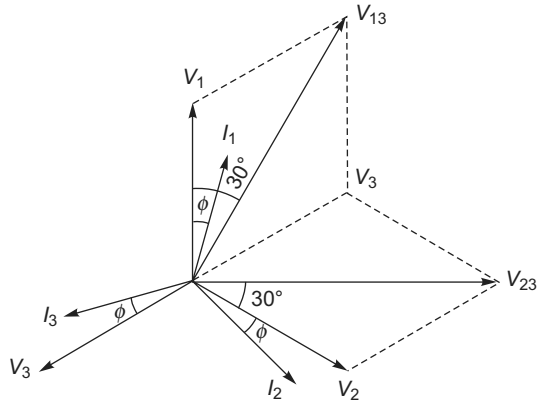


Fig. 19.15 Phase diagram for balanced star-connected load of Fig. 19.14

Consider a balanced load for the sake of simplicity.

Let $I_1 = I_2 = I_3 = I$; $V_1 = V_2 = V_3 = V$ (say) rms values

Line voltages $V_{13} = V_{23} = V_{12} = \sqrt{3}$ V

The phase currents lag the corresponding phasor voltages by an angle ϕ. The current I_1 flows through W_1 and voltage across its pressure coil is V_{13}. I_1 leads V_{13} by an angle $(30° - \phi)$.

$\therefore$ reading of $W_1 = V_{13}\, I_1 \cos(30° - \phi) = \sqrt{3}\, VI \cos(30 - \phi)$ (19.39)

The current through the wattmeter W_2 is I_2 and voltage across its pressure coil is V_{23}. I_2 lags V_{23} by an angle $(30° + \phi)$.

$\therefore$ reading of wattmeter $W_2 = V_{23} I_2 \cos(30° + \phi)$

$= \sqrt{3}\, VI \cos(30° + \phi)$ (19.40)

$\therefore$ sum of the wattmeter readings $= W_1 + W_2$

$= \sqrt{3}\, VI\ [\cos(30° - \phi) + \cos(30° + \phi)] = 3\, VI \cos\phi$ (19.41)

= three-phase power in the load

Difference of readings of two wattmeters is

$$W_1 - W_2 = \sqrt{3}\, VI\, [\cos(30° - \phi) - \cos(30° + \phi)]$$

$$= \sqrt{3}\, VI \sin\phi \quad (19.42)$$

$$\frac{W_1 - W_2}{W_1 + W_2} = \frac{\sqrt{3}\, VI \sin\phi}{3VI\cos\phi} = \frac{1}{\sqrt{3}} \tan\phi$$

or $$\phi = \tan^{-1} \sqrt{3}\, \frac{W_1 - W_2}{W_2 + W_2} \quad (19.43)$$

From Eq. (19.43), power factor (cos ϕ) may be found.

Some points worth noting are as follows:

(i) At unity power factor, the readings of the two wattmeters are equal.

(ii) When the power factor is 0.5, one of the wattmeters reads zero and the other reads total power.

(iii) With zero power factor, the readings of the two wattmeters are equal, but of opposite sign.

(iv) It should be noted that when the power factor is below 0.5, one of the wattmeters will give negative reading. Thus, to read the wattmeter, we must either reverse the current coil or the pressure coil connections. The wattmeter will then give a positive reading but this must be taken as negative for calculating the total power.

EXAMPLE 19.4 Two wattmeters are connected to measure power in a three-phase circuit. One of the wattmeters reads 500 W and the other points out in reverse direction. After reversing the voltage coil terminals, the reading of this wattmeter is found to be 200 W. Determine the power factor of the load and the total three-phase power of the circuit.

Solution

$W_1 = 500$ W $\qquad W_2 = -200$ W

$\therefore$ total power $= W_1 + W_2 = 300$ W

$$\tan \phi = \frac{500 - (-200)}{500 + (-200)} = \frac{700}{300} = \frac{7}{3}$$

$$\therefore \quad \phi = 66.8°$$

$$\text{pf} = \cos \phi = 0.39$$

19.4.6 Induction-Type Instruments

Induction-type instruments depend upon magnetic induction for operation and are used for ac measurement only.

$$\text{Deflecting torque } (T_d) \propto \phi_1, \phi_2 \tag{19.44}$$

$\therefore$ fluxes ϕ_1 and ϕ_2 both are proportional to some current I.

$\therefore$ $T_d \propto I^2$

The induction-type ammeter and voltmeter employ coil spring control and electromagnetic damping. The induction-type instrument is primarily used as a watt-hour meter or energy meter.

✦ Watt-hour Meter or Energy Meter

It is used for commercial measurement of electrical energy. Figure 19.16 shows the elements of a single-phase watt-hour meter.

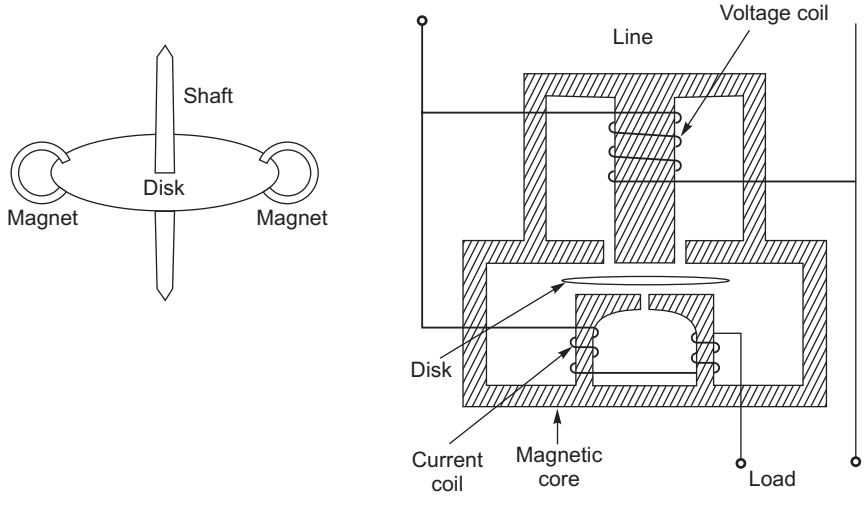


Fig. 19.16 Watt-hour meter or energy meter

The current coil is connected in series with the line, and the voltage coil is connected across the line. Both coils are wound on a metal of special design providing two magnetic circuits. A light aluminium disc is suspended in the air gap of the current coil field which causes eddy currents to flow in the disk. The reaction of the eddy currents and the field of voltage coil creates a torque on the disk, causing it to rotate. The number of rotations of the disk is proportional to the energy consumed by the load in a certain time interval.

At very light loads, the voltage component of the field produces zero torque that is not directly proportional to the load. Compensation for error is provided by inserting a shading coil with the meter operating at 10% of rated load. Two holes are drilled in the disk of the energy meter on the opposite side of the spindle to eliminate creeping on no load.

Electronic energy meters are now available and are gradually finding acceptability.

✦ Moving Iron Instrument

In this type of instrument, a plate of soft iron is the moving element of the system. This iron moves in a magnetic field produced by a stationary coil. The coil is excited by a current or voltage under measurement.

There are two types of moving iron instruments:

❑ ***Attraction Type*** In this type, a sheet of soft iron is attracted towards a solenoid.

❑ ***Repulsive Type*** In this type, two parallel strips of soft iron magnetised inside a solenoid repel each other.

The deflection of the pointer,

$$\theta = (I^2/2k)\, dL/d\theta \tag{19.45}$$

where k is the control spring constant, L is the instrument inductance and I is the initial current.

$$\therefore \quad \theta \propto I^2_{rms} \text{ for ac}$$

$$\propto I^2 \text{ for dc}$$

19.4.7 Attracted-iron Type

The schematic diagram is drawn in Fig. 19.17. As the current to be measured passes through the solenoid, magnetic flux created within it. The flux, causes the oblong soft iron disc to be attracted inwards and so the pointer moves through a certain angle. As the solenoid current increases, soft-iron disc is attracted more inwards increasing the pointer angle.

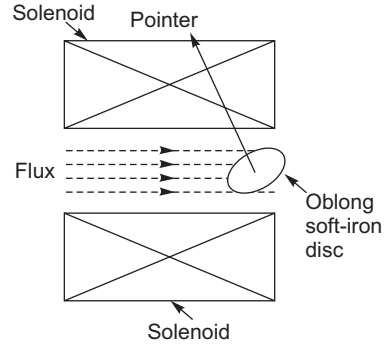


Fig. 19.17 Attracted-iron instrument

19.4.8 Repulsion-iron Type

Here two soft iron long pieces are placed inside the solenoid—one fixed and other mobile as shown in Fig. 19.18.

As the current to be measured is passed through the solenoid, the two soft-iron pieces get magnetised in the same direction and so repel each other. This causes the mobile piece to turn, and the spindle and pointer turns along with it, and the current value is read on a calibrated scale.

✦ Megger

Megger is an insulation testing instrument. It is used to measure very high resistances of the order of megaohms. This instrument works on the principle of an ohmmeter. The required deflecting torque is produced by both the system voltage and the current. Because of interaction

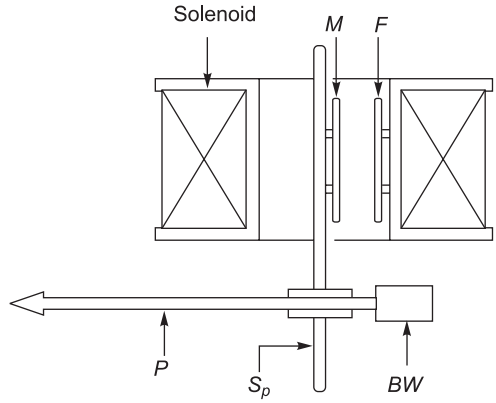


Fig. 19.18

between the magnetic fields produced by the voltage and the current, the deflecting torque is produced.

19.5 MULTIMETER OR VOM

The ammeter, the voltmeter, and the ohmmeter are all used as D'Arsonval movement. The difference between these instruments is in the circuit in which the basic movement is used. It is therefore obvious that a single instrument can be designed to perform the three measurement functions. This instrument, which contains a *function switch* to connect the appropriate circuits to the d'Arsonval movement, is often called a *multimeter* or *volt-ohm-milliammeter* (VOM).

A representative example of a commercial multimeter is shown in Fig. 19.19. The circuit diagram of this meter is given in Fig. 19.20. The meter is a combination of a dc milliammeter, a dc voltmeter, an ac voltmeter, a multirange ohmmeter, and an output meter.

Fig. 19.19 General-purpose multimeter. This instrument has been a familiar sight in electronics laboratories for many years (Courtesy, Simpson Electric Company)

Figure 19.21 shows the circuit for the dc voltmeter section, where the common input terminals are used for voltage ranges of 0–1.5 to 0–1,000 V. An external voltage jack, marked "dc 5,000 V," is used for dc voltage measurements to 5,000 V.

The basic movement of the multimeter of Fig. 19.19 has a full-scale current of 50 μA and an internal resistance of 2,000 Ω. The values of the multipliers are given in Fig. 19.21. Notice that

on the 5,000 V range, the range switch should be set to the 1,000–V position, but the test lead should be connected to the external jack marked "dc 5,000 V". The normal precautions for measuring voltages should be taken because of its fairly sensitivity (20 kΩ/V). The instruments are suitable for general service work in electronics field.

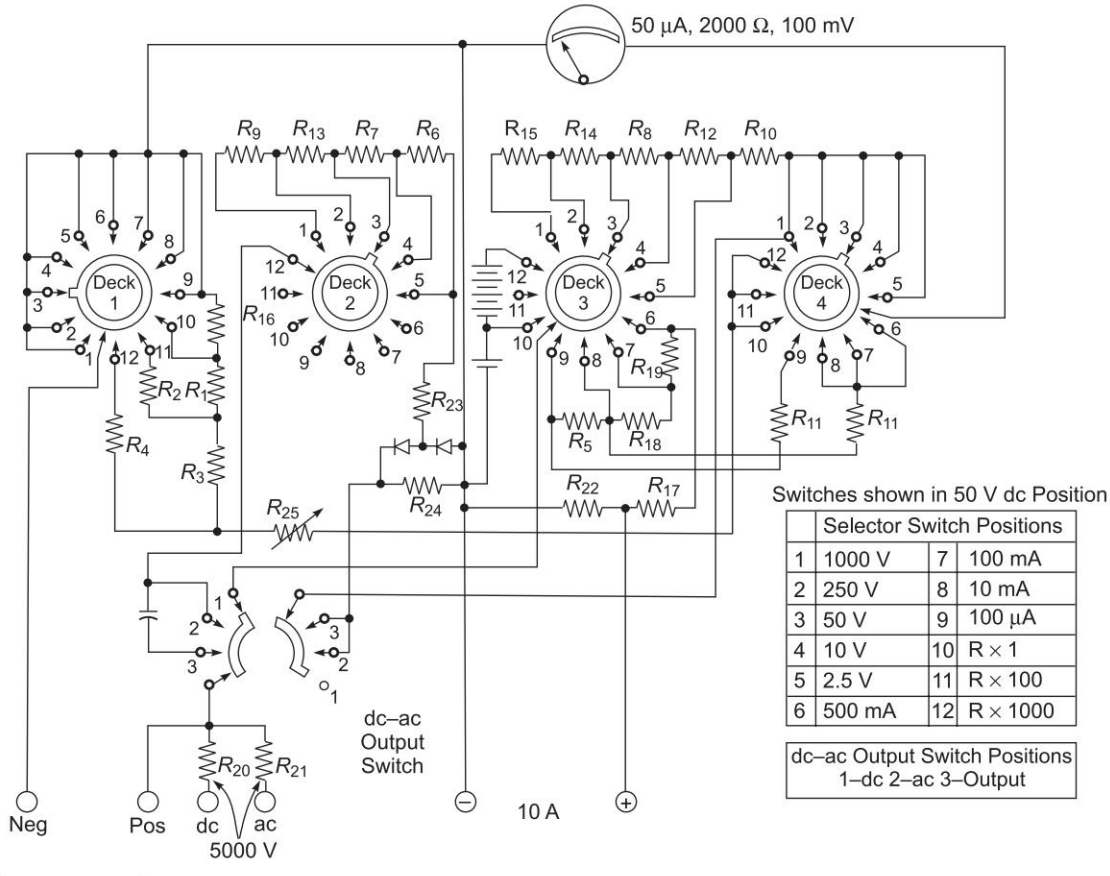


Fig. 19.20 Schematic diagram of the Simpson Model 260 multimeter (Courtesy, Simpson Electric Company)

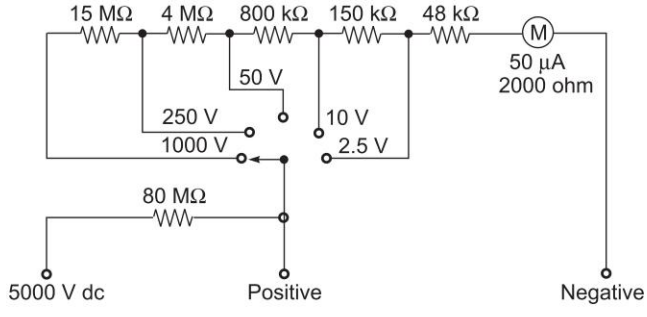


Fig. 19.21 dc voltmeter section of the Simpson Model 260 multimeter (Courtesy, Simpson Electric Company)

19.6 | OSCILLOSCOPE

The Cathode Ray Oscilloscope (CRO) is probably the most versatile tool for the development of electronic circuits and systems. The CRO allows the amplitude of electrical signals (e.g. V, I or P) to be displayed as a function of time. The CRO depends on the movement of an electron beam, which is bombarded (impinged) on a screen coated with a fluorescent material, to produce a visible spot. If the electron beam is deflected on both the conventional axes (X and Y axes), a two-dimensional display is produced. Typically, the X-axis of the oscilloscope is deflected at a constant rate, relative to time, and the vertical or Y-axis is deflected in response to an input stimulus such as voltage. This produces the time-dependent variation of the input voltage, which is very important to the design and development of electronic circuits.

The oscilloscope is basically an electron-beam voltmeter. The electron beam follows rapid variations in signal voltage and traces a visible path on the CRT (Cathode Ray Tube) screen which is the heart of the oscilloscope. Thus, rapid variations, pulsations or transients are reproduced and the analyst can observe the waveform as well as measure amplitude at any instant of time.

The oscilloscope can reproduce HF waves which are too fast for electromechanical devices to follow. Thus, it is a kind of recorder, which uses an electron beam instead of a pen. The oscilloscope is capable of displaying events that take place over periods of microseconds and nanoseconds.

A storage CRT can retain the display much longer, up to several hours after the image was first written on the phosphor. The retention feature will be useful while displaying the waveform of a very low frequency signal. A better method of trace storage is the digital storage oscilloscope. In this technique, the waveform to be stored is digitised, stored in a digital memory and retrieved for display on the storage oscilloscope. One very important feature of a digital storage oscilloscope is its ability to provide 'pretrigger view'. This means that the oscilloscope can display what happened before a trigger input is applied. This is useful when a failure takes place. To find the reason of the failure, it would be necessary to see different waveforms before the failure.

By combining a special fibre optic CRT with an oscillograph type paper drive (which panes the paper over the CRT face where it is exposed by light from the CRT phosphor), a recording oscilloscope with useful characteristics is obtained.

Oscilloscope Block Diagram

The heart of the oscilloscope is the cathode ray tube, which generates the electron beam, accelerates the beam to a high velocity, deflects the beam to create the image, and contains the phosphor screen where the electron beam eventually becomes visible. To accomplish these tasks, various electrical signals and voltages are required, and these requirements dictate the remainder of the blocks of the oscilloscope outline as shown in Fig. 19.22. The power-supply block provides the voltages required by the cathode ray tube to generate and accelerate the electron beam, as well as to supply the required operating voltages for the other circuits of the oscilloscope. Relatively high voltages are required by cathode ray tubes, of the order of a few thousand volts, for acceleration, as well as a low voltage for the heater of the electron gun,

which emits the electrons. Supply voltages for the other circuits are at various values, usually not more than a few hundred volts.

The laboratory oscilloscope has a time base which generates the correct voltage to supply the cathode ray tube to deflect the spot at a constant time-dependent rate. The signal to be viewed is fed to a vertical amplifier, which increases the potential of the input signal to a level that will provide a usable deflection of the electron beam. To synchronise the horizontal deflection with the vertical input, such that the horizontal deflection starts at the same point of the input vertical signal each time it sweeps, a synchronising or triggering circuit is used. This circuit is a link between the vertical input and the horizontal time base.

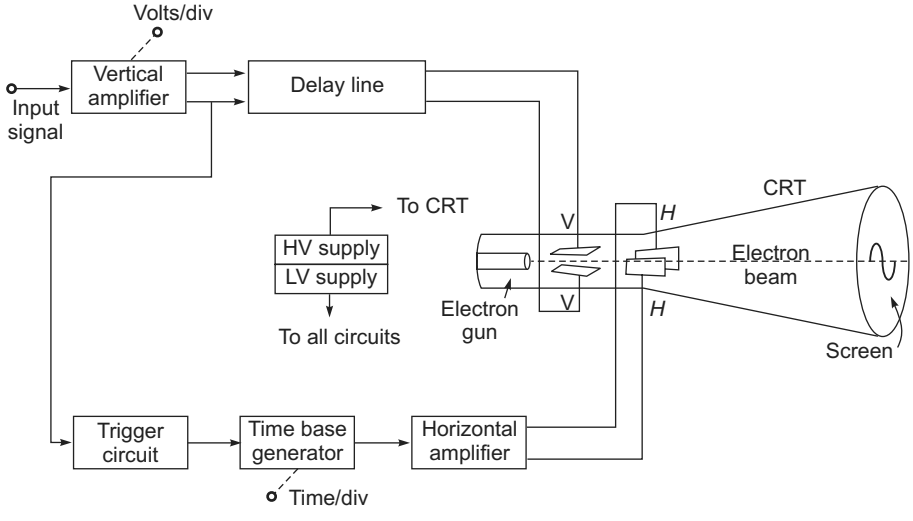


Fig. 19.22 Block diagram of a general-purpose oscilloscope

The linear deflection or horizontal sweep of a beam is accomplished by a *sweep generator*. The voltage output of sweep generator is shown in Fig. 19.23. Application of one cycle of this voltage difference, which increases linearly with time, to the horizontal plates causes the beam to be deflected linearly with time across the tube face. When the voltage suddenly falls to zero, as at points a, b, c, etc., the end of each sweep – the beam comes back to its initial position. The horizontal deflection of the beam is repeated periodically; the frequency of this periodicity is adjustable and can be controlled externally.

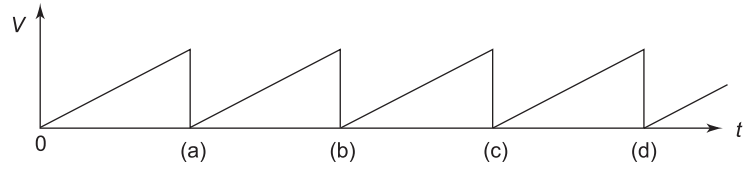


Fig. 19.23 Voltage output of sweep generator

To obtain steady traces at the output, an internal number of cycles of an unknown signal which is applied to the vertical plates must be synchronised with the cycle of the sweep generator. Hence, the pattern on the tube face repeats itself and appears to remain stationary.

19.7 VOLTAGE MEASUREMENT

Voltage is the amount of electric potential between two points in a circuit. Voltage, current and resistance all are internally related to each other. If we measure the voltage, rest of the values are obtained by mathematical calculations. An oscilloscope is mainly a voltage measuring device.

Oscilloscope displays peak-to-peak amplitude or voltage, that is, the absolute difference between the maximum amplitude of the signal to its minimum amplitude of the signal. After calculating high and low voltage points, or the amplitude, we can calculate the average of the minimum and maximum voltage.

19.8 CURRENT MEASUREMENT

Electrical current can be measured indirectly by attaching probes or resistors to the circuit. A resistor measures the voltage across the two points and using Ohm's law, by substituting the value of voltage and resistance calculated in the experiment, the value of electrical current can be calculated or we can use a clamp-on current probe with an oscilloscope to calculate the current.

19.9 FREQUENCY MEASUREMENT

Frequency of any signal can be measured using an oscilloscope by analysing the frequency spectrum on the screen. Frequency is defined as the number of times a cycle of an observed wave takes up in a second. The maximum frequency measured by an oscilloscope may vary but it is always in 100's of MHz range. To check the response of signals in a circuit, oscilloscope measures the rise and fall time of the wave.

✦ CRO Method

From the calibrated time base, display the wave and see the time interval for a cycle and calculate the frequency.

Frequency $f = (1/T)$ Hz (19.46)

where T is the time period in seconds.

✦ Dual-Trace Method (Fig. 19.24)

Display a known frequency wave on one channel to an unknown wave on the second channel. Count the number of cycles of both the waves.

$$f_{\text{unknown}} = (n/N) f_{\text{known}} \tag{19.47}$$

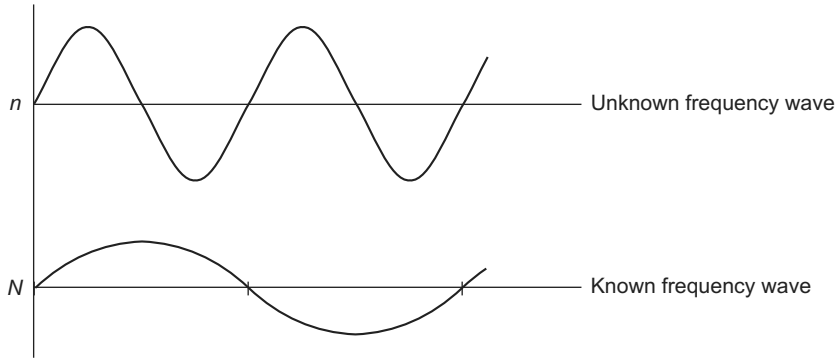


Fig. 19.24 Dual trace method

✦ Lissajous Pattern Method

An unknown frequency wave is applied to the *X*-plates and a known frequency wave is applied to the *Y*-plates of the CRO of Fig. 19.25. Adjust the known frequency wave until an elliptical loop appears on the screen as shown in Fig. 19.25. At this point, the two frequencies will be equal.

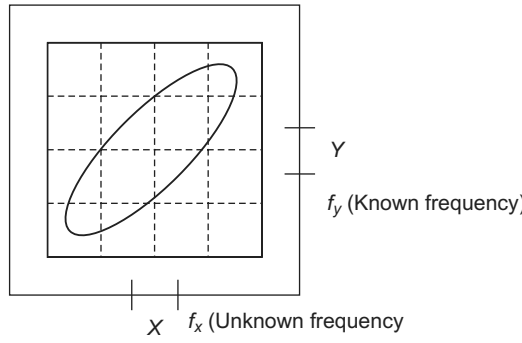


Fig. 19.25 Lissajous-pattern method

If f_x and f_y are equal then the Lissajous pattern could take one of the three shapes depending on the phase difference between the two signals as shown in Fig. 19.26.

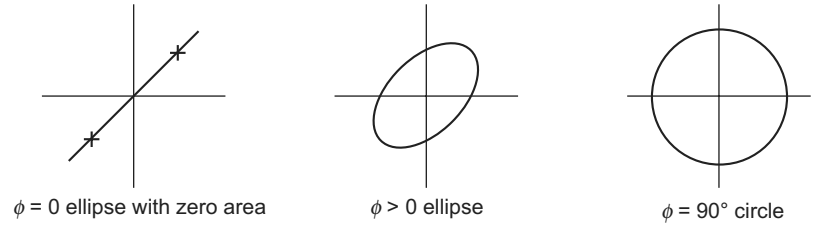


Fig. 19.26 Lissajous patterns

✦ Z Modulation Method

This method is specially useful for high frequency measurement. It could be used up to a known to unknown frequency ratio of 1:50.

19.10 | PHASE MEASUREMENT

✦ CRO Method

The two signals for which the phase difference is to be measured are applied to the *X*-plates and *Y*-plates to obtain a Lissajous pattern, are shown in Fig. 19.27. Then

$$\sin \phi = B/A = D/C$$

$\therefore$ Phase difference $\phi = \sin^{-1}(B/A) = \sin^{-1}(D/C)$

❑ ***Direct Reading Analogue Phase Meter*** (Fig. 19.28)

$$V_{av} = E_0\, t/T = E_0\, \phi°/360°$$

Since E_0 and 360° are constant, $V_{av} \propto \phi°$

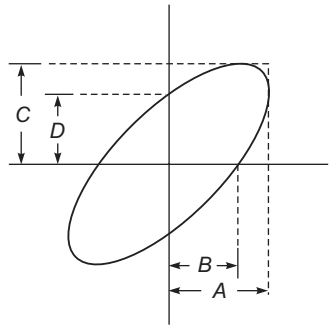


Fig. 19.27 CRO method

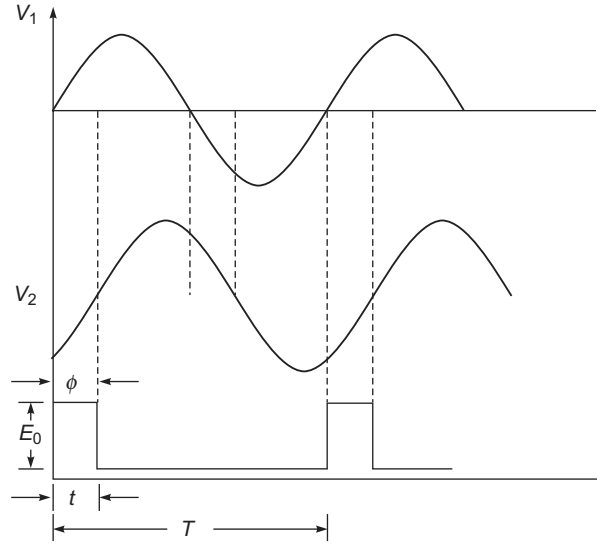


Fig. 19.28 Direct reading analogue phase meter

A multivibrator is being set at the edge of signal V_1 and gets reset at the edge of signal V_2, as shown in Fig. 19.29. Reading of the PMMC meter is calibrated in terms of the phase difference angle.

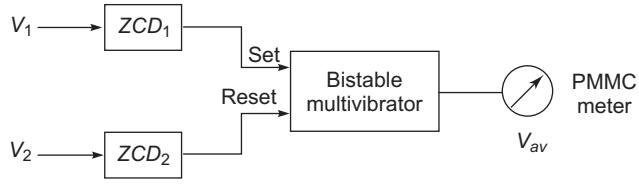


Fig. 19.29 Block diagram

19.11 DIGITAL INSTRUMENTS

Analog instruments display the quantity to be measured in terms of the deflection of a pointer. Digital instruments indicate the value of the measured in the form of a decimal number. The digital meters work on the principle of quantisation.

The advantages of digital instruments are as follows:

1. The readings are indicated directly in decimal numbers and, therefore, errors on account of human factors, such as errors due to parallax and approximation, are eliminated.
2. The readings may be carried to any significant figure by merely positioning the decimal point, i.e there is higher accuracy.
3. As compared to analog meters, digital instruments have a very high resolution.
4. Since output is in digital form, it may be directly fed into memory devices like tape recorders, printers and digital computers, etc. for storage and future computations.

19.11.1 Resolution in Digital Meters

The number of digits used in a digital meter determines the resolution. Thus, a three-digit Display Volt Meter (DVM) for a 0–1 V range will be able to indicate values from zero to 999 mV, with the smallest increment or resolution of 1 mV.

✦ Half Digit

In practice, a fourth digit, usually capable of indicating 0 or 1 only, is placed to the left of active digits. This permits going above 999–1999 to give an overlap between ranges for convenience. This is called *over-ranging*. This type of display is known as a half digit.

The resolution of a digital meter, however, is determined by the number of active or full digits used. If n is the number of full digits, then resolution is $1/10^n$.

For an 8-digit display, the resolution is 1 in 10^8, while for analog meters, in general, it is only 1 in 500.

19.11.2 Sensitivity of Digital Meters

Sensitivity is defined as the smallest change in the input which a digital meter is able to detect.

Sensitivity $S = (fs)_m \times R$ (19.48)

where $(fs)_m$ is the lowest full-scale value of the meter and R is the resolution expressed as decimal.

EXAMPLE 19.5 A 4½ digit voltmeter is used for voltage measurement.

(a) Find its resolution.
(b) How would 11.76 V be displayed on the 10 V range?
(c) How would 0.5434 be displayed on the 1 V range?
(d) How would 0.5434 be displayed on the 10 V range?

Solution

(a) Resolution = $1/10^4 = 0.0001$ or 0.01% V

(b) There are 5 digit places in a 4-digit display
$\therefore$ 11.76 V would be displayed as 11.760 V on its 10 V scale.

(c) Resolution on the 1 V range = $1 \times 0.0001 = 0.0001$ V
$\therefore$ on the 1 V range, any reading can be shown to the fourth decimal place.
Thus, 0.5434 V would be displayed as 0.5434 V on the 1 V range.

(d) Resolution on the 10 V range = $10 \times 0.0001 = 0.001$ V
Hence, on a 10 V range, readings can be displayed only up to the third decimal place.
$\therefore$ 0.5434 V will be shown as 00.543 V on a 10 V range. The digit 4 in the fourth decimal place will be lost. However, by employing a suitable range, e.g. 1 V, the digit 4 can be retained.

19.11.3 Digital Voltmeters

The schematic block diagram of a DVM is shown in Fig. 19.30.

An analog to digital converter (ADC) is the most critical block of a DVM. It decides the accuracy, resolution, etc.

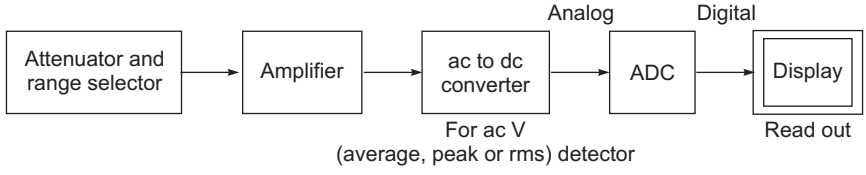


Fig. 19.30 A digital voltmeter

The types of DVM (according to the ADC principle) are as follows:

✦ Ramp-Type DVM

The principle behind ramp-type DVM is based on the measurement of the time it takes for a linear ramp voltage to rise from zero to the level of the input voltage, or to decrease from the level of the input voltage to zero.

Conversion from a voltage to a time interval is illustrated by the waveform diagram of Fig. 19.31(a). At the start of the measurement cycle, a ramp voltage is initiated, which is

continuously compared with the unknown input voltage. At the instant the ramp voltage is zero, the comparator 1 generates a pulse; another pulse is generated by the comparator 2, when the continuously increasing ramp voltage equals the unknown voltage. The start/stop circuit gives a pulse of width T using the two pulses.

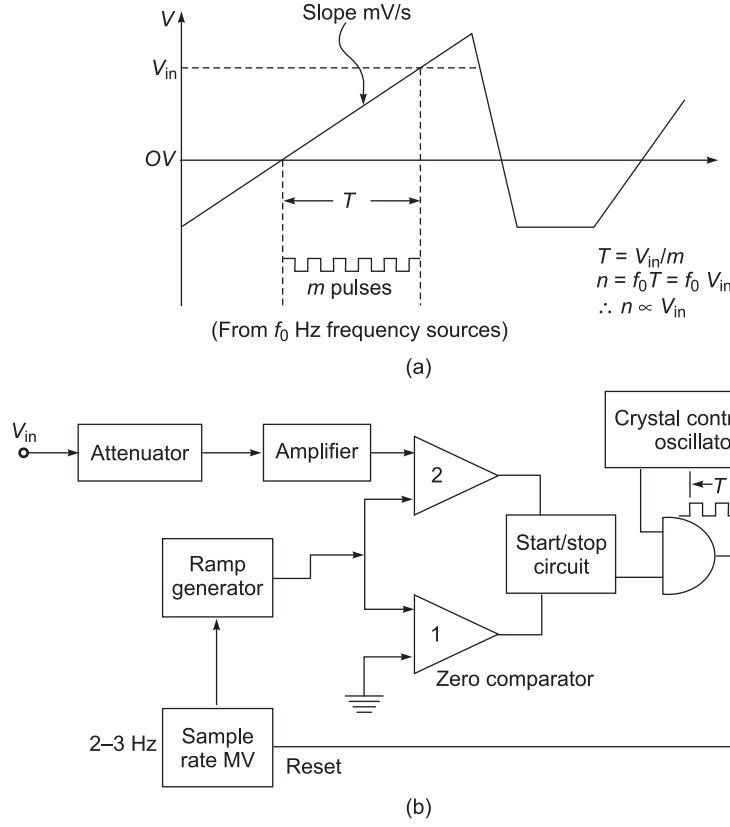


Fig. 19.31 (a) Waveforms for ramp-type DVM (b) Block diagram of a ramp-type DVM

The output of oscillator (f_0 Hz) is ANDed with the pulse of width T. Gate output goes to the number of decade counting units (dcUs), which totalise the number of pulses passed through the gate. The decimal number displayed by the readout is the input voltage.

✦ Dual-Slope-Type DVM

We change the capacitor for fixed time T_0 by V_{in}; then, it is discharged to fixed voltage V_{ref} and discharge time T is measured (see Fig. 19.32).

Charging: $nVp = -(1/RC) \int V_{in}\, dt$

where V_p = voltage across the capacitor after charging for time T_0

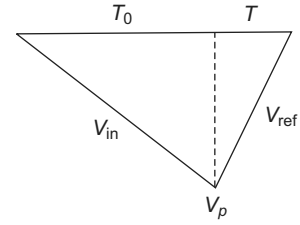


Fig. 19.32 V-T diagram of dual-slope-type DVM

R = charging resistance (in series with the capacitor)

Discharging: $V_p = -(1/RC) \int V_{ref}\, dt$

$$V_p = -(1/RC) \int V_{in} dt = -(1/RC) \int V_{ref} dt$$

$$T = V_{in} T_0 / V_{ref} \quad \text{or} \quad T \propto V_{in}$$

Both T_0 and T times are measured by pulses on the same frequency (f_0) with the same counter.

T_0: count by number of pulses

T: N count: pulses passed discharging

n: $V_{in} N_0 / V_{ref}$ or $n \propto V_{in}$

Figure 19.33 shows a complete dual-slope A/D converter. Electronic switches, usually FET switches, are used to switch the input of the integrator alternatively between the reference voltage and the unknown. Another pair of switches apply the integrator output to the automatic zero capacitor and ground the input for the automatic zero function. The switch timing and the counting of the clock pulses to determine the unknown voltage are under control of the control logic. The output is made available to the external electronics after the conversion is complete.

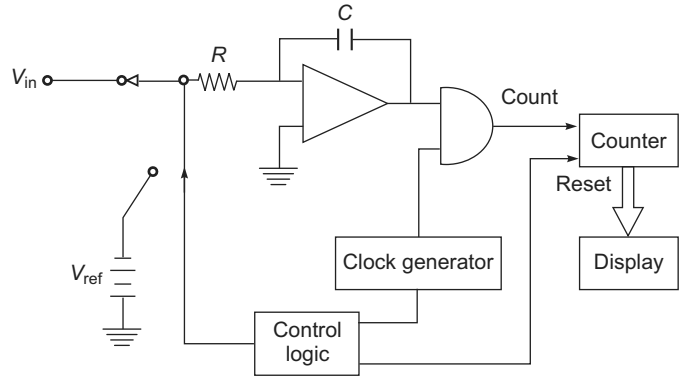


Fig. 19.33 Dual-slope A/D converter

EXAMPLE 19.6 A dual-slope integrating type of A/D converter has an integrating capacitor of 0.22 μF and a resistance of 100 Ω. If the reference voltage is 5 V and the output of the integrator is to remain below 10 V, find the maximum time the reference voltage can be integrated.

Solution Let T be the maximum time for which the reference voltage can be integrated. Then, the voltage across the capacitor after charging for time T is

$$V_p = (1/RC) \int V_{ref} dt$$

$V_p = 10$ V, $V_{ref} = 5$ V

$$\therefore \qquad 10 = (1/10^5 \times 0.22 \times 10^{-6}) \int 5 \, dt$$

Thus, $T = 44$ ms

✦ Successive-Approximation-Type DVM

This is the fastest compared to any other type of DVM. It is an electronic implementation of a technique called *binary regression*. This converter compares the analog input to a DAC reference voltage which is repeatedly divided in half. The process is shown in Fig. 19.34, where a three-digit binary number, representing the full voltage E_r, is divided in half (binary number 100), to the corresponding voltage $E_r/2$. A comparison of this reference voltage $E_r/2$ and the analog voltage is made. If the result of this comparison shows that this first approximation is too small (i.e. $E_r/2$ is smaller than the analog input) then the next comparison is made against $E_r/4$ (binary number 010). After four successive approximations, the digital number is resolved.

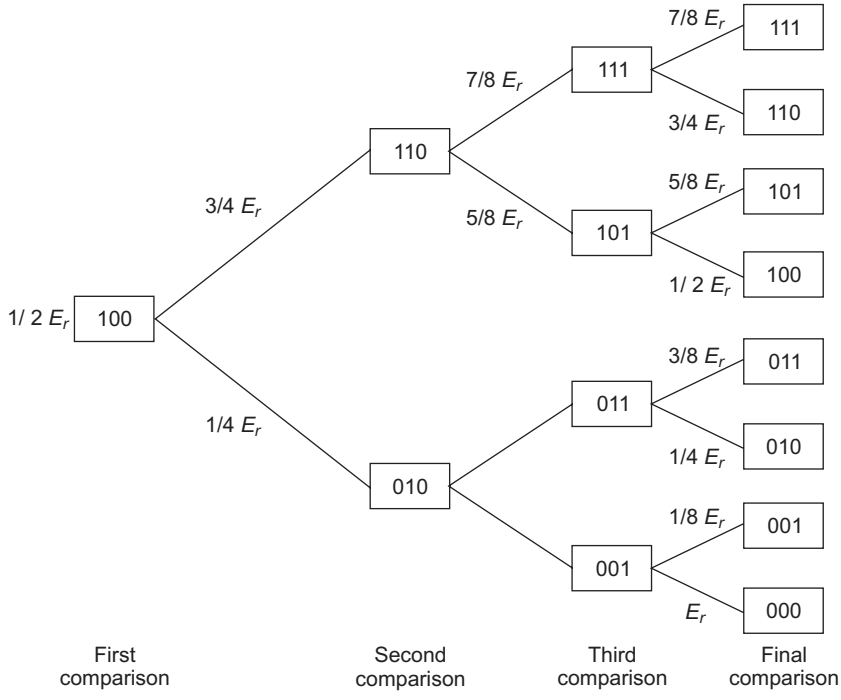


Fig. 19.34 Operation of successive approximation A/D converter

At the start of the conversion cycle, both the control resistor and the distribution register are set with a 1 in MSB and 0 in all bits of less significance. Thus, the distribution register shows 1000 and this causes the output voltage at the D/A convertor section to be one half of reference supply. At the same time, a pulse enters the time-delay circuitry. By the time the D/A converter and the comparator have settled, this delayed pulse is gated with the comparator output.

When the next MSB is set in the control register by the action of the timing circuit, the MSB remains in the state 1 or it is reset to 0, depending on the comparator output. The single 1 in the distribution register is shifted to the next position and keeps track of the comparison mode.

The procedure repeats itself (see Fig. 19.35) until the final approximation has been made and the distribution register indicates the end of the conversion.

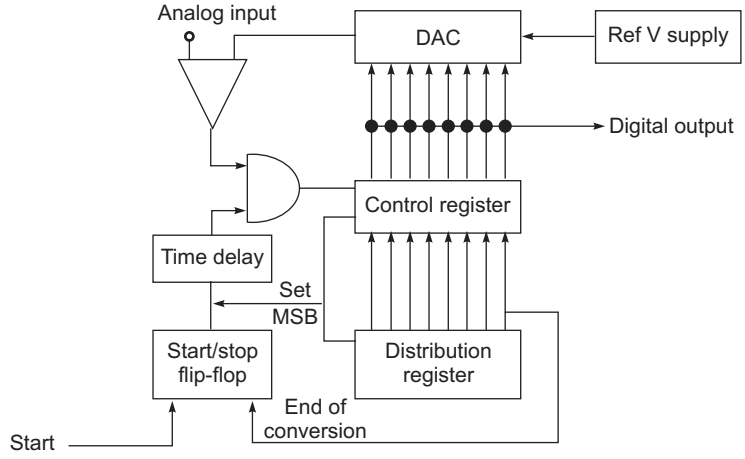


Fig. 19.35 Block diagram of the successive approximation A/D converter

EXAMPLE 19.7 Find the successive approximation A/D output for a 4-bit converter to a 3.217 V input if the reference is 5 V.

Solution For a 4-bit converter, only 4 digits can be shown. Input 3.217 V is approximately equal to 3.25 V. The output can be shown in tabular form as follows:

Pulse number	Value represented	Output (binary number)
1	2.5	1
2	1.25	0
3	0.625	1
4	0.3125	0

$\therefore$ the 4-bit representation is 1010.

EXAMPLE 19.8 A series type ohm-meter is designed to operate with a 6 V battery with a circuit diagram as shown in Fig. 19.36. The meter movement has an internal resistance of 2 kΩ and requires a current of 100 μA for full-scale deflection. The value of R_1 = 49 kΩ.

(a) Assuming the battery voltage has fallen to 9.5 V, calculate the value of R_2 required to obtain zero reading in the meter.

(b) Under the condition mentioned in part (a), an unknown resistor is connected to the meter causing 60% meter deflection. Calculate the value of the unknown resistance.

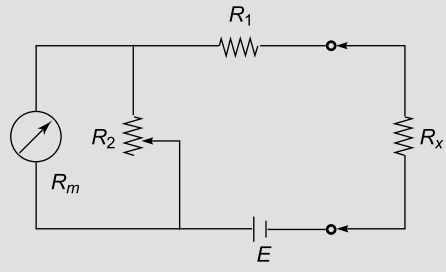


Fig. 19.36

Solution

(a) $R_x = 0$ for zero meter reading, $R_m = 2\ \text{k}\Omega$, $R_1 = 49\ \text{k}\Omega$,

$$I_{\text{fsd}} = 100\ \mu\text{A}$$

Voltage across movement = $2000 \times 100\ \mu\text{A} = 0.2\ \text{V}$

Then $IR_1 + 0.2 = 5.9$

$$I = 116.32\ \mu\text{A}$$

$$I_{\text{sh}} = 116.32\ \mu\text{A} - I_{\text{fsd}} = 16.32\ \mu\text{A}$$

$$R_2 = 0.2\ \text{V}/16.32\ \mu\text{A} = 12.25\ \text{k}\Omega$$

(b) for 60% deflection

$$I = 60 \times 116.32\ \mu\text{A}/100 = 69.8\ \mu\text{A}$$

$$R_{\text{eq}}\ (\text{for meter}) = 5.9/69.8\ \mu\text{A} = 84.53\ \text{k}\Omega$$

$$R_x = 84{,}530 - 49000 - 2000 \times 12.25/11.25$$

$$= 33.81\ \text{k}\Omega$$

EXAMPLE 19.9 The resistance of the pressure coil branch of the wattmeter W in the circuit of Fig. 19.37 is $R_p\ \Omega$. In the position 2 of the switch, an inductive reactance of $R_p\Omega$ is connected in series with the pressure coil branch. If the readings of the wattmeter in switch positions 1 and 2 are W_1 and W_2 respectively, determine the reactive power taken by the load in terms of W_1 and W_2. Neglect current coil impedance and pressure coil reactance.

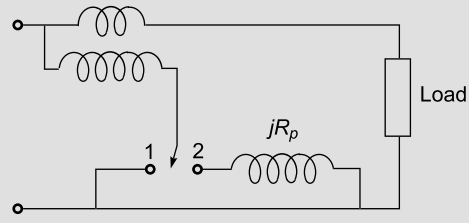


Fig. 19.37

Solution The wattmeter reading at the position 1 is

$$W_1 = (VI_c/k\,R_P)\cos\phi\, dM/d\theta \qquad \text{(i)}$$

The wattmeter reading at the position 2 is

$$W_2 = (VI_c/k\,R_P)\cos(\phi - \alpha)\cos\alpha\,\frac{dM}{d\theta} \qquad \text{(ii)}$$

where $\alpha = \tan^{-1}(\omega L/R_p)$

$\omega L = Rp$, reactance in series with pressure coil

$\alpha = 45°$

Dividing Eq. (ii) by Eq. (i), we get

$$W_2/W_1 = \cos 45° \cos(\phi - 45°)/\cos\phi$$

$$= (1 + \tan\phi)/2$$

$$\therefore \quad \tan\phi = (2W_2 - W_1)/W_1$$

Now reactive power $(P_r) = W_1 \sin\phi$

Thus, $$P_r = W_1(2W_2 - W_1)/\sqrt{4W_2^2 + 2W_1^2 - 4W_1W_2}$$

SUMMARY

- Moving Iron (MI) instrument is used to measure both ac and dc, but it gives rms value of ac and average value of dc.
- The two types of MI meters are:

  1. Attraction type 2. Repulsion type.
- Megger is an insulation testing instrument. It is used to measure very high resistance of the order of mega ohms.
- Instrument transformers are used in ac systems for the measurement of current, voltage, power and energy.

  T*ypes*: Current transformer, Potential transformer.
- Wheatstone bridge is used to measure resistance.
  Maxwell bridge is used to measure inductance.
  Hay bridge is used to measure high inductance.
  Schering bridge is used to measure high inductance.
  Owen's bridge is used to measure inductance.
  Anderson's bridge is used to measure inductance
  Wien's bridge is used to measure frequency.
- The cause of errors in measuring instruments is distributed capacitance and self-capacitance.
- **Sensitivity of digital meters:**

  It is defined as the smallest change in the input which a digital meter is able to detect.
  Sensitivity S = (Fs) $m \times$ R
  (Fs) m = lowest full-scale value of meter
  R = resolution expressed in decimal.
- An electrical transducer is a device which converts non-electrical input into electrical output.

EXERCISES

➥ Review Questions

1. What are the basic elements of a measuring instrument?
2. How can we get true values from measurements?
3. Why is the controlling torque needed in a measuring instrument?
4. How can the range of an ammeter/voltmeter be extended?
5. How can an electrodynamometer be converted into an ammeter/voltmeter?
6. Can we measure three-phase power with a single wattmeter and how?
7. In three-phase measurement using the two wattmeter method, if the power factor is zero, what are the two watt readings?
8. A thermocouple instrument measures average value or rms value? Why?
9. How can an instrument transformer extend the range of the ac measuring instrument?
10. Which will give accurate reading, bridge or meter measurement?
11. What are the properties of CRO which make it superior to the other measuring instruments?
12. What are the uses of vertical and horizontal plates provide in a CRO?
13. What is the purpose of a triggering circuit in a CRO?
14. What is the deflection sensitivity of a CRO?

➥ Problems

1. A 0–150 V voltmeter has a guaranteed accuracy of 1% full-scale reading, and the voltage measured by this instrument is 85 V. Calculate the limiting error in percentage.
2. The coil of a moving coil galvanometer has 250 turns and a resistance of 150 Ω. The coil dimensions are 2 cm × 2.5 cm. The strength of the uniform magnetic field is 0.12 Wb/m^2. The inertia constant of the moving system is 1.6×10^{-7} kg m^2 and the control torque constant is 2.4×10^{-6} Nm per radian. Assuming that the damping is entirely electromagnetic, determine the value of the resistance to be connected across the galvanometer terminals to obtain critical damping of the moving system.
3. For the instrument mentioned in Example 19.2, find the values of shunt resistances to convert the interament into a multirange ammeter reading upto 10 A, 15 A.
4. Figure 19.38 displays the lissajous patterns for a case where voltage of the same frequency out of different phases are connected to the *Y* and *X* plates of the oscilloscope. Find the phase difference in each case. The spot generating the patterns moves in a clockwise direction.

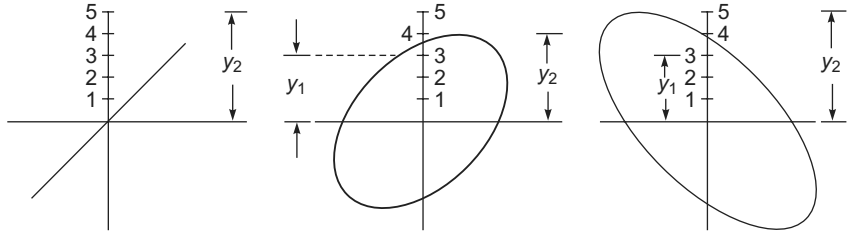


Fig. 19.38

5. The lowest range of a 4.5 digit voltmeter is 10 mV full scale. Find the sensitivity of the meter.
6. A certain 3 digit DVM has an accuracy specification of 0.5% of reading ± 2 digits.
   (a) What is the possible error in volts, when the instrument is reading 6.00 V on its 10 V range?

(b) What is the possible error in volts, when the instrument is reading 0.20 V on the 10 V range?
(c) What percentage of the reading is the possible error in the case of (b)?

➥ Multiple-Choice Questions

1. In an instrument, the error when reading at half scale is
(a) less than the full-scale error (b) equal to the full-scale error
(c) greater than the full-scale error (d) equal to half of the full-scale error

2. Which of the following meters has the best accuracy?
(a) Moving-iron meter (b) Moving-coil meter
(c) Rectifier-type meter (d) Thermocouple meter

3. A resistor of 10 kΩ with a tolerance of 5% is connected in parallel with a 5 kΩ resistor of 10% tolerance. What is the tolerance limit for the series network?
(a) 5% (b) 6.67% (c) 10% (d) 8.33%

4. A resistor of 10 kΩ with a tolerance of 5% is connected in parallel with a 5 kΩ resistor of 10% tolerance. What is the tolerance limit for the parallel network?
(a) 5% (b) 6.67% (c) 10% (d) 8.33%

5. A moving-coil instrument is used
(a) for low-frequency ac only (b) for both ac and dc circuits
(c) in dc circuits only (d) for measuring high-frequency ac

6. Moving-iron instruments have a scale, which is
(a) uniform (b) squared (c) log (d) none of these

7. Electrostatic instruments are used as
(a) voltmeters only (b) ammeters only
(c) both voltmeters and ammeters (d) wattmeters only

8. Dynamometer-type instruments can be used for
(a) ac only (b) dc only (c) both ac and dc (d) none of these

9. Dielectric loss can be measured by
(a) an energy meter (b) the Wheatstone bridge
(c) an electrostatic meter (d) none of the above

10. Which of the following ranges of a meter requires the smallest shunt resistance?
(a) 0–10 mA (b) 0–100 mA (c) 0–1 A (d) 0–10 A

11. CRO is an instrument which acts as a/an
(a) ammeter (b) voltmeter (c) wattmeter (d) watt-hour meter

12. The sweep generator of a CRO is used to produce
(a) sinusoidal voltage for the horizontal deflection of electron beam
(b) saw tooth voltage for the vertical deflection of electron beam
(c) saw tooth voltage for the horizontal deflection of electron beam
(d) sinusoidal voltage for the vertical deflection of electron beam

13. In an oscilloscope, in terms of division on screen, the voltage of the is
(a) peak-to-peak voltage (b) average voltage
(c) RMS voltage (d) maximum voltage

ANSWERS

✦ Multiple-Choice Questions

1. (b) **2.** (a) **3.** (b) **4.** (d) **5.** (c) **6.** (b) **7.** (a) **8.** (c) **9.** (c) **10.** (a)
11. (b) **12.** (c) **13.** (a)

APPENDIX A

Frequency Response of the FET Amplifier

A.1 FREQUENCY RESPONSE OF THE FET AMPLIFIER

The typical frequency response of an amplifier is presented in the form of a graph that shows output amplitude (or, more often, voltage gain) plotted versus log frequency. The typical plot of the voltage gain is shown in Fig. A.1. The gain is null at zero frequency, then rises as frequency increases, levels off for further increases in frequency, and then begins to drop again at high frequencies. The frequency response of an amplifier can be divided into three frequency regions.

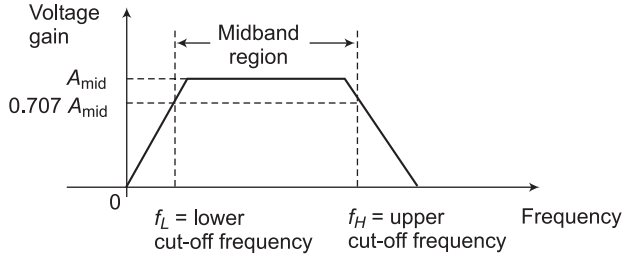


Fig. A.1 Diagram of voltage gain versus frequency for an amplifier

The frequency response begins with the lower frequency region designated between 0 Hz and lower cut-off frequency. At lower cut-off frequency, f_L, the gain is equal to 0.707 A_{mid}. A_{mid} is a constant midband gain obtained from the midband frequency region. The third region, the upper frequency region, covers frequency between upper cut-off frequency and above.

Similarly, at upper cut-off frequency, f_H, the gain is equal to 0.707 A_{mid}. After the upper cut-off frequency, the gain decreases with frequency increase and dies off eventually.

The Lower Frequency Response

Since the impedance of coupling capacitors increases as frequency decreases, the voltage gain of a FET amplifier decreases as frequency decreases. At very low frequencies, the capacitive reactance of the coupling capacitors may become large enough to drop some of the input voltage or output voltage. Also, the emitter-bypass capacitor may become large enough so that it no longer shorts the emitter resistor to ground. Approximately, the following equations can be used to determine the lower cut-off frequency of the amplifier, where the voltage gain drops 3 dB from its midband value (= 0.707 times the midband A_{mid}):

(1) $f_1 = 1/(2\pi r_{in} C_1)$ where:
f_1 = lower cut-off frequency due to C_1
C_1 = input coupling capacitance
r_{in} = input resistance of the amplifier

(2) $f_2 = 1/(2\pi r_{out} C_2)$ where:
f_2 = lower cut-off frequency due to C_2
C_2 = output coupling capacitance
r_{out} = output resistance of the amplifier

Provided that f_1 and f_2, are not close in value, the actual lower cut-off frequency is approximately equal to the largest of the two.

The Upper Frequency Response

Transistors have inherent shunt capacitances between each pair of terminals. At high frequencies, these capacitances effectively short the ac signal voltage.

A.2 | RIGHT-HAND RULE

In mathematics and physics, the **right-hand rule** is a common mnemonic for understanding notation conventions for vectors in three dimensions. It was invented for use in electromagnetism by the British physicist John Ambrose Fleming in the late 19th century.

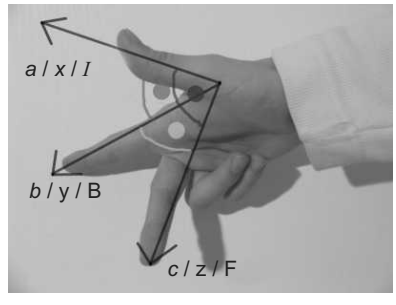


Fig. A.2 Right-hand rule

When choosing three vectors that must be at right angles to each other, there are two distinct solutions. Thus, when expressing this idea in mathematics, one must remove the ambiguity of which solution is meant.

There are variations on the mnemonic depending on context, but all variations are related to the one idea of choosing a convention.

Right-handed and Left-handed Coordinates

Let the *X* and *Y* axes define a horizontal plane with the *X*-axis pointing toward you. Then the *Z*-axis can either point up (right-handed) or down (left-handed). If the coordinates are right-handed and you place your right fist on the plane then your fingers will curl from the first or *X*-axis to the second or *Y*-axis and your thumb will point along the *Z*-axis. If the coordinates were left-handed, the fingers of your left hand would curl from the first axis to the second and your left thumb would point along the *Z*-axis.

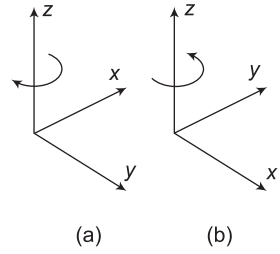


Fig. A.3 (a) Left-handed orientation (b) Right-hand orientation

If any one axis is reversed, the handedness changes. To preserve handedness, one of the other axes must also reverse which is equivalent to switching the labels on the other two axes. Note that the handedness of coordinates has no meaning unless the axes are labelled in sequence: 1,2,3 or *x*, *y*, *z*.

✦ Applications

The first form of the rule is used to determine the direction of the cross product of two vectors. This leads to widespread use in physics, wherever the cross product occurs. A list of physical quantities whose directions are related by the right-hand rule is given below. (Some of these are related only indirectly to cross products, and use the second form.)

- The angular velocity of a rotating object and the rotational velocity of any point on the object
- A torque, the force that causes it, and the position of the point of application of the force
- A magnetic field, the position of the point where it is determined, and the electric current (or change in electric flux) that causes it
- A magnetic field in a coil of wire and the electric current in the wire
- The force of a magnetic field on a charged particle, the magnetic field itself, and the velocity of the object

- The vorticity at any point in the field of flow of a fluid
- The induced current from motion in a magnetic field (known as Fleming's right-hand rule)
- The x, y and z unit vectors in a Cartesian coordinate system can be chosen to follow the right-hand rule
- Right-handed coordinate systems are often used in rigid body physics and kinematics.

A.3 | LEFT-HAND RULE

In certain situations, it may be useful to use the opposite convention, where one of the vectors is reversed and so creates a left-handed triad instead of a right-handed triad.

An example of this situation is for left-handed materials. Normally, for an electromagnetic wave, the electric and magnetic fields, and the direction of propagation of the wave obey the right-handed rule. However, left-handed materials have special properties, notably the **negative refractive index**. It makes the direction of propagation point in the opposite direction.

De Graaf's translation of Fleming's left-hand rule—which uses thrust, field and current—and the right-hand rule, is the FBI rule. The **FBI rule** changes thrust into F (Lorentz force), B (direction of the magnetic field) and I (current). The FBI rule is easily remembered by US citizens because of the commonly known abbreviation for the Federal Bureau of Investigation.

Fleming's left-hand rule is a rule for finding the direction of the thrust on a conductor carrying a current in a magnetic field.

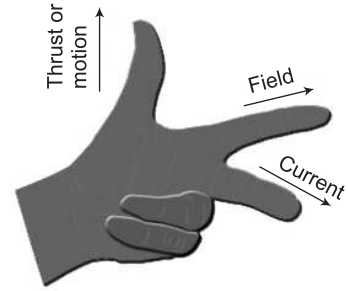


Fig. A.4 Left-hand rule

✦ Symmetry

Vector	Right hand	Right hand	Right hand	Left hand	Left hand	Left Hand
a, x or I	Thumb	Fingers or palm	First or index	Thumb	Fingers or palm	First or index
b, y or B	First or index	Thumb	Fingers or palm	Fingers or palm	First or index	Thumb
c, z or F	Fingers or palm	First or index	Thumb	First or index	Thumb	Fingers or palm

References

- *http://en.wikipedia.org/wiki/Right-hand_rule*
- *http://en.wikipedia.org/wiki/Fleming%27s_left-hand_rule_for_motors*
- *http://en.wikipedia.org/wiki/Hall_effect*
- *http://www.electronics-tutorials.ws/amplifier/amp_6.html*

APPENDIX B

Capacitive Sensors

There are several physical effects, which result in the direct generation of electrical signals in response to non-electrical influences and thus can be used in direct sensors. Examples are thermoelectric (Seebeck) effect, piezoelectricity, and photo effect. Let us take two isolated conductive objects of arbitrary shape (plates) and connect them to the opposite poles of a battery. The plates will receive equal amounts of opposite charges; i.e. a negatively charged plate will receive additional electrons while there will be a deficiency of electrons in the positively charged plate. Now, let us disconnect the battery. If the plates are totally isolated and exist in a vacuum, they will remain charged theoretically infinitely long. A combination of plates, which can hold an electric charge is called a *capacitor*. If a small *positive* electric test charge, q_0, is positioned between the charged objects, it will be subjected to an electric force from the positive plate to the negative. The positive plate will repel the test charge and the negative plate will attract it, resulting in a combined push-pull force. Depending on the position of the test charge between the oppositely charged objects, the force will have a specific magnitude and direction, which is characterized by vector **f**.

The capacitor may be characterized by q, the magnitude of the charge on either conductor, and by V , the positive potential difference between the conductors. It should be noted that q is not a net charge on the capacitor, which is zero. Further, V is not the potential of either plate, but the potential difference between them. The ratio of charge to voltage is constant for each capacitor: $q/V = C$.

This fixed ratio, C, is called the *capacitance* of the capacitor. Its value depends on the shapes and relative position of the plates. C also depends on the medium in which the plates are immersed.

Capacitance is a very useful physical phenomenon in a sensor designer's toolbox. It can be successfully applied to measure distance, area, volume, pressure, force, and so forth. The

following background establishes fundamental properties of the capacitor and gives some useful equations.

Parallel-plate Capacitor

In a parallel-plate capacitor, in which the conductors take the form of two plane parallel plates of area A separated by a distance d. If d is much smaller than the plate dimensions, the electric field between the plates will be uniform, which means that the field lines (lines of force **f**) will be parallel and evenly spaced. To calculate the capacitance, we must relate V , the potential difference between the plates, to q, the capacitor charge:

$$C = q/V \tag{B.1}$$

Alternatively, the capacitance of a flat capacitor can be found from

$$C = \varepsilon_0 A/d \tag{B.2}$$

The above formula is important for the capacitive sensor's design. It establishes a relationship between the plate area and the distance between the plates. Varying either of them will change the capacitor's value, which can be measured quite accurately by an appropriate circuit. It should be noted that the above equations hold only for capacitors of the parallel type. A change in geometry will require modified formulas. The ratio A/d may be called a *geometry factor* for a parallel-plate capacitor.

Cylindrical Capacitor

A cylindrical capacitor consists of two coaxial cylinders of radii a and b and length l. For the case when $l \gg b$, we can ignore fringing effects and calculate capacitance from the following formula:

$$C = 2\pi\varepsilon_0 l/\ln(b/a).$$

In this formula, l is the length of the overlapping conductors and $2\pi l[\ln(b/a)]^{-1}$ is called a *geometry factor* for a coaxial capacitor. A useful displacement sensor can be built with such a capacitor if the inner conductor can be moved in and out of the outer conductor. According to the above equation, the capacitance of such a sensor is in a linear relationship with the displacement, l.

Capacitive Occupancy Detectors

Being a conductive medium with a high dielectric constant, a human body develops a coupling capacitance to its surroundings. This capacitance greatly depends on such factors as body size, clothing, materials, type of surrounding objects, weather, and so forth. However wide the coupling range is, the capacitance may vary from a few picofarads to several nanofarads. When a person moves, the coupling capacitance changes, thus making it possible to discriminate static objects from the moving ones. In effect, all objects form some degree of a capacitive coupling with respect to one another. If a human (or, for that purpose, anything) moves into the vicinity of the objects whose coupling capacitance with each other has been previously established, a new capacitive value arises between the objects as a result of the presence of an intruding body.

Capacitive Sensors

The capacitive displacement sensors have very broad applications, they are employed directly to gauge displacement and position and also as building blocks in other sensors where displacements are produced by force, pressure, temperature, and so forth. The ability of capacitive detectors to sense virtually all materials makes them an attractive choice for many applications. Equation (B.2) states that the capacitance of a flat capacitor is inversely proportional to the distance between the plates. The operating principle of a capacitive gauge, proximity, and position sensors is based on either changing the geometry (i.e. a distance between the capacitor plates) or capacitance variations in the presence of conductive or dielectric materials. When the capacitance changes, it can be converted into a variable electrical signal. As with many sensors, a capacitive sensor can be either monopolar (using just one capacitor) or differential (using two capacitors), or a capacitive bridge can be employed (using four capacitors). When two or four capacitors are used, one or two capacitors may be either fixed or variable with the opposite phase.

Index

C

D

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

X

Z