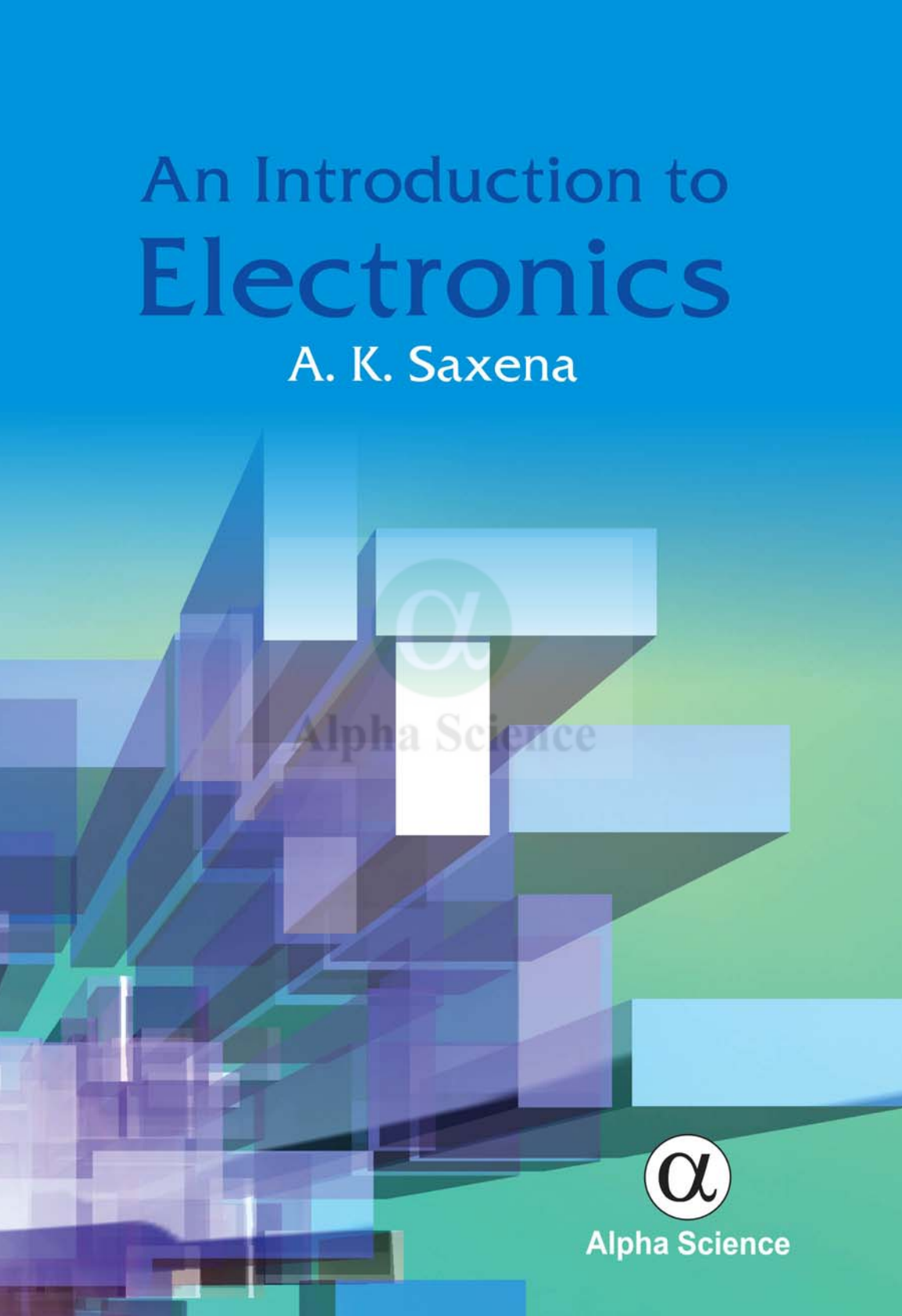
An Introduction to Electronics

A. K. Saxena



Alpha Science

An Introduction to Electronics

An Introduction to Electronics

A.K. Saxena

Alpha Science International Ltd.
Oxford, U.K.

An Introduction to Electronics
552 pgs. | 461 figs. | 10 tbls.

A.K. Saxena
Department of Physics
A.P.S. University
Rewa (M.P.)

ALPHA SCIENCE INTERNATIONAL LTD.
7200 The Quorum, Oxford Business Park North
Garsington Road, Oxford OX4 2JZ, U.K.

www.alphasci.com

ISBN 978-1-84265-860-4
E-ISBN 978-1-78332-063-9
Printed in India

Preface

This book is intended to serve as a textbook of Electronic for B.Sc. (Physics) students of all colleges and universities. For convenience of the students, the book has been divided into twelve chapters. The book covers circuit analysis (network analysis and network theorems), semi-conductor diodes, clippers and champers, two terminal devices and their applications, Bipolar Junction Transistors (BJT), Amplifiers, Two port networks and hybrid parameters, RC coupled and multistage amplifiers, Feedback in amplifiers (various types of feedback), sinusoidal oscillators, Multi-vibrators, Three terminal devices, Unijunction Transistor (UJT), Field Effect Transistor (FET); biasing and small signal models, FET amplifiers, MOSFETs, Modulation and demodulation (various types), and in last, noise and types of noise, noise in active circuits (BJT and FET), noise figure, noise power spectral density and effective noise temperature.

I am thankful to the authors and publishers of various books which have been helpful in providing an idea for the manuscript.

I also wish to express thanks to Mr. N.K. Mehra (Narosa Publishing House) for bringing out the book in a short time.

I hope that the book would be helpful to the students of the concerned course.

A.K. Saxena

Contents

Preface *v*

1. Circuit Analysis **1.1-1.96**

1.1	Ohm's Law, Resistivity and Conductivity	1.1
1.2	Resistances in Series	1.2
1.3	Resistances in Parallel	1.2
1.4	Division of Current in Parallel Circuits	1.3
1.5	Duality between Series and Parallel Circuits	1.4
1.6	Duality in a.c. Circuits	1.5
1.7	Star-Delta Connection	1.7
1.8	Network Definitions	1.10
1.9	Kirchhoff's Laws	1.13
1.10	Reduction of a Complicated Network	1.19
1.11	Conversions between T and π Sections	1.22
1.12	Thevenin's Theorem	1.27
1.13	Norton's Theorem	1.32
1.14	Superposition Theorem	1.35
1.15	Maximum Power Transfer Theorem	1.39
1.16	Wheatstone Bridge	1.50
1.17	Alternating Currents	1.52
1.18	Sinusoidal Waveform	1.53
1.19	RMS Value	1.54
1.20	Average Value	1.55
1.21	Units of Frequency and Period	1.56
1.22	Phasors	1.57
1.23	Complex Notation of Phasors	1.58
1.24	Polar Notation of Phasors	1.59
1.25	A.C. Voltage Applied to a Pure Resistance	1.60

1.26	A.C. Voltage Applied to a Pure Inductance	1.61
1.27	A.C. Voltage Applied to a Capacitance	1.61
1.28	Frequency Response of the Basic Elements	1.62
1.29	A.C. Power in the Basic Circuit Elements	1.63
1.30	Impedance	1.63
1.31	Series Connection of Basic Elements	1.64
1.32	Resistance and Inductance in Series	1.64
1.33	Resistance and Capacitance in Series	1.67
1.34	Inductance and Capacitance in Series	1.69
1.35	Resistance, Inductance and Capacitance in Series	1.71
1.36	Admittance and Susceptance	1.73
1.37	Kirchhoff's Laws for a.c. Networks	1.78
1.38	Superposition Theorem (for a.c. Networks)	1.79
1.39	Thevenin's Theorem (for a.c. Networks)	1.80
1.40	Norton's Theorem (for a.c. Networks)	1.81
1.41	Maximum Power Transfer Theorems – General Case (for a.c. Network)	1.82
1.42	A.C. Bridges	1.85
1.43	Maxwell's Inductance Bridge	1.86
1.44	Maxwell-Wien Bridge or Maxwell's L/C Bridge	1.88
1.45	Anderson Bridge	1.90
1.46	De Sauty Bridge	1.91
1.47	Schering Bridge	1.92
1.48	Wien Parallel Bridge	1.94
	Questions	*1.95*

2. Semiconductor Diodes **2.1-2.65**

2.1	Semiconductors	2.1
2.2	Formation of Energy Bands in Solids	2.2
2.3	Energy Bands in Diamond	2.2
2.4	Energy Bands in Metals, Insulators and Semiconductors	2.3
2.5	Intrinsic Semiconductors	2.5
2.6	Extrinsic Semiconductors	2.7
2.7	Mass Action Law	2.8
2.8	Charge Densities in an Extrinsic Semiconductors	2.9
2.9	Drift and Diffusion Currents	2.10
2.10	Fermi Level in an Intrinsic Semiconductor	2.11

2.11	Fermi Level in an Extrinsic Semiconductor	2.13
2.12	Current in a Semiconductor (Conductivity and Mobility)	2.14
2.13	Einstein Relationship	2.17
2.14	The Continuity Equation	2.17
2.15	Saturation of Drift Velocity	2.21
2.16	Fabrication of *P-N* Junctions	2.22
2.17	Basic Structure of *P-N* Junction	2.25
2.18	Built in Potential Barrier	2.28
2.19	Transition (or Space-Charge) Capacitance	2.31
2.20	A *P-N* Junction Under Forward and Reverse Bias (Diode Characteristics)	2.35
2.21	The Energy Band Diagrams	2.36
2.22	Current Flow Mechanism: The Rectifier Equation for a Forward Biased Diode	2.37
2.23	Application of the Continuity Equations to the Abrupt *P-N* Junction with Constant Current	2.39
2.24	Charge Carrier Distribution and Current Density in a Forward Biased *P-N* Junction	2.42
2.25	Junction Breakdown	2.45
2.26	The Ideal Diode	2.47
2.27	Static and Dynamic Resistance of a Diode	2.47
2.28	Load Line Concept	2.50
2.29	Dynamic Characteristics	2.51
2.30	Transfer Characteristic	2.51
2.31	Clippers	2.52
2.32	Clampers	2.60
2.33	*P-N* Diode Applications	2.63
	Questions	*2.64*

3. Two Terminal Devices and their Applications **3.1-3.64**

3.1	Introduction (D.C. Power Supply)	3.1
3.2	Rectifiers	3.2
3.3	Halfwave Rectifiers	3.2
3.4	Fullwave Rectifier	3.8
3.5	Peak Inverse Voltage	3.11
3.6	Comparison of Halfwave and Fullwave Rectifier	3.11
3.7	Bridge Rectifier	3.12

3.8	Filter Circuits	3.16
3.9	Series Inductor Filter	3.17
3.10	Shunt Capacitor Filter	3.19
3.11	Choke-Input or *L*-Section Filter	3.23
3.12	Capacitor-Input or π-Filter	3.26
3.13	Salient Features of *L*-Section and π-Filters	3.28
3.14	Zener Diodes	3.28
3.15	Zener Diode Specifications	3.29
3.16	Zener Diode Equivalent Circuit	3.30
3.17	The Voltage Regulator Circuit	3.31
3.18	Zener Diode Analysis	3.32
3.19	Design of a Voltage Regulator Circuit	3.35
3.20	Zener Diode Applications	3.37
3.21	Tunnel Diode	3.41
3.22	Varactor Diode	3.47
3.23	Photoconductivity	3.50
3.24	Photodiode	3.54
3.25	Light-Emitting Diode	3.56
3.26	Photovoltaic or Solar-Cell	3.59
	Questions	*3.63*
4. The Bipolar Junction Transistors		**4.1-4.18**
4.1	Introduction	4.1
4.2	The *PNP* and *NPN* Transistors	4.1
4.3	Transistor Operation	4.3
4.4	Emitter Efficiency γ, Base Transport Factor β and the Collector Efficiency δ	4.5
4.5	Common-Base Configuration	4.7
4.6	Base-Width Modulation and Punch through Effect	4.9
4.7	Alpha (α)	4.11
4.8	The Transistor Common Base Circuit as Amplifier	4.11
4.9	Common-Emitter Configuration	4.14
4.10	Beta (β)	4.16
4.11	Relationship between β and α	4.17
	Questions	*4.18*
5. Transistor Biasing and Amplifiers		**5.1-5.42**
5.1	Why Bias a Transistor?	5.1

5.2	The Load Line and Operating Point	5.1
5.3	Selection of Operating Point	5.2
5.4	Inherent Variations of Transistor Parameters	5.6
5.5	Stabilization	5.6
5.6	Essentials of a Transistor Biasing Circuit	5.8
5.7	Stability Factor	5.8
5.8	Methods of Transistor Biasing	5.9
5.9	Base Resistor Method	5.9
5.10	Biasing with Feedback Resistor	5.12
5.11	Bias Circuit with Emitter Resistor	5.14
5.12	Voltage Divider Bias	5.17
5.13	Analysis of Voltage-Divider Bias using Thevenin Theorem	5.19
5.14	Stability and Stability Factor of Voltage Divider Bias Circuit	5.23
5.15	Practical Design Approach using Voltage-Divider Bias	5.25
5.16	Mid-Point Biasing	5.27
5.17	Practical Circuit of Transistor Amplifier	5.28
5.18	Phase Reversal	5.30
5.19	D.C. and A.C. Equivalent Circuits	5.32
5.20	Load Line Analysis	5.35
5.21	Output from Transistor Amplifier	5.40
	Questions	*5.41*

6. Two Port Networks and Amplifiers **6.1-6.53**

6.1	Two Port Network	6.1
6.2	Controlled Sources	6.4
6.3	Hybrid-Parameter Equivalent Circuit	6.6
6.4	An Equivalent Circuit for the BJT	6.6
6.5	Graphical Determination of *h*-Parameters	6.9
6.6	Experimental Determination of *h*-Parameters	6.9
6.7	Performance of a Linear Circuit in terms of *h*-Parameters	6.12
6.8	Mathematical Analysis of Common-Emitter Amplifier using Hybrid Parameters	6.14
6.9	Mathematical Analysis of Common Base Amplifier using Hybrid Parameters	6.18
6.10	Mathematical Analysis of Common Collector Transistor using Hybrid Parameters	6.22
6.11	Conversion of the *h*-Parameters	6.25

6.12	Values of *h*-Parameters of a Typical BJT under CE, CB and CC Configurations	6.26
6.13	Classification of Amplifiers	6.26
6.14	Multistage Amplifiers	6.28
6.15	R-C (Resistance-Capacitance) Coupled Transistor Amplifiers	6.39
6.16	Comparison of Different Types of Couplings	6.50
6.17	Variations of Hybrid Parameters of a Transistor	6.51
	Questions	*6.52*
7. Feedback in Amplifiers		**7.1-7.24**
7.1	Introduction	7.1
7.2	Feedback Equation	7.2
7.3	Meaning of the Loop-Gain, $A\beta$	**7.4**
7.4	Some Fundamental Assumptions	7.4
7.5	Salient Features of Negative Feedback Amplifier	7.5
7.6	Types of Feedback Connections	7.10
7.7	Advantages of Negative Feedback	7.21
7.8	Disadvantages of Negative Feedback System	7.21
	Questions	*7.23*
8. Sinusoidal Oscillators		**8.1-8.21**
8.1	Introduction	8.1
8.2	Kinds of Oscillators	8.2
8.3	Operation of Oscillator	8.3
8.4	Essentials of an Oscillator Circuit	8.6
8.5	Types of Transistor Oscillators	8.9
8.6	Colpitt's Oscillator	8.10
8.7	Hartley Oscillator	8.12
8.8	Phase Shift Oscillator	8.14
8.9	Wien Bridge Oscillator	8.17
	Questions	*8.20*
9. Multivibrators		**9.1-9.18**
9.1	Introduction	9.1
9.2	Transistor as a Switch	9.1
9.3	Fixed Current Bias Transistor Switching Circuit	9.4
9.4	Multivibrators	9.5
9.5	Astable Multivibrators	9.7

9.6 Monostable Multivibrators 9.12
9.7 Bistable Multivibrator 9.15
Questions *9.18*

10. Three Terminal Devices **10.1-10.47**
10.1 Introduction 10.1
10.2 Unijuction Transistor (UJT) 10.1
10.3 UJT Relaxation Oscillator 10.6
10.4 Field-Effect Transistor 10.8
10.5 The Pinch-Off Effect in FET 10.14
10.6 JFET Parameters 10.16
10.7 Variation of Transconductance with Drain Current and Gate Voltage 10.18
10.8 Channel JFET 10.18
10.9 Essentials of FET Modelling 10.19
10.10 D.C. Load Line and Bias Point 10.23
10.11 FET Biasing 10.24
10.12 FET Small Signal Models 10.28
10.13 Common Source JFET Amplifier 10.31
10.14 Common Drain JFET Amplifier, or Source Follower 10.33
10.15 Common Gate JFET Amplifier 10.35
10.16 Applications of FETs 10.37
10.17 MOSFET 10.37
10.18 Comparison between NMOS and PMOS 10.40
10.19 Comparison between JFETs and MOSFETs 10.40
10.20 Silicon Controlled Rectifier 10.42
Questions *10.46*

11. Modulation and Demodulation **11.1-11.54**
11.1 Introduction 11.1
11.2 Modulation 11.2
11.3 Amplitude Modulation 11.4
11.4 Types of Amplitude Modulation 11.13
11.5 Vestigial Sideband (VSB) Modulation Systems 11.15
11.6 Comparison of Various AM Systems 11.16
11.7 Basic Principle of AM Generation 11.17
11.8 Block Diagram of AM (Amplitude Modulator) Transmitter 11.17
11.9 Transistor AM Modulators 11.18

11.10	Square Law Diode Modulator	11.23
11.11	Generation of DSBSC Signal	11.24
11.12	Single Sideband Generator	11.26
11.13	Frequency Modulation	11.29
11.14	Comparison of FM and AM	11.35
11.15	Phase Modulation	11.35
11.16	Pulse Modulation Systems	11.39
11.17	Comparison between FM and PM	11.45
11.18	PM Signal Generation	11.46
11.19	Demodulation or Detection	11.47
11.20	AM Detection	11.47
11.21	AM Demodulators (or Detectors)	11.48
	Questions	*11.53*
12. Noise		**12.1-12.28**
12.1	Introduction	12.1
12.2	Noise: Classification and Origin	12.1
12.3	Generalisation of Thermal Noise Relationship	12.14
12.4	Thermal Noise Calculation for a Single Noise Source	12.16
12.5	White Noise and its Mathematical Representation	12.17
12.6	Multiple Noise Sources: Superposition of Power Spectral Density	12.18
12.7	Excess Resistor Noise	12.19
12.8	Noise in Active Circuits	12.20
12.9	Noise Figure of an Amplifier	12.21
12.10	Available Noise Power Spectral Density	12.24
12.11	Available Power Density	12.25
12.12	Available Noise Power Density of an R-L-C Network	12.26
12.13	Effective Noise Temperature	12.27
	Questions	*12.28*
Index		*I.1-I.4*

1 Circuit Analysis

Electric circuits are the foundation on which the study of most electrical phenomena are based. Generally, electric circuits involve networks with voltages, currents, resistances, capacitances, inductances etc. Before considering methods of network analysis, we shall review some basics of electric currents in the following sections.

1.1 OHM'S LAW, RESISTIVITY AND CONDUCTIVITY

Ohm's Law This law applies to electric current through good conductors and may be stated as follows.

The ratio of potential difference (V) between any two points on a conductor to the current flowing between them is constant, provided, the temperature of the conductor does not change.

i.e., $$\frac{V}{I} = \text{constant} \equiv R$$

the resistance of the conductor between the two points considered. The resistance R

(i) varies directly as length l of the conductor
(ii) varies inversely as cross-section A
(iii) depends on the nature of the material
(iv) depends on the temperature of the conductor

If temperature is kept constant, we can say

$$R \propto \frac{l}{A} \quad \text{or} \quad R = \rho \frac{l}{A}$$

where ρ is a constant depending on the nature of the material and in knows as its resistivity. The unit of resistivity is ohm-metre (Ω-m) or ohm-cm (Ω-cm).

Conductance G is reciprocal of resistance.* Whereas resistance of a conductor measures the *opposition* to the flow of current, the conductance measures the *inducement* which it offers to the flow.

* In a.c. circuits, it is defined slightly differently

$$R = \rho \frac{l}{A}; \quad G = \frac{1}{R} = \frac{A}{\rho l} = \sigma A/l$$

where σ is the conductivity of the conductor. The unit of conductance is Siemens (S). Earlier this unit was called mho.

Conductivity $$\sigma = G \frac{l}{A} = \frac{G \text{ Siemens} \times l \text{ metre}}{A \text{ metre}^2}$$

$$= G \frac{l}{A} \text{ Siemens} \times \text{metre} \quad \text{or} \quad G \frac{l}{A} \text{ (S/m)}.$$

1.2 RESISTANCES IN SERIES

Figure 1.1 shows three resistances R_1, R_2 and R_3 in series. There is a progressive fall in potential as we go from point A to D.

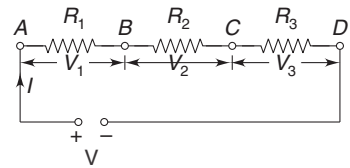


Fig. 1.1

$$V = V_1 + V_2 + V_3$$

or $$IR = IR_1 + IR_2 + IR_3$$

(from Ohms law)

or $$R = R_1 + R_2 + R_3$$ (R is the equivalent resistance)

Also $$\frac{1}{G} = \frac{1}{G_1} + \frac{1}{G_2} + \frac{1}{G_3}$$ (i.e. resistances are additive)

According to voltage divider rule, various voltage drops are

$$V_1 = V \frac{R_1}{R}, \quad V_2 = V \frac{R_2}{R}; \quad V_3 = V \frac{R_3}{R}$$

1.3 RESISTANCES IN PARALLEL

Figure 1.2 shows three resistances R_1, R_2, R_3 in parallel:

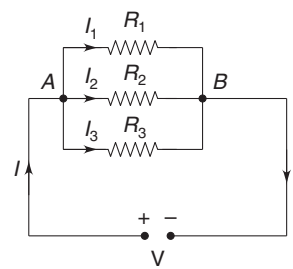


Fig. 1.2

Total current $$I = I_1 + I_2 + I_3$$

or $$\frac{V}{R} = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$$

$\therefore$ $$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

where R is the equivalent resistance between A and B. Also

$$G = G_1 + G_2 + G_3$$

i.e., conductances are additive.

1.4 DIVISION OF CURRENT IN PARALLEL CIRCUITS

Figure 1.3, two resistances are joined in parallel across a voltage V. The current in each branch as given by Ohm's law is

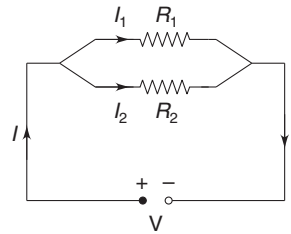


Fig. 1.3

$$I_1 = \frac{V}{R_1},\ I_2 = \frac{V}{R_2}$$

$$\therefore \quad \frac{I_1}{I_2} = \frac{R_2}{R_1}$$

As $\frac{1}{R_1} = G_1$ and $\frac{1}{R_2} = G_2$

$$\therefore \quad \frac{I_1}{I_2} = \frac{G_1}{G_2}$$

i.e., the division of current in the branches of a parallel circuit is directly proportional to the ratio of the conductances and inversely proportional to the resistances. We may also express the branch currents in terms of the total current:

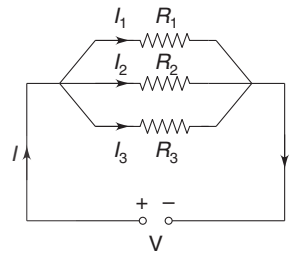


Fig. 1.4

$$I_1 + I_2 = I,$$

$$\therefore \quad I_2 = I - I_1$$

or $$\frac{I_1}{I - I_1} = \frac{R_2}{R_1}$$

or $$I_1 R_1 = R_2 (I - I_1)$$

or $$I_1 (R_1 + R_2) = R_2 I$$

or $$I_1 = I \cdot \frac{1/G_2}{\frac{1}{G_1} + \frac{1}{G_2}} = \frac{G_1 G_2}{G_2} \frac{1}{G_1 + G_2}$$

or $$I_1 = I \frac{G_1}{G_1 + G_2}$$

and $$I_2 = I - I_1 = I \left[1 - \frac{G_1}{G_1 + G_2}\right]$$

or $$I_2 = I \frac{G_2}{G_1 + G_2}$$

This current divider rule has direct application in solving electric circuits by Norton's theorem.

Now, consider the case of three resistors in parallel connected across a voltage V (Fig. 1.4).

Total current $I = I_1 + I_2 + I_3$. Let the equivalent resistance be R. Then

$$V = IR$$

Also $$V = I_1 R_1$$

$\therefore$ $$IR = I_1 R_1$$

or $$\frac{I}{I_1} = \frac{R_1}{R}$$

or $$I_1 = \frac{IR}{R_1} \tag{1.1}$$

Now, $$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

or $$R = \frac{R_1 R_2 R_3}{R_2 R_3 + R_3 R_1 + R_1 R_2}$$

Using Eq. (1.1)

$$I_1 = I \left(\frac{R_2 R_3}{R_1 R_2 + R_2 R_3 + R_3 R_1} \right)$$

$$= I \frac{G_1}{G_1 + G_2 + G_3}$$

Similarly,

$$I_2 = I \left(\frac{R_1 R_3}{R_1 R_2 + R_2 R_3 + R_3 R_1} \right)$$

$$= I \frac{G_2}{G_1 + G_2 + G_3}$$

$$I_3 = I \left(\frac{R_1 R_2}{R_1 R_2 + R_2 R_3 + R_3 R_1} \right)$$

$$= I \frac{G_3}{G_1 + G_2 + G_3}$$

1.5 DUALITY BETWEEN SERIES AND PARALLEL CIRCUITS

There is a certain peculiar pattern of relationship between series and parallel circuits. For instances, in a series circuit, current is the same whereas in a parallel circuit, voltage is the same. Also, in a series circuit, individual voltages are added and in a parallel circuit, individual currents are added. Thus, while comparing series and parallel circuits, we see that voltage takes the place of current and current takes the place of voltage. Such a behaviour is known as *duality* and the two circuits are said to be duals of each other. (see Table 1.1).

Table 1.1

Series Circuit	Parallel Circuit
$I_1 = I_2 = I_3 = \ldots$	$V_1 = V_2 = V_3 = \ldots$
$V_T = V_1 + V_2 + V_3 + \ldots$	$I_T = I_1 + I_2 + I_3 + \ldots$
$R_T = R_1 + R_2 + R_3 + \ldots$	$G_T = G_1 + G_2 + G_3 + \ldots$
$I = \frac{V_1}{R_1} = \frac{V_2}{R_2} = \frac{V_3}{R_3} = \ldots$	$V = \frac{I_1}{G_1} = \frac{I_2}{G_2} = \frac{I_3}{G_3} = \ldots$
Voltage Divider Rule:	Current Divider Rule:
$V_1 = V_T \frac{R_1}{R_T}$, $V_2 = V_T \frac{R_2}{R_T}$ etc.	$I_1 = I_T \frac{G_1}{G_T}$, $I_2 = I_T \frac{G_2}{G_T}$ etc.

1.6 DUALITY IN A.C. CIRCUITS

Ordinarily a voltage is considered the driving force, and a current the response of the circuit. However, except for a mental difficulty induced by habit, there is no reason why a current cannot be considered the driving force, and a voltage the response. The reversal of philosophy here is merely that due to the interchange of the dependent and independent variables in the problem, and leads to the concept of *duality* in networks.

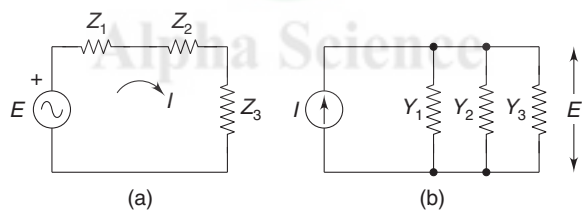


Fig. 1.5 *Dual networks*

Some dual relations involving exchange of current for voltage are illustrated by the following pairs of equations:

$$e = R_i \qquad i = Ge$$

$$e = L\, di/dt \qquad i = C\, de/dt$$

$$e = \frac{1}{C} \int i\, dt \qquad i = \frac{1}{L} \int e\, dt$$

These show that resistance and conductance, inductance and capacitance, also have dual relationships. Duality also appears as a relation between networks. Referring to the simple circuit of Fig. 1.5(a), the loop equation is

$$I\,(Z_1 + Z_2 + Z_3) = E \tag{1.2}$$

whereas the node voltage equation for (b) in

$$E\,(Y_1 + Y_2 + Y_3) = I \tag{1.3}$$

Since these relations are of identical mathematical form, but with interchanged dependent and independent variables, they are *dual networks*. Because of the identical mathematical form they must have identical *forms of solutions*. That is, (a) will behave with respect to element voltage as (b) does with respect to element current.

The voltage drops in one circuit are analogous to the branch currents in the other circuit. In one case network equations may be most easily written by summation of voltage drops, and in the other case the equations are lost easily written by summing the currents at the junctions. Thus it can be seen that the property of duality actually comes from, and is contained in, Kirchhoff's two circuit laws (Section 1.9).

The circuit-dual concept can be carried further by referring to the circuits of Fig. 1.6, where it can be shown that a parallel *GLC* circuit driven by a current source is the dual of a series *RLC* circuit excited by a voltage source. The meals equation for the series case may be written

$$Ri + L\frac{di}{dt} + \frac{1}{C}\int i\,dt = e \tag{1.4}$$

and for the dual parallel circuit,

$$Ge + C\frac{de}{dt} + \frac{1}{L}\int e\,dt = i \tag{1.5}$$

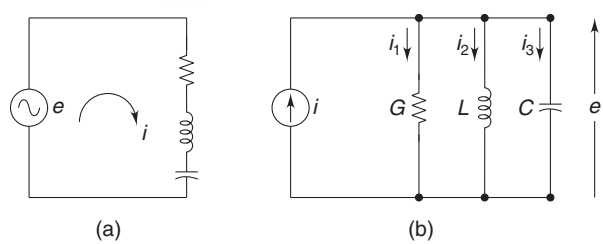


Fig. 1.6 *Series and parallel networks as dual*

illustrating the various dual relations.

It should be emphasized that duals are not equivalent in performance. Duality implies only that the equations of circuit performance are similar in mathematical form, and will behave similarly, but with interchanged variables. For instance, in (a) of Fig. 1.6, the sum of the voltages across L and C will go to zero at resonance; in (b) the sum of the currents i_2 and i_3 into the L and C branches will go to zero at parallel resonance.

Further cases of duality will be discussed as they arise.

1.7 STAR-DELTA CONNECTION

Like series and parallel connections, the resistances may be connected in 'star' (or *Y*) and 'delta' (Δ) connection as shown in Fig. 1.7(a) and 1.7(b).

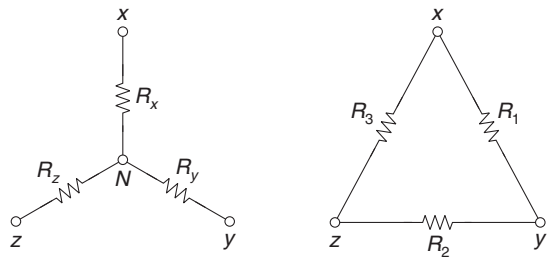


Fig. 1.7 *Connection of resistances in star and delta connection*

Circuits shown in Figs. 1.7(a) and (b) are identical provided their respective resistances from the terminals *XY*, *YZ* and *ZX* are equal.

In the star connection,

$$R_{X-Y} = R_X + R_Y$$

$$R_{Y-Z} = R_Y + R_Z$$

and $$R_{Z-X} = R_Z + R_X \tag{1.6}$$

Similarly, in delta connection, the resistance seen from the terminals *XY*, *YZ* and *ZX* are respectively.

$$\left.\begin{aligned} R_{X-Y} &= R_1 \parallel (R_2 + R_3) = \frac{R_1 (R_2 + R_3)}{R_1 + R_2 + R_3} \\ R_{Y-Z} &= R_2 \parallel (R_1 + R_3) = \frac{R_2 (R_1 + R_3)}{R_1 + R_2 + R_3} \\ R_{Z-X} &= R_3 \parallel (R_1 + R_2) = \frac{R_3 (R_1 + R_2)}{R_1 + R_2 + R_3} \end{aligned}\right\} \tag{1.7}$$

Next, we equate the resistance in star and delta across appropriate terminals. We get

$$R_X + R_Y = \frac{R_1 (R_2 + R_3)}{R_1 + R_2 + R_3} \tag{1.8}$$

$$R_Y + R_Z = \frac{R_2 (R_1 + R_3)}{R_1 + R_2 + R_3} \tag{1.9}$$

$$R_Z + R_X = \frac{R_3 (R_1 + R_2)}{R_1 + R_2 + R_3} \tag{1.10}$$

Subtracting Eq. (1.9) from Eq. (1.8), we get

$$R_X - R_Z = \frac{R_1 R_2 + R_1 R_3}{R_1 + R_2 + R_3} - \frac{R_2 R_1 + R_2 R_3}{R_1 + R_2 + R_3} \tag{1.11}$$

Then, adding Eq. (1.10) to Eq. (1.11)

$$2R_X = \frac{R_1 R_2 + R_1 R_3 - R_2 R_1 - R_2 R_3 + R_3 R_1 + R_3 R_2}{(R_1 + R_2 + R_3)}$$

$$= \frac{2 (R_1 R_3)}{R_1 + R_2 + R_3}$$

$$\therefore \quad R_X = \frac{R_1 R_3}{R_1 + R_2 + R_3}$$

In a similar way,

$$R_Y = \frac{R_1 R_2}{R_1 + R_2 + R_3}$$

and

$$R_Z = \frac{R_2 R_3}{R_1 + R_2 + R_3}$$

Thus, we see, if the resistances in Δ connected resistive network are known, we can find the equivalent star network when

$$\left.\begin{aligned} R_X &= \frac{R_1 R_3}{R_1 + R_2 + R_3} \\ R_Y &= \frac{R_1 R_2}{R_1 + R_2 + R_3} \\ R_Z &= \frac{R_2 R_3}{R_1 + R_2 + R_3} \end{aligned}\right\} \tag{1.12}$$

R_X, R_Y and R_Z being the equivalent resistances in the star network of the resistances connected in delta network and designated as R_1, R_2 and R_3.

Let us now multiply each two Eqs. of (1.12) and add the three. We get

$$R_X R_Y + R_Y R_Z + R_Z R_X = \frac{R_1^2 R_2 R_3 + R_2^2 R_1 R_3 + R_3^2 R_1 R_2}{(R_1 + R_2 + R_3)^2} \tag{1.13}$$

Let the Eq. (1.13) be divided by R_X. Then we get

$$R_Y + R_Z + \frac{R_Y R_Z}{R_X} = \frac{R_1 R_2 R_3 (R_1 + R_2 + R_3)}{R_X (R_1 + R_2 + R_3)^2}$$

$$= \frac{R_1 R_2 R_3}{(R_1 + R_2 + R_3) R_X}$$

or

$$R_Y + R_Z + \frac{R_Y R_Z}{R_X} = \frac{R_1 R_2 R_3}{(R_1 + R_2 + R_3)} \times \frac{(R_1 + R_2 + R_3)}{R_1 R_3}$$

$$= R_2 \quad \text{[Substituting the value of } R_X \text{ from Eq. (1.12)]}$$

In a similar way, dividing Eq. (1.13) by R_Y we get

$$R_X + R_Z + \frac{R_Z R_X}{R_Y} = R_3$$

and $$R_Y + R_X + \frac{R_X R_Y}{R_Z} = R_1$$

Thus, we get find R_1, R_2 and R_3 (i.e. the equivalent Δ network provided R_X, R_Y and R_Z are known in a given star network.

The equations are given as

$$\left.\begin{aligned} R_1 &= R_X + R_Y + \frac{R_X R_Y}{R_Z} \\ R_2 &= R_Y + R_Z + \frac{R_Y R_Z}{R_X} \\ R_3 &= R_X + R_Z + \frac{R_X R_Z}{R_Y} \end{aligned}\right\} \quad (1.14)$$

Example 1.1 *With reference to Fig. 1.8 convert Π connected resistors to equivalent T connection.*

Solution Figure 1.8 represents the Π connection of three resistors. The component resistors of equivalent *T* connection (Fig. 1.9) are shown below following the principle of Δ to star conversion.

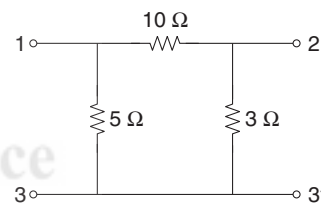


Fig. 1.8

Here, $$R_X = \frac{10 \times 5}{10 + 5 + 3} = 2.78\ \Omega;$$

$$R_Y = \frac{10 \times 3}{10 + 5 + 3} = 1.67\ \Omega;$$

$$R_Z = \frac{5 \times 3}{10 + 5 + 3} = 0.83\ \Omega.$$

It may be noted that star connection is frequently called as *Y* (Y_{ve}) connection while delta connection is also called as mesh connection. In case, the resistances are replaced by impedances, the result remains the same i.e., only the resistances in the Eqs. (1.12) and (1.14) are to be replaced by respective impedances.

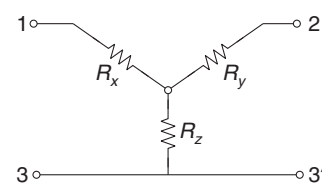


Fig. 1.9

Also, observing Figs. 1.10(a) and (b), it may be indicated in this context that the star connection may also be termed as *T* connection while the delta connection can be termed as Π (pie) connection.

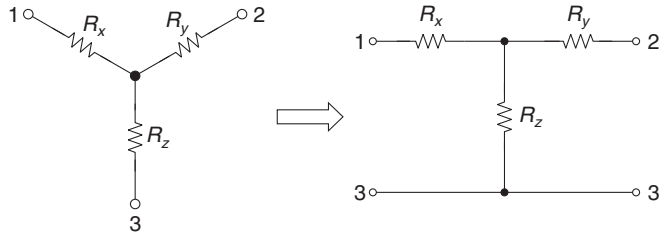


Fig. 1.10 *(a) T connection*

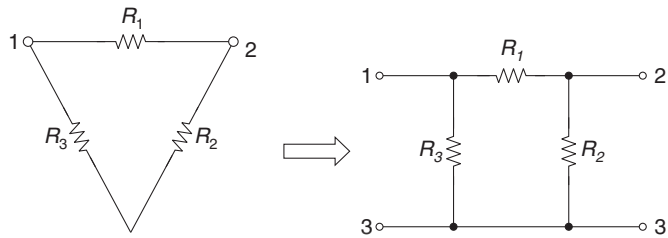


Fig. 1.10 *(b) Pie connection*

There are certain theorems, which when applied to solve electric networks, either simplify the network itself or render their analytical solution very easy. These theorems can also be applied to an ac system with the only difference that impedances replace the ohmic resistances of dc system. Before discussing network analysis it is desirable to introduce certain network definitions which are described below.

1.8 NETWORK DEFINITIONS

1. **Circuit** A circuit is closed conducting path through which an electric current either flows or is intended to flow.
2. **Circuit Elements or Parameters** Any individual circuit component (inductor, resistor, capacitor, generator, etc.) with two terminals by which it may be connected to other electric components is called a circuit element or parameter (These parameters may be lumped or distributed).
3. **Lumped Networks** One in which physically separate resistors, inductors, capacitors can be represented.
4. **Distributed Network** A network in which the resistors, inductors, capacitors, cannot be electrically separated and individually isolated as separate elements. A transmission line is such a network.
5. **Branch** A group of elements, usually in series and having two terminals.

6. **Potential Source** A hypothetical generator which maintains its value of potential independent of the output current. As an ac source, it will be indicated by a circle enclosing a wavy line.
7. **Current Source** A hypothetical generator which maintains an output current independent of the voltage across its terminals. The current source will be indicated by a circle enclosing an arrow, the arrow indicating the assumed reference current direction.
8. **Linear Circuit** A linear circuit is one whose parameters are constant i.e., they do not change with voltage or current.
9. **Non-linear Circuit** It is that circuit whose parameters change with voltage or current.
10. **Bilateral Circuit** A bilateral circuit is one whose properties or characteristics are the same in either direction. The usual transmission line is bilateral, because it can be made to perform its function equally well in either direction.
11. **Unilateral Circuit** It is that circuit whose properties or characteristics change with the direction of its operation. A diode rectifier is a unilateral circuit, because it cannot perform rectification in both directions.
12. **Network** An electric network is any inter-connection of circuit elements or branches. In Fig. 1.11(a), is shown a complicated network of elements or impedances and sources, where each branch may include *R L C* or other types of elements. Such a network having two distinct pairs of terminals is called a *four-terminal* network. If one of the 1, 1 terminals is common to the 2, 2 pair, as at (b), the circuit is a *three-terminal* network, and if the 2, 2 terminals are short circuited, it becomes a *two-terminal* network.

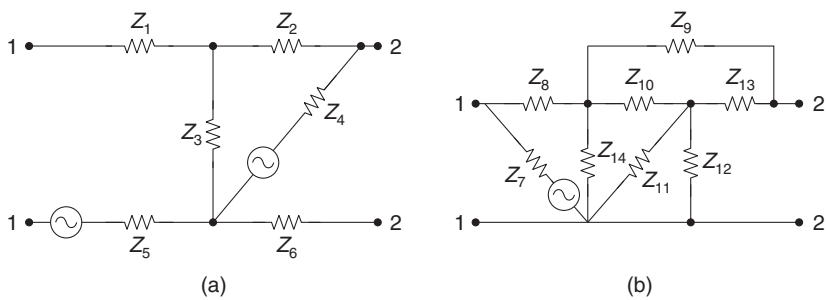


Fig. 1.11 *Possible forms of networks: (a) a four-terminal network; (b) a three-terminal network*

13. **Passive Network** A network containing circuit elements without energy sources.

14. **Active Network** A network containing generators or energy sources as well as other elements.
15. **Linear Element** A circuit element is linear if the relation between current and voltage involves a constant coefficient as in

$$e = Ri, \quad e = L \cdot \frac{di}{dt}, \quad e = \frac{1}{C} \int i \, dt$$

Iron-cored reactors and incandescent lamps are examples of elements that are not linear, or in which the coefficient is a function of current.

16. **Linear Networks** Linear networks are those in which the differential equation relating the instantaneous current and voltage is a linear equation with constant coefficients.
17. **Node (or Junction)** A terminal of any branch of network, or a terminal common to two or more branches.
18. **Loop** A closed path in a circuit in which no element or node is encountered more than once.
19. **Mesh** It is a loop that contains no other loop within it. For example, the circuit of Fig. 1.12 has seven branches, six nodes, three loops and two meshes.

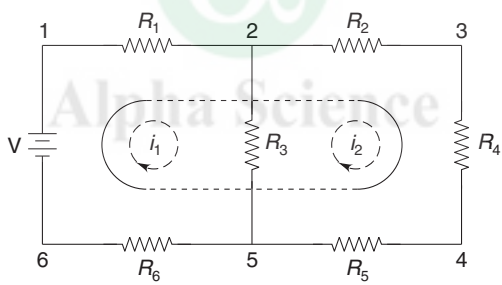


Fig. 1.12

We will now discuss various network theorems which are of great help in solving complicated networks (A network is said to be completely solved when all voltages and all currents in its different elements are determined). There are two general approaches to network analysis:

(i) Direct Method Here, the network is left in its original form while determining its different voltages and currents. Such methods are usually restricted to fairly simple circuits and include Kirchhoff's laws, Loop-analysis, superposition theorem, Nodal analysis, Reciprocity theorem etc.

(ii) Network Reduction Method Here, the original network is converted into a much simpler equivalent circuit for rapid calculation of different

quantities. This method can be applied to simple as well as complicated networks. Examples of this method are Delta/Star and Star/Delta conversions, Thevenin's Theorem and Norton's Theorem etc.

1.9 KIRCHHOFF'S LAWS

These laws are more comprehensive than Ohm's law and are used for solving electrical networks which may not be readily solved by Ohm's law. There are two Kirchhoff's laws, the first law deals with flow of current and is popularly known as Kirchhoff's current law (KCL) while the second one deals with voltage drop in a closed network and is known as Kirchhoff's voltage law (KVL).

1.9.1 Kirchhoff's Current Law

The algebraic sum of currents at any node of a circuit is zero. In Fig. 1.13 as per KCL

$$i_1 + i_2 - i_3 - i_4 - i_5 = 0$$

(the direction of incoming currents to a node being positive, the outgoing currents should be taken negative).

Thus $i_1 + i_2 = i_3 + i_4 + i_4$

i.e., the algebraic sum of currents entering a node must be equal to the algebraic sum of currents leaving a node.

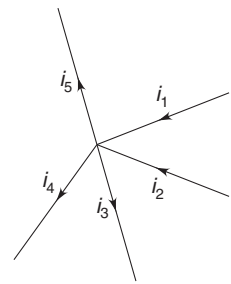


Fig. 1.13

Example 1.2 *Find the magnitude and direction of the unknown currents in Fig. 1.14. Given i_1 = 10 A, i_2 = 6 A, i_5 = 4 A.*

Solution By observation, it is evident that

$$i_1 = i_7$$

$\therefore$ $$i_7 = 10 \text{ A}$$

At node A, from KCL,

$$i_1 = i_2 + i_4$$

or $$10 = 6 + i_4$$

or $$i_4 = 4 \text{ Amp.}$$

At node B, utilizing KCL,

$$i_2 = i_3 + i_5$$

or $$i_3 = i_2 - i_5 = 6 - 4 = 2 \text{ Amp}$$

Similarly, at node C,

$$i_7 = i_5 + i_6$$

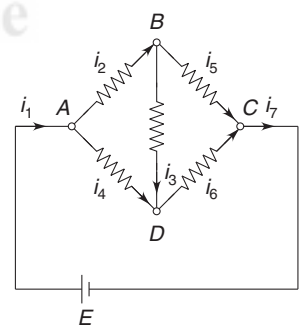


Fig. 1.14

$$\therefore \quad i_6 = i_7 - i_5$$

$$= 10 - 4$$

or $$i_6 = 6 \text{ Amp}$$

Thus, all the unknown currents of the problem have been determined. Then, branch currents are

$$i_1 = i_7 = 10 \text{ A}$$

$$i_2 = 6 \text{ A}$$

$$i_3 = 2 \text{ A}$$

$$i_4 = 4 \text{ A}$$

$$i_5 = 4 \text{ A}$$

$$i_6 = 6 \text{ A}$$

Note that the same problem can be solved with the assumed direction of currents reversed, and it would not influence the final magnitude and direction of currents.

Current Division

A parallel circuit acts as a current divider as the current divides in all branches in a parallel circuit. In Fig. 1.15, the current I has been divided into I_1 and I_2 in two parallel branches with resistances R_1 and R_2 while V represents the drop across R_1 or R_2.

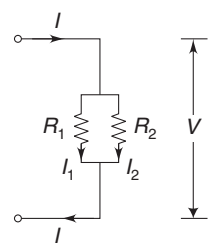


Fig. 1.15

Obviously $I_1 = \dfrac{V}{R_1}$ and $I_2 = \dfrac{V}{R_2}$. Let R represents the total resistance and is given by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

or $$R = \frac{R_1 R_2}{R_1 + R_2}$$

Also $$I = V/R = \frac{V(R_1 + R_2)}{R_1 R_2} \qquad (1.15)$$

But $V = I_1 R_1 = I_2 R_2$

$$\therefore \quad I = \frac{I_1 R_1 (R_1 + R_2)}{R_1 R_2} = \frac{I_1}{R_2}(R_1 + R_2) \qquad (1.16)$$

Also $$I = \frac{I_2 R_2 (R_1 + R_2)}{R_1 R_2} = \frac{I_2}{R_1}(R_1 + R_2) \qquad (1.17)$$

Thus Eq. (1.16)

$$I_1 = I\left(\frac{R_2}{R_1 + R_2}\right) \tag{1.18}$$

and Eq. (1.17)

$$I_2 = I\left(\frac{R_1}{R_1 + R_2}\right) \tag{1.19}$$

Thus, it may be said that the current in any of the parallel branches is equal to the ratio of opposite branch resistance to the total resistance multiplied by the total current.

Example 1.3 *In Fig. 1.16, find "v" and the magnitude and direction of the unknown currents in the branches xn, yn and zn.*

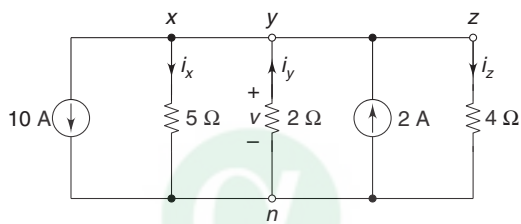


Fig. 1.16

Solution Let the unknown currents in branches *xn*, *yn* and *zn* be i_x, i_y and i_z respectively and their directions have been shown in Fig. 1.16.

At node "*y*" form KCL,

$$10 + i_x + i_z = i_y + 2$$

or
$$i_z - i_y + i_z = -8 \tag{i}$$

Again, from Ohm's Law,

$$i_x = \frac{v}{5}\text{ A},$$

$$i_y = -\frac{v}{2}\text{ A}$$

and
$$i_z = \frac{v}{4}\text{ A}$$

Then, from (i),

$$\frac{v}{5} + \frac{v}{2} + \frac{v}{4} = -8.$$

or
$$v = -8 \times \frac{20}{19} = -8.42\text{ V}$$

[– ve sign of *v* indicates node *n* to be + ve and at higher potential]

$\therefore$ $$v = -8.42 \text{ V}$$

$$i_x = -\frac{8.42}{5} = -1.684 \text{ A}$$

(i.e. current flows from "n" to "x")

$$i_y = -\frac{(-8.42)}{2} = 4.21 \text{ A}$$

(i.e. current flows from "n" to "y")

$$i_z = -\frac{8.42}{4} = -2.1 \text{ A}$$

(i.e. current flows from "n" to "z")

Figure 1.17 represents actual direction of currents in the redrawn circuit.

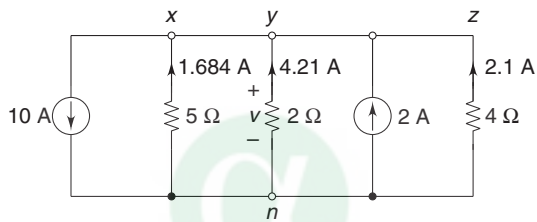


Fig. 1.17

Example 1.4 *What is the power absorbed in the resistors of Fig. 1.18? Assume* $i_0 = 1$ *A.*

Also find the current from 2 V source.

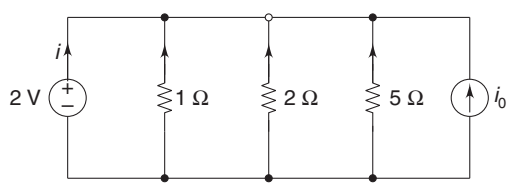


Fig. 1.18

Solution Let us redraw the circuit as shown in Fig. 1.19 with the currents i_1, i_2, and i_5 as assumed through the respective resistances.

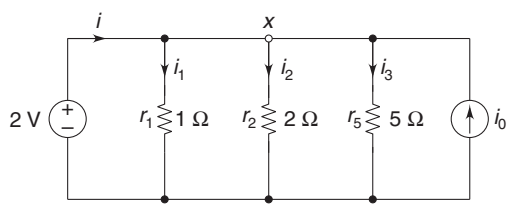


Fig. 1.19

Obviously, $i_1 = \frac{2}{1} = 2$ A; $i_2 = \frac{2}{2} = 1$ A; $i_5 = \frac{2}{5} = 0.4$ A.

Also the powers absorbed by the respective resistors are as follows:

for 1 Ω resistor: $p_1 = i_1^2 \times r_1 = 4$ W

for 2 Ω resistor: $p_2 = i_2^2 \times r_1 = 2$ W

for 5 Ω resistor: $p_5 = i_5^2 \times r_5 = 0.8$ W

Then the total power absorbed in the circuit is (4 + 2 + 0.8) W or 6.8 W.

Again, applying KCL at node "x",

$$i + i_0 = i_1 + i_2 + i_5 = 3.4 \text{ A}.$$

As it is, given that $i_0 = 1$ A, hence,

$$i = 3.4 - i_0 = 2.4 \text{ A}.$$

∴ We finally obtain the power absorbed in resistors as follows:

$$p_1 = 4 \text{ W}, \quad p_2 = 2 \text{ W}$$

and $p_5 = 0.8$ W while the voltage source current is 2.4 A.

[**Check:** the power delivered by the voltage source is vi i.e., 4.8 W. The power delivered by the current source is vi_0, i.e., 2 W. The total source power is then 4.8 + 2.0 = 6.8 W while the net power consumed has been obtained as 6.8 W. This shows conservation of power].

1.9.2 Kirchhoff's Voltage Law (KVL)

The algebraic sum of voltages (or voltage drops) in any closed path of network that is traversed in a single direction is zero.

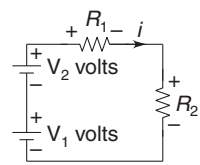


Fig. 1.20 *Application of KVL with series connected voltage sources*

Both the figures (Fig. 1.20 and Fig. 1.21) show closed circuits (also termed as mesh). As per KVL, in Fig. 1.20

$$-V_1 + (-V_2) + iR_1 + iR_2 = 0 \tag{1.20}$$

[Here, the assumed current i causes a + ve drop of voltage when flowing from + ve to – ve potential while the current flowing from – ve to + ve potential]

or $$i(R_1 + R_2) = V_1 + V_2$$

giving $$i = \frac{V_1 + V_2}{(R_1 + R_2)}$$

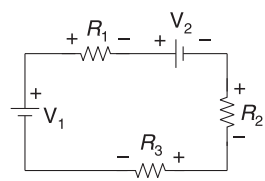


Fig. 1.21 *Application of KVL while voltage sources are connected in opposite polarity*

On the other hand in Fig. 1.21, from KVL, with assumed direction of the circuit current i,

$$-V_1 + iR_1 + V_2 + iR_2 + iR_3 = 0 \tag{1.21}$$

or $$i\,(R_1 + R_2 + R_3) = V_1 - V_2$$

giving $$i = \frac{V_1 - V_2}{R_1 + R_2 + R_3}$$

It may be noted that the voltage V_1 is – ve in both the Eqs. (1.20) and (1.21) while V_2 is – ve in Eq. (1.20) but + ve in Eq. (1.21). This is because of the fact that in Fig. 1.20 and in Fig. 1.21, with assumed directions of current, the current in the source (V_1) flows from – ve to + ve polarity while for V_2, the current flows from – ve to + ve polarity in Fig. 1.20 and from + ve to – ve polarity in Fig. 1.21.

In this context it may be noted that for dependent sources in that circuit, KVL can also be applied. *In case of calculation of power of any source, when the current enters the source, the power is absorbed by the source while the source delivers the power if the current is coming out of the source.*

1.9.3 Voltage Division in Series Circuits

Let us consider a voltage source E with resistance r_1 and r_2 in series across it (Fig. 1.22).

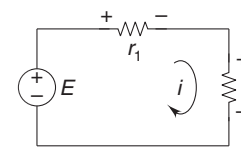


Fig. 1.22

Obviously, the current in the loop will be

$$i = \frac{E}{r_1 + r_2}$$

and voltage across $$r_1 = i\,r_1 = \frac{Er_1}{r_1 + r_2}$$

while voltage across $$r_2 = i\,r_2 = \frac{Er_2}{r_1 + r_2} \tag{1.22}$$

Thus, the voltage across a resistor in series circuit is equal to the value of that resistor times the total impressed voltage across the series elements divided by the total resistance of the series elements.

Example 1.5 *Design a voltage divider circuit with two resistances r_1 and r_2 such that the drop v_1 (drop across r_1) is twice the drop v_2 (drop across r_2) in the circuit of Fig. 1.23.*

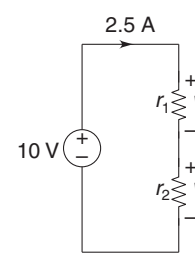


Fig. 1.23

Solution Using Ohm's Law,

$$(r_1 + r_2) = r_t = \frac{10}{2.5} = 4\ \Omega$$

[$\because v = 10$ volts, $i = 2.5$ Amps.]

Since $$v_1 = 2v_2,$$

$$r_1 = 2r_2$$

Hence, $r_t = r_1 + r_2$

$= 2r_2 + r_2 = 3r_2$

However, r_t being 4 Ω,

$$r_2 = \frac{4}{3} = 1.33\ \Omega$$

and $r_1 = 2r_2 = 2.66\ \Omega$

$\therefore$ $r_1 = 2.66\ \Omega$ while $r_2 = 1.33\ \Omega$

[Check: with $r_1 = 2.66\ \Omega$, $v_1 = 2.5 \times 2.66 = 6.65$ V and $v_2 = 1.33 \times 2.5 = 3.325$ V. It is then evident that $v_1 = 2v_2$]

1.10 REDUCTION OF A COMPLICATED NETWORK

Many circuits will be found complicated and difficult to analyse. If they can be reduced or simplified it is frequently possible better to understand their performance. This reduction can be accomplished through use of certain types of simple networks that are *equivalent* in performance.

One passive network is said to be equivalent to a second network if the second can be substituted for the first without change in currents and voltages appearing at the network terminals. It must be emphasized that an equivalent network is not identical internally but has only identical values of external voltages and currents at the terminals. Thus two networks are equivalent if each can be placed in a box with terminals and if the boxes can be substituted one for the other in a circuit without change in the circuit operation. Thus it may be possible to find a simple network which may be substituted for a more complicated network.

Assume that the box of Fig. 1.24(a) contains a linear passive electric network, no matter how complicated in internal connection. Certain external terminals may be designated 1,1 and 2,2; and impedance measurements may be made at these terminals. The input impedance may be measured at either pair of terminals with any selected value of impedance connected to the other pair. Convenient impedance values which may be chosen are open circuit and short-circuit. The impedance measured across 1,1 terminals with the 2,2 terminals open-circuited may be designated Z_{1oc}.

Likewise, the measurement at the 1,1 terminals with the 2,2 terminals short-circuited might be called Z_{1sc}. Similar measurements taken at the 2,2 terminals would result in Z_{2oc} and Z_{2sc}, giving four such possible measurements.

Since these are measurements of terminal impedance, any other network which can be found with identical measurements should be equivalent. The simplest form for a network equivalent to the one in the box may now be

considered. There are eight quantities, the magnitudes and phase angles of V_1, V_2, I_1, and I_2, which must be established by the new network, in order that the conditions for equivalence be satisfied. However, if a load be connected to the 2,2 terminals and called Z_R, then

$$V_2 = I_2 Z_R$$

so that although there are apparently four phase angles and four magnitudes to be specified, only six of these are functions of the network, one magnitude and one phase angle being determined by the load impedance Z_R. With three magnitudes and three phase angles to be determined, a minimum of three adjustable impedances is needed to set up a network which may be equivalent to that in the box. Only two arrangements of three-element circuits are possible; these are shown at (b) and (c) of Fig. 1.24 as the T (Y) and π (delta) networks.

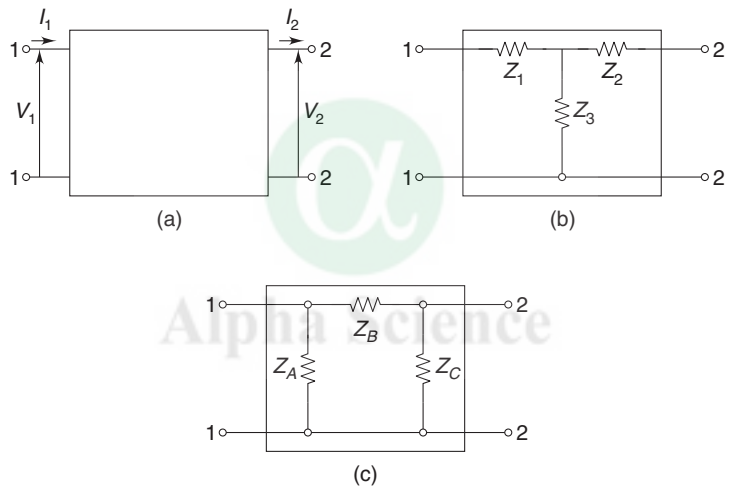


Fig. 1.24 *(a) Any complicated network, with terminal voltages and currents indicated; (b) a T network which may be made equivalent to the network in the box of (a); (c) a π network equivalent to (b) and (a).*

Consider the T network of (b), Fig. 1.24 as a possible simplified circuit, that by proper selection of values of Z_1, Z_2, and Z_3 might be made to have the same external currents and voltages as the original network of (a). The T network would then be said to be equivalent to the network at (a). If the T section is equivalent, the open- and short-circuit measurements must be identical. If these measurements are made on the T circuit, the results will be

$$Z_{1oc} = Z_1 + Z_3 \tag{1.23}$$

$$Z_{1sc} = Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3} \tag{1.24}$$

$$Z_{2oc} = Z_2 + Z_3 \tag{1.25}$$

$$Z_{2sc} = Z_2 + \frac{Z_1 Z_3}{Z_1 + Z_3} \tag{1.26}$$

A *T* equivalent for any passive network having a given set of open and short-circuit measurements can then be found by solving the above equations for Z_1, Z_2, and Z_3. Since there are four equations and only three unknowns, one equation and one measurement are not needed.

Subtracting Eq. (1.24) from Eq. (1.25),

$$Z_{1oc} - Z_{1sc} = Z_3 - \frac{Z_2 Z_3}{Z_2 + Z_3} = \frac{Z_3^2}{Z_2 + Z_3}$$

After recognition that the denominator is Z_{2oc}, from Eq. (1.25), there results

$$Z_3 = \pm \sqrt{Z_{2oc} (Z_{1oc} - Z_{1sc})}$$

Use of the above equation leads to

$$Z_1 = Z_{1oc} - Z_3 \tag{1.27}$$

$$Z_2 = Z_{2oc} - Z_3 \tag{1.28}$$

$$Z_3 = \pm \sqrt{Z_{2oc} (Z_{1oc} - Z_{1sc})} \tag{1.29}$$

as the values for a *T* networks which will be equivalent in performance to the original network. Choice of the + or – sign on the radical of Z_3 will lead to two different *T* networks. These will be equivalent as to current and voltages under the definition of equivalence, except for an ambiguity of 180° in output phase angle. That this is inherent in the method is illustrated by the simple expedient of reversing the output leads of the network in Fig. 1.24(a). The open and short-circuit impedances will remain unchanged, but there will be a change of 180° in the output phase conditions. Thus the method cannot resolve this ambiguity without additional data as to the phase relations of the original network.

If it is desired to find an equivalent π section, the same three measurements on the π circuit of (c), Fig. 1.24 would be

$$Z_{1oc} = \frac{Z_A (Z_B + Z_C)}{Z_A + Z_B + Z_C} \tag{1.30}$$

$$Z_{2oc} = \frac{Z_C (Z_A + Z_B)}{Z_A + Z_B + Z_C} \tag{1.31}$$

$$Z_{1sc} = \frac{Z_A Z_B}{Z_A + Z_B} \tag{1.32}$$

Multiplying Eqs. (1.31) and (1.32) gives

$$Z_{2oc} Z_{1sc} = \frac{Z_A Z_B Z_C}{Z_A + Z_B + Z_C} \tag{1.33}$$

and subtracting Eq. (1.32) from Eq. (1.30);

$$Z_{1oc} - Z_{1sc} = \frac{Z_C Z_A^2}{(Z_A + Z_B + Z_C)(Z_A + Z_B)}$$

Use of Eq. (1.31) leads to

$$Z_{2oc}(Z_{1oc} - Z_{1sc}) = \frac{Z_A^2 Z_C^2}{(Z_A + Z_B + Z_C)^2} \tag{1.34}$$

Values may then be obtained for the three π-secion branches as

$$Z_A = \frac{Z_{2oc} Z_{1sc}}{Z_{2oc} - \sqrt{Z_{2oc}(Z_{1oc} - Z_{1sc})}} \tag{1.35}$$

$$Z_B = \frac{Z_{2oc} Z_{1sc}}{\sqrt{Z_{2oc}(Z_{1oc} - Z_{1sc})}} \tag{1.36}$$

$$Z_C = \frac{Z_{2oc} Z_{1sc}}{Z_{1oc} - \sqrt{Z_{2oc}(Z_{1oc} - Z_{1sc})}} \tag{1.37}$$

These three equations permit designing a π network equivalent to any complicated network.

The results of the preceding paragraphs may now be summarized as follows:

Any linear, bilateral, passive electrical network can be represented, at a single frequency by a T or π network.

1.11 CONVERSIONS BETWEEN *T* AND π SECTIONS

It is convenient to be able to make conversions directly from *T* to equivalent π sections or vice versa, without the necessity of employing the external open- and short-circuit measurements of Section 1.10. For the *T* and π sections of (a) and (b), Fig. 1.25, these measurements are

T Section	*π Section*
$Z_{1oc} = Z_1 + Z_3$	$Z_{1oc} = \dfrac{Z_A(Z_B + Z_C)}{Z_A + Z_B + Z_C}$
$Z_{2oc} = Z_2 + Z_3$	$Z_{2oc} = \dfrac{Z_C(Z_A + Z_B)}{Z_A + Z_B + Z_C}$

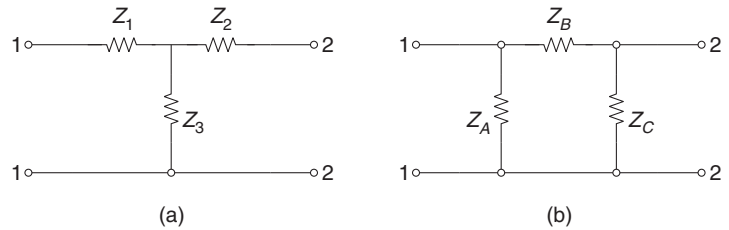


Fig. 1.25 *(a) A T network; (b) a π network which may be made equivalent to the T network of (a)*

$$Z_{1sc} = Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3} \qquad Z_{1sc} = \frac{Z_A Z_B}{Z_A + Z_B}$$

If the two circuits are to be equivalent, the external measurements must be equivalent. If the two networks were placed in boxes and only the terminals brought out, it would then be impossible to distinguish the T from the π circuit at a given frequency. Equating the measurements leads to

$$Z_1 + Z_3 = \frac{Z_A (Z_B + Z_C)}{Z_A + Z_B + Z_C} \tag{1.38}$$

$$Z_2 + Z_3 = \frac{Z_C (Z_A + Z_B)}{Z_A + Z_B + Z_C} \tag{1.39}$$

$$Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3} = \frac{Z_A Z_B}{Z_A + Z_B} \tag{1.40}$$

Subtracting Eq. (1.40) from Eq. (1.38) yields

$$Z_3 - \frac{Z_2 Z_3}{Z_2 + Z_3} = \frac{Z_A (Z_B + Z_C)}{Z_A + Z_B + Z_C} - \frac{Z_A Z_B}{Z_A + Z_B}$$

$$\frac{Z_3^2}{Z_2 + Z_3} = \frac{Z_A^2 Z_C}{(Z_A + Z_B + Z_C)(Z_A + Z_B)}$$

Equation (1.39) should be inserted for the denominator of the left-hand term, giving a value for Z_3. All branches of the T network follow as

$$Z_1 = \frac{Z_A Z_B}{Z_A + Z_B + Z_C} \tag{1.41}$$

$$Z_2 = \frac{Z_B Z_C}{Z_A + Z_B + Z_C} \tag{1.42}$$

$$Z_3 = \frac{Z_A Z_C}{Z_A + Z_B + Z_C} \tag{1.43}$$

If the impedances of the π circuit are known, the above equations may be used to determine the equivalent T network.

The inverse transformation from T to π can be obtained if Eq. (1.39) is substituted in Eq. (1.40), resulting in

$$Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1 = \frac{Z_A Z_B Z_C}{Z_A + Z_B + Z_C}$$

Successive use of Eqs. (1.43), (1.41), and (1.42) in the right-hand side gives

$$Z_A = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_2} \tag{1.44}$$

$$Z_B = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_3} \tag{1.45}$$

$$Z_C = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_1} \tag{1.46}$$

thus permitting a π circuit equivalent to a given T network to be designed.

The various relations for transformation between T and π networks can be summarized through Fig. 1.26. The impedance for a T arm may be obtained from the π by noting that the T impedance is calculated by using the product of the two adjacent π network impedances divided by the sum of all three π impedances. Likewise, a π-section impedance can be found by use of the sum of the double products of the T branches divided by the impedance of the T branch opposite the desired π arm.

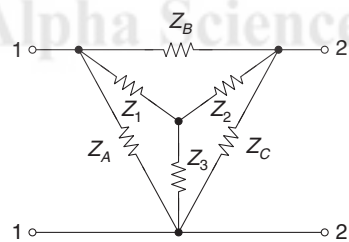


Fig. 1.26 *A diagram for use in remembering the T-to-π and the π-to-T transformation relations*

If Eqs. (1.41), (1.42), (1.43) for the branches of the T are written in terms of admittances, the following are obtained:

$$Y_1 = \frac{Y_A Y_B + Y_B Y_C + Y_C Y_A}{Y_C} \tag{1.47}$$

$$Y_2 = \frac{Y_A Y_B + Y_B Y_C + Y_C Y_A}{Y_A} \tag{1.48}$$

$$Y_3 = \frac{Y_A Y_B + Y_B Y_C + Y_C Y_A}{Y_B} \tag{1.49}$$

Comparison of these equations with those for the branches of the π network in Eqs. (1.44), (1.45), (1.46) is interesting, and shows that the *π and T networks are duals*. This is further reinforced if Eqs. (1.44), (1.45) and (1.46) are rewritten in terms of admittances, when dual equations for the *T* network are obtained.

Example 1.6 *Determine the value of R and current through it in Fig. 1.27(a), if current through branch AO is zero.*

Solution The given circuit can be redrawn as shown in Fig. 1.27(b). As seen, it is nothing else but wheatstone bridge circuit. As we know, when current through branch *AO* is zero, the bridge is said to be balanced. In that case, products of resistances of opposite arms of the bridge become equal.

$$\therefore \qquad 4 \times 1.5 = 1 \times R$$

or
$$R = 6\ \Omega.$$

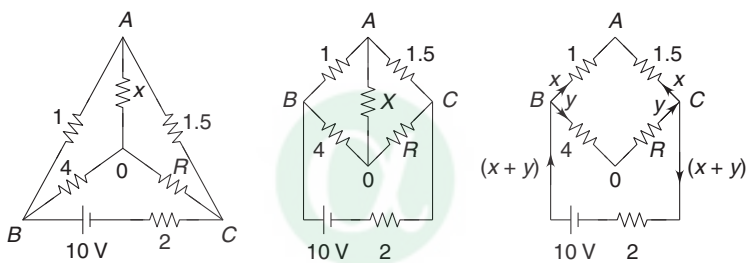


Fig. 1.27

Under condition of balance, it makes no difference if resistance *X* is removed thereby giving us the circuit of Fig. 1.27(c). Now, there are two parallel paths between points *B* and *C* of resistances (1 + 1.5) = 2.5 Ω and (4 + 6) = 10 Ω.

$$\therefore \qquad R_{BC} = 10 \parallel 2.5 = \frac{10 \times 2.5}{10 + 2.5}$$

$$= \frac{25}{12.5} = 2\ \Omega$$

Total circuit resistance = 2 + 2 = 4 Ω

Total circuit current = $\frac{10}{4}$ = 2.5 A.

This current gets divided into two parts at point *B*. The current through *R* is

$$y = 2.5 \times \frac{2.5}{12.5} = 0.5 \text{ A}.$$

Example 1.7 *In the circuit shown in Fig. 1.28, calculate the value of the unknown resistance R and the current flowing through it when the current in branch OC is zero.*

Solution if current through R-ohms is I amp, AO branch carries the same current, since, current through the branch CO is zero. This also means that the nodes C and O are at the equal potential. Then equating voltage-drops, we have

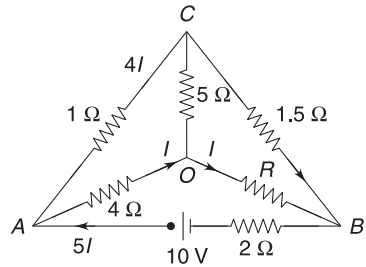


Fig. 1.28

$$V_{AO} = V_{AC}.$$

This means branch AC carries a current of 4 I. This current of 4 I also flows through the branch CB. Equating the voltage-drops in branches OB and CB,

$$1.5 \times 4\, I = R\, I$$

$$\therefore \quad R = 6\ \Omega$$

At node A, applying KCL, a current of 5 I flows through the branch BA from B to A.

Applying KVL around the loop $BAOB$,

$$I = 0.5 \text{ Amp.}$$

Example 1.8 *Obtain the branch currents in the unbalanced bridge circuit of Fig. 1.29. Also determine the voltage drop across AC and the equivalent resistance between terminals A and C in the bridge.*

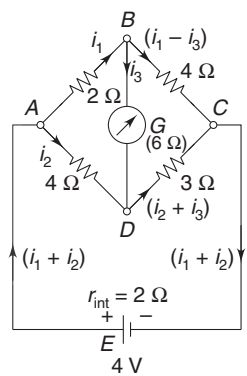


Fig. 1.29

Solution The assumed directions of currents have been shown in Fig. 1.29.

Applying KVL, in loop $ABDA$,

$$2i_1 + 6i_3 - 4i_2 = 0$$

or $$i_1 + 3i_3 - 2i_2 = 0$$

i.e., $$i_1 - 2i_2 + 3i_3 = 0 \qquad \text{(i)}$$

From loop $BCDB$, application of KVL yields

$$4\,(i_1 - i_3) - 3\,(i_2 + i_3) - 6i_3 = 0$$

or $$4i_1 - 3i_2 - 13i_3 = 0 \qquad \text{(ii)}$$

Also, using KVL, from loop $ABCEA$,

$$2i_1 + 4\,(i_1 - i_3) + 2\,(i_1 + i_2) - 4 = 0$$

or $$8i_1 + 2i_2 - 4i_3 = 4$$

i.e., $$4i_1 + i_2 - 2i_3 = 2 \qquad \text{(iii)}$$

Solving these Eqs. (i), (ii) and (iii), we get

$$i_1 = 451.64 \text{ mA}$$

$$i_2 = 322.58 \text{ mA}$$

$$i_3 = 64.51 \text{ mA}$$

and $$(i_1 + i_2) = i = 774.19 \text{ mA}.$$

Then the branch currents are as follows:

Current in branch AB $(i_1) = 451.61$ mA

Current in branch AD $(i_2) = 322.58$ mA

Current in branch BD $(i_3) = 64.51$ mA

Current in branch BC $(i_1 - i_3) = 387.\ 1$ mA

Current in branch DC $(i_2 + i_3) = 387.09$ mA

Current in external circuit $(i_1 + i_2) = 774.19$ mA

Again, the internal voltage drop of the cell being $[r_{int}\ (i_1 + i_2)]$, its value comes out to be $(2 \times 774.10)\ 10^{-3}$ i.e., 1.55 V.

Equivalent resistance between A and C being given by the ratio of p.d. between point A and C to the current between these two points, the numerical value of the equivalent resistance becomes

$$r_{equiv.\ (AC)} = \frac{2.452}{0.77419} = 3.17\ \Omega$$

1.12 THEVENIN'S THEOREM

This theorem is possibly the most extensively used network theorem. It is applicable where it is desired to determine the current through or voltage across any one element in a network without going through the rigorous method of solving a set of network equations.

1.12.1 Statement of Thevenin's Theorem*

Any two terminal bilateral linear d.c. circuit can be replaced by an equivalent circuit consisting of a voltage source and a series resistor.

* A more general statement of Thevenin's theorem is that any linear active network consisting of impendent and or dependent voltage and current source (S) and linear bilateral network elements can be repl;aced by an equivalent circuit consisting of a voltage source in series with a resistance, the voltage source being the open circuited voltage across the open circuited load terminals and the resistance being the internal resistance of the source network looking through the open circuited load terminals.

1.12.2 Explanation

Let us consider a simple d.c. circuit as shown in Fig. 1.30(a). We are to find I_L by Thevenin's theorem.

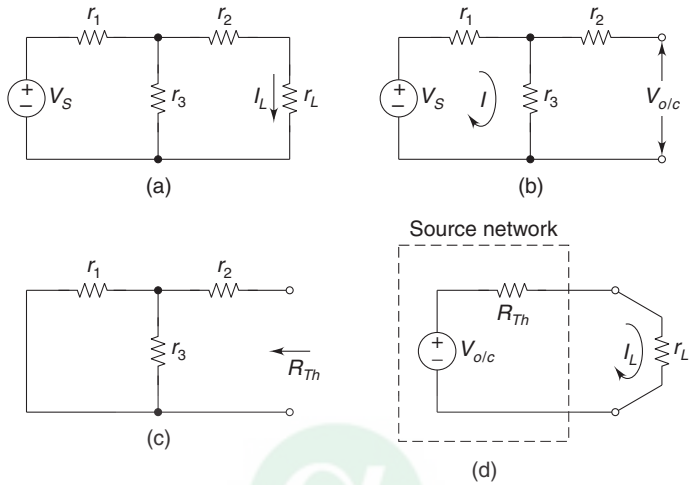


Fig. 1.30 *(a) A simple d.c. circuit, (b) finding of $V_{o/c}$, (c) finding of R_{th} (d) finding of I_L forming Thevenin equivalent circuit.*

In order to find the equivalent voltage source, r_L is removed (Fig. 1.30(b)) and V_{oc} is calculated.

$$V_{oc} = I\, r_3 = \frac{V_S}{r_1 + r_3} \cdot r_3.$$

Next, to find the internal resistance of the network (*Thevenin's resistance* or *equivalent resistance*) in series with V_{oc}, the voltage source is removed (deactivated) by a short circuit (as the source does not have any internal resistance as shown in Fig. 1.30(c).

As per Thevenin's theorem, the equivalent being Fig. 1.30(c),

$$I_L = \frac{V_{oc}}{R_{Th} + r_L}\ A.$$

1.12.3 Steps for Solving a Network Using Thevenin's Theorem

Step 1 Remove the load resistor (R_L) and find the open circuit voltage (V_{oc}) across the open circuited load terminals.

Step 2 Deactivate the constant sources (for voltage source, remove it by internal resistance and for current source, delete the source by open circuit) and find the internal resistance. (Thevenin's resistance) of the source side looking through the open circuited load terminals. Let this resistance be R_{th}.

Step 3 Obtain Thevenin's equivalent circuit (Fig. 1.31) by placing R_{th} in series with V_{oc}.

Step 4 Reconnect R_L across the load terminals as shown in Fig. 1.31.

Obviously I (the load current)

$$= \frac{V_{oc}}{R_{Th} + R_L}$$

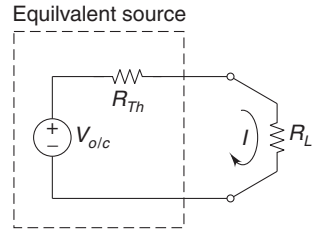


Fig. 1.31 *Thevenin's equivalent network*

1.12.4 Different Methods of Finding R_{th}

(*Thevenin's equivalent resistance or Internal impedance*)

(a) *For impendent sources*. The most common method of finding R_{Th}, the internal resistance of any linear, bilateral network containing independent current or voltage sources is to deactivate the source by internal resistance i.e., for independent current source, deactivate it by removing the source and for voltage source, deactivate it by shorting it (assuming the internal resistance of the voltage source being zero). Then find the internal resistance of the network looking through the load terminals kept open circuited. This method is illustrated in the following examples.

(b) *For the circuits containing dependent sources in addition to or in absence of independent source.*

1st Method Find V_{oc} across the open circuited load terminals by conventional network analysis. Next, short the load terminals and find the short-circuit (I_{sc}) through the shorted terminals.

The internal resistance of the source network is then obtained as

$$R_{Th} = \frac{V_{oc}}{I_{sc}}$$

2nd Method Remove the load resistance and apply a d.c. driving voltage v_{dc} at the open-circuited load terminals. Keep the other independent sources deactivated during this time (i.e. short the voltage source terminals and open the current source terminals). A d.c. driving current i_{dc} will flow in the circuit from the load terminals due to application of v_{dc}.

The internal resistance of the source network is then obtained as

$$R_{Th} = \frac{v_{dc}}{i_{dc}}$$

Both the two methods highlighted above have been illustrated in the following examples.

Example 1.9 *In the network of Fig. 1.32, find the current through the 10 Ω resistor Thevenin's Theorem.*

Solution Let the resistance r_4 (10 Ω) be removed and the circuit is exhibited in Fig. 1.32(a).

At node *C*, application of KCL yields

$$I_1 + I_3 - I_2 = 0$$

or $$\frac{V_1 - V_{oc}}{r_1} + \frac{V_3 - V_{oc}}{r_3} - \frac{V_{oc} - V_2}{r_2} = 0$$

[assuming the open-circuit voltage across the terminals $x - y$ in Fig. 1.32(a) to be V_{oc}; abviously, the potential at *C* node is V_{oc}]

i.e., $$\frac{10 - V_{oc}}{2} + \frac{20 - V_{oc}}{1} - \frac{V_{oc} + 12}{5} = 0$$

or $$-0.5\ V_{oc} - V_{oc} - 0.2\ V_{oc} = 2.4 - 20 - 5$$

or $$1.7\ V_{oc} = +22.6\ \text{V}$$

$\therefore$ $$V_{oc} = 13.29\ \text{V}.$$

Next, the independent voltage source are removed by short-circuits (Fig. 1.32(b)).

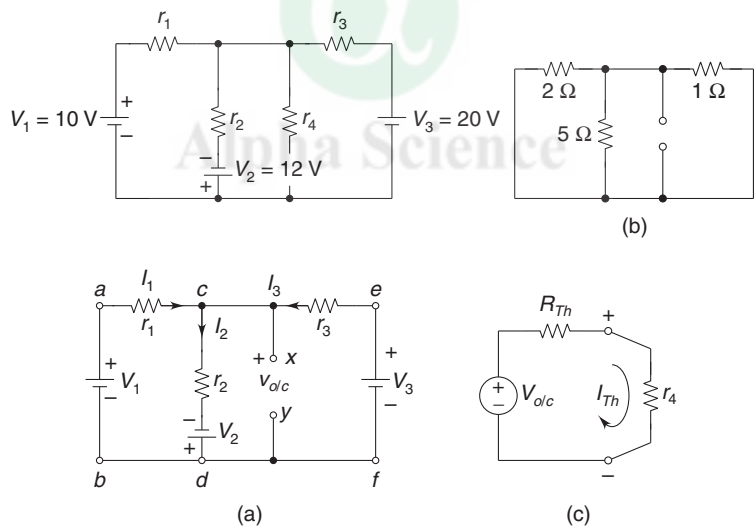


Fig. 1.32

Here, $$\frac{1}{R_{Th}} = \frac{1}{2} + \frac{1}{5} + \frac{1}{1}$$

or $$R_{Th} = \frac{10}{17}\ \Omega.$$

Example 1.10 *Thevenize the bridge circuit across a-b in Fig. 1.33.*

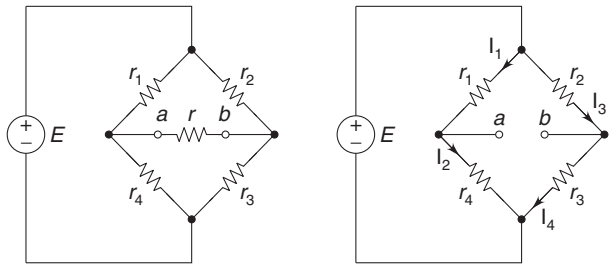


Fig. 1.33(a)

Solution Let r_L be removed and the branch currents be named. Figure 1.33(a) represents the circuit configuration.

With terminals *a-b* open circuited,

$$I_1 = I_2 \quad \text{and} \quad I_3 = I_4$$

Also, $$I_1 = \frac{E}{r_1 + r_4}; \tag{i}$$

and $$I_3 = \frac{E}{r_2 + r_3}; \tag{ii}$$

Again, $$V_{a-b} (= V_{oc}) = I_3 r_2 - I_1 r_1 \tag{iii}$$

Substitution of Eqs. (i) and (ii) in Eq. (iii) yields

$$V_{oc} = \frac{E}{r_2 + r_3} r_2 - \frac{E}{r_1 + r_4} r_1$$

$$= \frac{E r_2 (r_1 + r_4) - E r_1 (r_2 + r_3)}{(r_2 + r_3)(r_1 + r_4)}$$

$$= \frac{E (r_1 r_2 + r_4 r_2 - r_1 r_2 - r_1 r_3)}{(r_2 + r_3)(r_1 + r_4)}$$

or $$V_{oc} = \frac{E (r_4 r_2 - r_1 r_3)}{(r_2 + r_3)(r_1 + r_4)} \tag{iv}$$

V_{oc} in Eq. (iv) represents the Thevenin voltage across *a-b*. Next, the voltage source is deactivated such that Figs. 1.33(b) and (c) represent the circuit configuration.

Obviously, $r_1 \parallel r_2$ and $r_3 \parallel r_4$.

$\therefore$ the resistance through *a-b* is

$$R_{ab} = (r_1 \parallel r_2) + (r_3 \parallel r_4)$$

$$= \frac{r_1 r_2}{r_1 + r_2} + \frac{r_3 r_4}{r_3 + r_4}$$

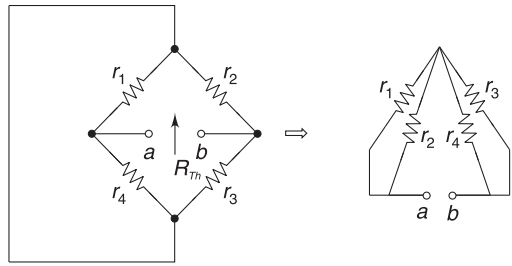


Fig. 1.33

$$= \frac{r_1\, r_2\, (r_3 + r_4) + r_3\, r_4\, (r_1 + r_2)}{(r_1 + r_2)\, (r_3 + r_4)}$$

$$= \frac{r_1\, r_2\, r_3 + r_1\, r_2\, r_4 + r_3\, r_4\, r_1 + r_3\, r_4\, r_1}{(r_1 + r_2)\, (r_3 + r_4)}$$

$\therefore$ In the given bridge circuit,

$$V_{oc} = \frac{E\, (r_4\, r_2 - r_1\, r_3)}{(r_2 + r_3)\, (r_1 + r_4)}$$

and

$$R_{int} = \frac{r_1\, r_2\, r_3 + r_1\, r_2\, r_4 + r_3\, r_4\, r_1 + r_3\, r_4\, r_2}{(r_1 + r_2)\, (r_3 + r_4)}$$

1.13 NORTON'S THEOREM

Norton Theorem is converse of Thevenin's theorem. It consists of equivalent current source instead of equivalent voltage source as done in Thevenin's theorem. The determination of internal resistance of the source network is identical in both the theorems. However, in final stage, i.e., in the Norton equivalent circuit, the current generator is placed in parallel to the internal resistance unlike to that in Thevenin's theorem where the equivalent voltage source was placed in series with the internal resistance.

1.13.1 Statement of Norton's Theorem

A linear active network consisting of independent and or dependent voltage and current sources and linear bilateral network elements can be replaced by an equivalent circuit consisting of a current source in parallel with a resistance, the current source being the short circuited current across the load terminal and the resistance being the internal resistance of the source network looking through the open-circuited load terminals.

1.13.2 Explanation

In order to find the current through r_L, the load resistance (Fig. 1.34(a)) by Norton's theorem, let us replace r_L, by short-circuit (Fig. 1.34 (b)).

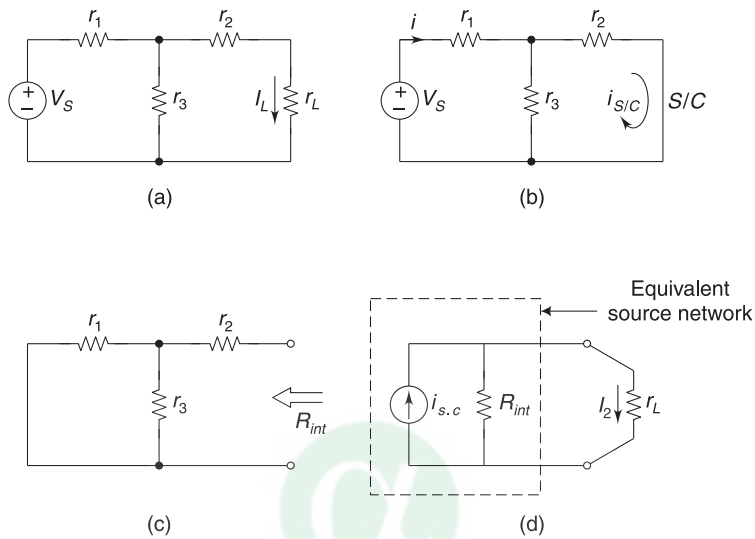


Fig. 1.34 *(a) A simple d.c. network; (b) finding of $i_{s/c}$; (c) Finding of R_{th} (or R_{int}); (d) Norton's equivalent circuit*

Obviously $$i = \frac{V_s}{r_1 + \dfrac{r_2\, r_3}{r_2 + r_3}} \quad \text{and} \quad i_{sc} = i\,\frac{r_3}{r_3 + r_2}$$

Next, the short-circuit is removed and the independent source is deactivated as done in Thevenin's theorem (Fig. 1.34(c)).

Here, $$R_{int} = r_2 + \frac{r_1\, r_3}{r_1 + r_3}.$$

As per Norton's theorem, the equivalent source circuit would contain a current source in parallel to the internal resistance, the current source being the short-circuited current across the shorted terminals of the load resistor (Fig. 1.34 (d)).

Obviously, $$I_L = i_{sc}\,\frac{R_{int}}{R_{int} + r_L}.$$

It may be noted here that determination of R_{int} for the source system in Norton Theorem is identical to that of Thevenin's theorem already described.

1.13.3 Steps for Solving a Network Utilising Norton's Theorem

Step 1 Remove the load resistor and find the internal resistance of the source network by deactivating the constant sources. This procedure is exactly same as described for Thevenin's theorem. Let this resistance be R_{int}.

Step 2 Next, short the load terminals and find the short-circuit current flowing through the shorted load terminals using conventional network analysis. Let this current be i_{sc}.

Step 3 Norton's equivalent circuit is drawn by keeping R_{int} in parallel to i_{sc} as shown in Fig. 1.34(d).

Step 4 Reconnect the load resistor (R_L) across the load terminals and the current through it (I_L) is then given by

$$I_L = i_{sc} \frac{R_{int}}{R_{int} + R_L}.$$

Example 1.11 *Find Norton's equivalent circuit to the left of terminals x-y in the network of Fig. 1.35.*

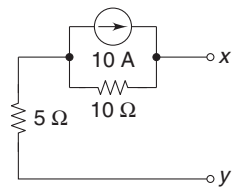


Fig. 1.35

Solution Let us first short the terminals *x-y* (Fig. 1.35(a))

Here, I_{sc} is the current through the 5 Ω resistor.

$$\therefore \quad I_{sc} = 10 \times \frac{10}{10 + 5} = 6.67 \text{ A}$$

[by current divider rule]

To determine the equivalent resistance of the circuit of Fig. 1.35 looking through *x-y*, the constant source is deactivated as shown in Fig. 1.35(b).

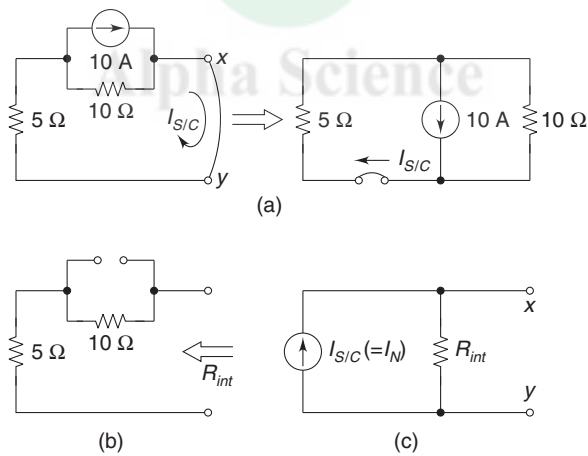


Fig. 1.35(a), (b), (c)

Here, $R_{int} = 10 + 5 = 15\ \Omega$

Norton's equivalent circuit has been shown in Fig. 1.35(c).

Here, $I_N = 6.67$ A; $R_{int} = 15\ \Omega$

1.14 SUPERPOSITION THEOREM

This theorem finds use in solving a network where two or more sources are present and connected not in series or in parallel.

1.14.1 Statement of Superposition Theorem

If a number of voltage or current sources are acting simultaneously in a linear network, the resultant current in any branch is the algebraic sum of the currents that would be produced in it, when each source acts alone replacing all other independent sources by their internal resistances.

1.14.2 Explanation

In Fig. 1.36 to apply superposition theorem, let us first take the source V_1 alone at first replacing V_2 by short-circuit (Fig. 1.36(a)).

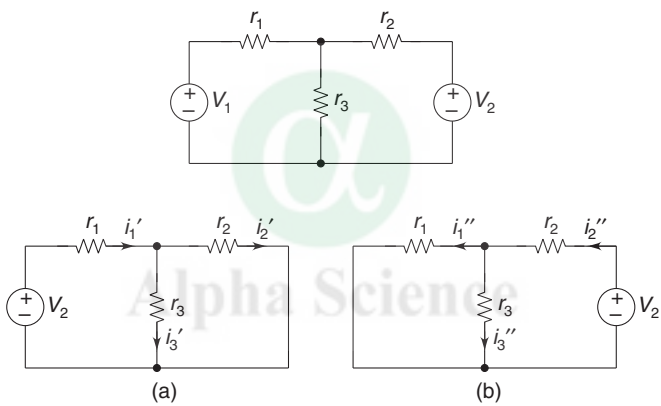


Fig. 1.36

Here,
$$i_1' = \frac{V_1}{\frac{r_2\, r_3}{r_2 + r_3} + r_1}$$

$$i_2' = i_1' \frac{r_3}{r_2 + r_3}$$

and
$$i_3' = i_1' - i_2'$$

Next, removing V_1 by short-circuit, let the circuit be energized by V_2 only (Fig. 1.36(b)).

Here,
$$i_2'' = \frac{V_2}{\frac{r_1\, r_3}{r_1 + r_3} + r_2}$$

and
$$i_1'' = i_2'' \frac{r_3}{r_1 + r_3}$$

Also, $i_3'' = i_2'' - i_1''$.

As per superposition theorem,

$$i_3 = i_3' + i_3''$$
$$i_2 = i_2' - i_2''$$
$$i_1 = i_1' - i_1''$$

[it may be noted that during application of superposition, the direction of currents calculated for each source should be taken care of].

1.14.3 Steps for Solving a Network using the Principle of Superposition

Step 1 Take only one independent source of voltage/current and deactivate the other independent voltage/current sources. (For voltage sources, remove the source and short circuit the respective circuit terminals and for current sources, just delete the source keeping the respective circuit terminals open). Obtain branch currents.

Step 2 Repeat the above step for each of the independent sources.

Step 3 To determine the net branch current utilising superposition theorem, just add the currents obtained in step-1 and step-2 for each branch. If the currents obtained in step-1 and step-2 are in same direction, just add them; on the other hand, if the respective currents are directed opposite in each step, assume the direction of the clockwise current to be +ve and subtract the current obtained in the next step from the original current. The net current in each branch is then obtained.

Example 1.12 *In Fig. 1.37 when $V_0 = 0$, $I = 2$ A; find I when $V_0 = 10$ V.*

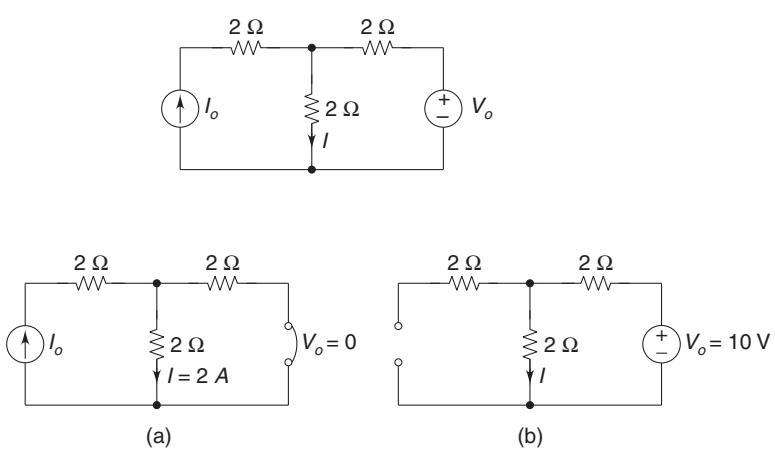


Fig. 1.37

Solution When $V_0 = 0$, the circuit will look like that shown in Fig. 1.37(a). This time it is evident that $I = 2$ A.

Next, $V_0 = 10$ V is applied.

Utilising the concept superposition, the circuit diagram is shown in Fig. 1.37(b) for the voltage source alone. This time $I = 10/(2 + 2) = 2.5$ A.

$\therefore$ Net current through the required resistor i.e., $I = 2 + 2.5 = 4.5$ A as per superposition theorem.

Example 1.13 *In the circuit of Fig. 1.38, find the current through the 5 Ω resistor using the principle of superposition.*

Solution Setting the voltage source deactivated first, current through 5 Ω resistor is zero.

[$\therefore$ With reference to Fig. 1.38(a), the current source will pass the entire current through the shorted path i.e. the deactivated voltage source path and no current would flow through 5 Ω resistor]

i.e., $I' = 0$.

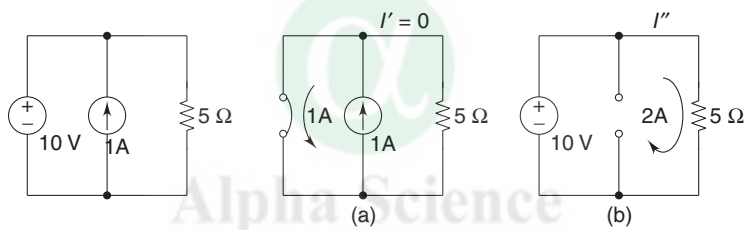


Fig. 1.38

Next, deactivating the current source, the current through the 5 Ω resistor is only 10/5 (= 2) A (= I'').

Using the principle of superposition, I, the current through the 5 Ω resistor is

$$I' + I'' = 0 + 2 = 2 \text{ A}.$$

Example 1.14 *Find I in the circuit shown in Fig. 1.39.*

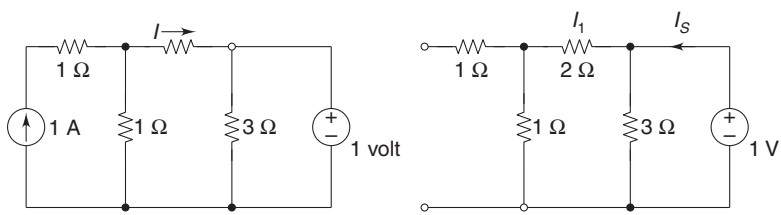


Fig. 1.39(a)

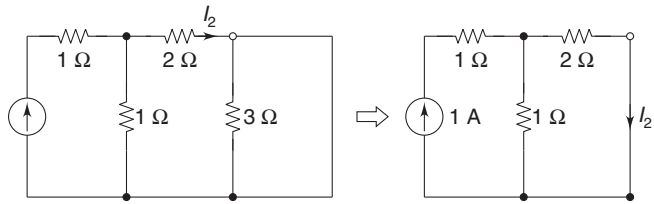


Fig. 1.39(b)

Solution Principle of superposition is applied by taking 1 V source only at first (Fig. 1.39(a)).

$$I_S = \frac{1 \text{ V}}{[(1 + 2) \parallel 3] \ \Omega} = (1/1.5) \text{ A}$$

$$I_1 = I_S \frac{3}{3 + 2 + 1}$$

$$= \frac{1}{1.5} \times \frac{3}{6} = \frac{1}{3} \text{ A} \qquad \text{[by current division formula]}$$

Next, let us assume the current source only (Fig. 1.39(b)),

$$I_2 = 1 \times \frac{1}{1 + 2} = \frac{1}{3} \text{ A} \qquad \text{[by current division formula]}$$

It may be observed that utilising the principle of superposition, the net response, when both the sources (1 A and 1 V) are present, the current through 2 Ω resistor is

$$I = (I_1 - I_2)$$

$$= \frac{1}{3} - \frac{1}{3} = 0 \qquad [I_1 \text{ and } I_2 \text{ being directed reverse}]$$

Example 1.15 *Find v by Superposition Theorem (Fig. 1.40).*

Solution Taking the 10 V source only, with reference to Fig. 1.40(a),

$$v_1 = 10 \times i$$

$$= 10 \frac{10}{5 + 10}$$

$$= 6.67 \text{ V.}$$

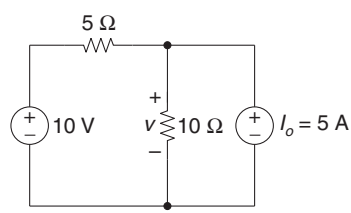


Fig. 1.40

Taking the 5 A source only, with reference to Fig. 1.40(b),

$$i_{10} = 5 \frac{5}{5 + 10} = 1.67 \text{ A}$$

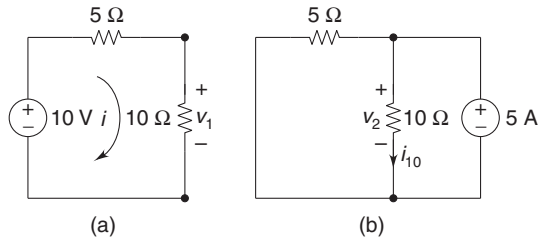


Fig. 1.40 (a), (b)

$\therefore \quad v_2 = 1.67 \times 10 = 16.70 \text{ V}$

By Superposition Theorem,

$$v = v_1 + v_2 = 6.67 + 16.70 = 23.37 \text{ V}.$$

1.15 MAXIMUM POWER TRANSFER THEOREM

This theorem is used to find the value of load resistance for which there would be maximum amount of power transfer from source to load.

1.15.1 Statement of Maximum Power Transfer Theorem

A resistance load, being connected to a d.c. network, receives maximum power when the load resistance is equal to the internal resistance (Thevenin's equivalent resistance) of the source network as seen from the load terminals.

1.15.2 Explanation

A variable resistance R_L is connected to a d.c. source network as shown in Fig. 1.41(a) while Fig. 1.41(b) represents the Thevenin voltage V_0 and Thevenin resistance R_{Th} of the source network. The aim is to determine the value of R_L such that it receives maximum power from the d.c. source.

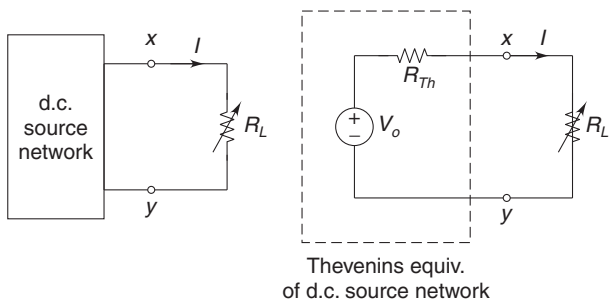


Fig. 1.41 *(a) Load connected to the d.c. source network; (b) Equivalent source network and load*

With reference to Fig. 1.41(b),

$$I = \frac{V_0}{R_{Th} + R_L}$$

while the power delivered to the resistive load is

$$P_L = I^2 R_L = \left(\frac{V_0}{R_{Th} + R_L}\right)^2 \times R_L$$

P_L can be maximized by varying R_L and hence, maximum power can be delivered when $(dP_L/dR_L) = 0$.

However,
$$\frac{dP_L}{dR_L} = \frac{1}{[(R_{Th} + R_L)^2]^2}$$
$$\left[(R_{Th} + R_L)^2 \frac{d}{dR_L}(V_0^2 R_L) - V_0^2 R_L \frac{d}{dR_L}(R_{Th} + R_L)^2\right]$$
$$= \frac{1}{(R_{Th} + R_L)^4}[(R_{Th} + R_L)^2 V_0^2 - V_0^2 R_L \times 2 (R_{Th} + R_L)]$$
$$= \frac{V_0^2 (R_{Th} + R_L - 2R_L)}{(R_{Th} + R_L)^3} = \frac{V_0^2 (R_{Th} - R_L)}{(R_{Th} + R_L)^3}$$

But
$$\frac{dP_L}{dR_L} = 0$$

∴ Finally,

$$\frac{V_0^2 (R_{Th} - R_L)}{(R_{Th} + R_L)^3} = 0$$

which gives $(R_{Th} - R_L) = 0$

or $R_{Th} = R_L$

Hence, it has been proved that power transfer from a d.c. source network to a resistive network is maximum when the internal resistance of the d.c. source network is equal to the load resistance.

Again, with $R_L = R_{Th}$, the system being perfectly matched for load and source, the power transfer becomes maximum and this amount of power ($P_{\max}$) can be obtained as

$$P_{\max} = \frac{V_0^2 R_{Th}}{(R_{Th} + R_{Th})^2} = \frac{V_0^2}{4 R_{Th}}$$

(it may be noted that this is the power consumed by the load).

Obviously, the power transfer by the source would be also $\frac{V_0^2}{4 R_{Th}}$, the load power and source power being the same.

The total power supplied is thus

$$P = 2\,\frac{V_0^{\,2}}{4\,R_{Th}} = \frac{V_0^{\,2}}{2\,R_{Th}}$$

During maximum power transfer the efficiency η becomes

$$\eta = \frac{P_{\max}}{P} \times 100 = 50\%$$

The concept of maximum power transfer by making the source resistance equal to the load resistance has wide application in communication circuits where the magnitude of power transfer is sufficiently small. To achieve maximum power transfer, then the source and load resistances are matched and with flow of maximum power from source to load, low efficiency of 50% is tolerated. On the other hand in electric power transmission systems, the load resistance being sufficiently greater than the source resistance, it is difficult to achieve the condition of maximum power transfer ordinarily. Moreover, in power system, emphasis is given to keep voltage drops and line losses to a minimum value and hence operation of the power system, operating with bulk power transmission capability, becomes uneconomical if it be operated with only 50% efficiency just for the sake of maximum power transfer. Hence in power transmission systems this criterion of maximum power transfer is seldom realized.

1.15.3 Steps to Solve Problems Relating Maximum Power Transfer Theorem

Following steps may be executed to solve the problems relating to maximum power transfer theorem.

Step 1 Remove the load resistance and find Thevenin's resistance (R_{Th}) of the source network looking through the open-circuited load terminals.

Step 2 As per maximum power transfer theorem, this R_{Th} is the load resistance of the network i.e., $R_L = R_{Th}$ that allows maximum power transfer.

Step 3 Find the Thevenin's voltage (V_0) across the open-circuited load terminals.

Step 4 Maximum power transfer is given by $\dfrac{V_0^{\,2}}{4\,R_{Th}}$.

Example 1.16 *Find the value of R in the circuit of Fig. 1.42 such that maximum power transfer takes place. What is the amount of this power?*

Solution Let R be replaced first and the open-circuit voltage be V_{oc} (Fig. 1.42(a)).

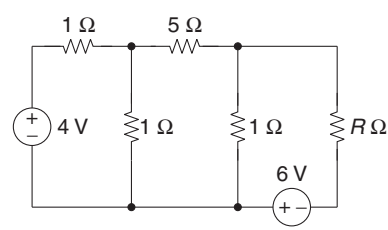


Fig. 1.42

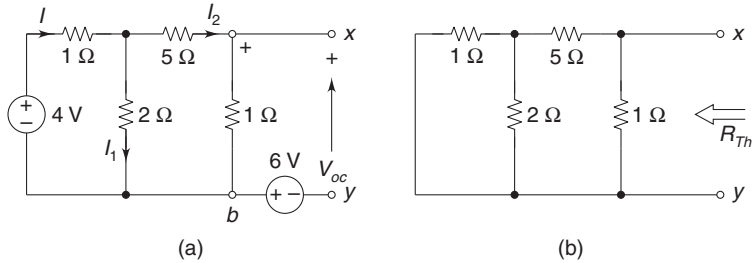


Fig. 1.42

Here, $$I = \frac{4 \text{ V}}{[\{(5 + 1) \parallel 2\} + 1] \ \Omega}$$

$$= \frac{4}{\frac{5}{2}} = \frac{8}{5} \text{ A}$$

$\therefore$ $$I_2 = I \frac{2}{2 + 5 + 1} = \frac{8}{5} \times \frac{1}{4} = \frac{2}{5} \text{ A}$$

The drop across *a-b* branch is then

$$V_{a\text{-}b} = \frac{2}{5} \times 1 = \frac{2}{5} \text{ V}$$

Obviously, $$V_{oc} = V_{a\text{-}b} + 6 \text{ V} = \frac{2}{5} + 6 = \frac{32}{5} \text{ V}$$

or $$V_{oc} = 6.4 \text{ V}$$

To find internal resistance of the circuit across *x-y*, with reference to Fig. 1.42(b),

$$R_{Th} = \{1 \parallel 2 + 5\} \parallel 1 = \frac{17}{3} \parallel 1 = \frac{17}{20} \ \Omega$$

$$= 0.85 \ \Omega$$

As per maximum power transfer theorem,

$$R = R_{Th} = 0.85 \ \Omega$$

and $$P_{\max} \text{ (max. power)} = \frac{V_{oc}^2}{4R} = \frac{(6.4)^2}{4 \times 0.85} = 12 \text{ W}$$

Example 1.17 *Determine the branch currents in the network of Fig. 1.43, when the value of each branch resistance is 1 Ω.*

Solution Let the current directions be as shown in Fig. 1.43.

Apply Kirchhoff's Second law to the closed-circuit *ABDA*, we get

$$5 - x - z + y = 0 \quad \text{or} \quad x - y + z = 5 \qquad \text{(i)}$$

Similarly, circuit *BCDB* gives

$$-(x-z)+5+(y+z)+z=0$$

or $$x-y-3z=5 \qquad \text{(ii)}$$

Lastly, from circuit *ADCEA*, we get

$$-y-(y+z)+10-(x+y)=0$$

or $$x+3y+z=10 \qquad \text{(iii)}$$

From Eq. (i) and (ii), we get,

$$z=0$$

Substituting $z = 0$ either in Eq. (i) or (ii) and in Eq. (iii), we get

$$x-y=5 \qquad \text{(iv)}$$

$$x+3y=10 \qquad \text{(iv)}$$

Subtracting Eq. (v) from (iv), we get

$$-4y=-5$$

or $$y=5/4=1.24 \text{ A}$$

Equation (iv) gives $x = 25/4$ A $= 6.25$ A

Current in branch AB = current branch $BC = 6.25$ A

Current in branch $BD = 0$; current in branch AD = current in branch $DC = 1.25$ A; current in branch $CEA = 6.25 + 1.25 = 7.5$ A.

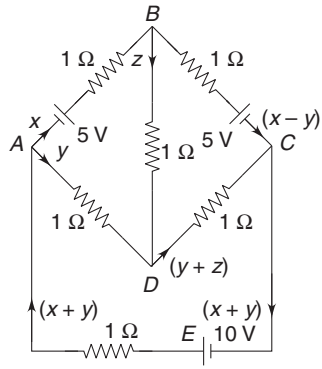


Fig. 1.43

Example 1.18 *By using Superposition Theorem, find the current in resistance R shown in Fig. 1.44(a)*

$R_1 = 0.005\ \Omega$, $R_2 = 0.004\ \Omega$, $R = 1\ \Omega$, $E_1 = 2.05$ V, $E_2 = 2.15$ V

Internal resistances of cells are negligible.

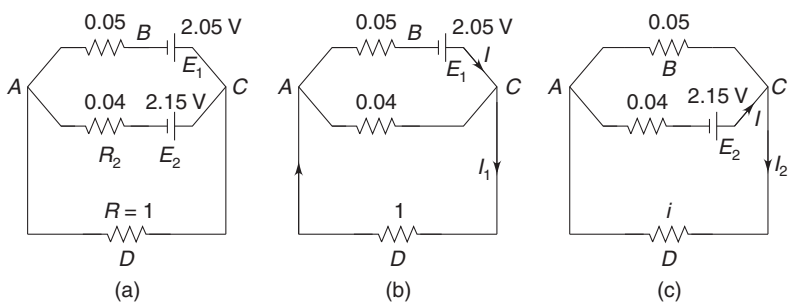


Fig. 1.44

Solution In Fig. 1.44(b), E_2 has been removed. Resistances of 1 Ω and 0.04 Ω are in parallel across points A and C. $R_{AC} = 1 \parallel 0.04 = 1 \times 0.04/1.01 = 0.038\ \Omega$. This resistance is in series with 0.05 Ω. Hence, total resistance offered to battery

$E_1 = 0.05 + 0.038 = 0.088\ \Omega$. $I = 2.05/0.088 = 23.3$ A. Current through 1 Ω resistance, $I_1 = 23.3 \times 0.04/1.04 = 0.896$ A from C to A.

When E_1 is removed, circuit becomes as shown in Fig. 1.44(c). Combined resistance of paths CBA and CDA is $= 1 \parallel 0.05 = 1 \times 0.05/1.05 = 0.048\ \Omega$. Total resistance offered to E_2 is $= 0.04 + 0.048 = 0.088\ \Omega$. Current $I = 2.15/0.088 = 24.4$ A. Again $I_2 = 24.4 \times 0.05/1.05 = 1.16$ A.

The current through 1 Ω resistance when both batteries are present

$$= I_1 + I_2 = 0.896 + 1.16 = 2.056 \text{ A}.$$

Example 1.19 *Use Superposition theorem to find current I in the circuit shown in Fig. 1.45(a). All resistances are in ohms.*

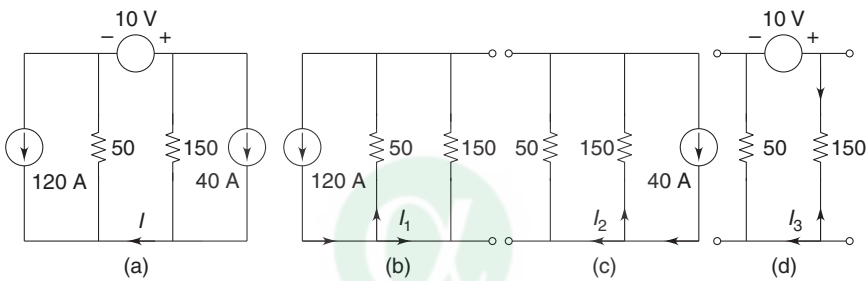


Fig. 1.45

Solution In Fig. 1.45(b), the voltage source has been replaced by a short and the 40 A current sources by an open. Using the current-divider rule, we get

$$I_1 = 120 \times 50/200 = 30 \text{ A}.$$

In Fig. 1.45(c), only 40 A current sources has been considered. Again, using current-divider rule

$$I_2 = 40 \times 150/200 = 30 \text{ A}.$$

In Fig. 1.45(d), only voltage source has been considered. Using Ohm's law,

$$I_3 = 10/200 = 0.05 \text{ A}$$

Since I_1 and I_2 cancel out,

$$I = I_3 = 0.005 \text{ A}$$

Example 1.20 *Use Superposition Theorem to find I in the circuit of Fig. 1.46.*

Solution At a time, one source acts and the other is de-activated, for applying Superposition theorem. If I_1 represents the current in 5 ohm resistor due to 20 V source, and I_2 due to 30 V source,

$$I = I_1 + I_2$$

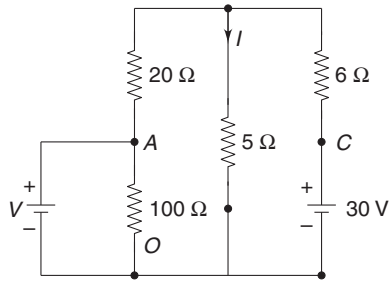


Fig. 1.46

Due to 20 V source, current into node *B*

$$= 20/(20 + 5/6) = 0.88 \text{ amp}$$

Out of this, $I_1 = 0.88 \times 6/11 = 0.48$ amp

Due to 30 V source, current into node *B*

$$= 30/(6 + 5/20) = 3 \text{ amp}$$

Out of this, $I_2 = 3 \times 20/25 = 2.4$ amp

Hence, $I = 2.88$ amp

Alternatively, Thevenin's theorem can be applied at nodes *BD* after removing 5 ohms resistor from its position. Following the procedure to evaluate V_{Th} and R_{Th}.

Thevenin voltage,

$$V_{Th} = 27.7 \text{ volts}$$

and $$R_{Th} = 4.62 \text{ ohms}$$

Current, $$I = 27.7/(4.62 + 5) = 2.88 \text{ amp}$$

Example 1.21 *A bridge network ABCD has arms AB, BC, CD and DA of resistances 1, 1, 2 and 1 ohm respectively. If the detector AC has a resistance of 1 ohm, determine by star/delta transformation, the network resistance as viewed from the battery terminals.*

Solution As shown in Fig. 1.47(b), delta *DAC* has been reduced to its equivalent star.

$$R_D = \frac{2 \times 1}{2 + 1 + 1} = 0.5 \ \Omega,$$

$$R_A = \frac{1}{4} = 0.25 \ \Omega,$$

$$R_C = \frac{2}{4} = 0.5 \ \Omega$$

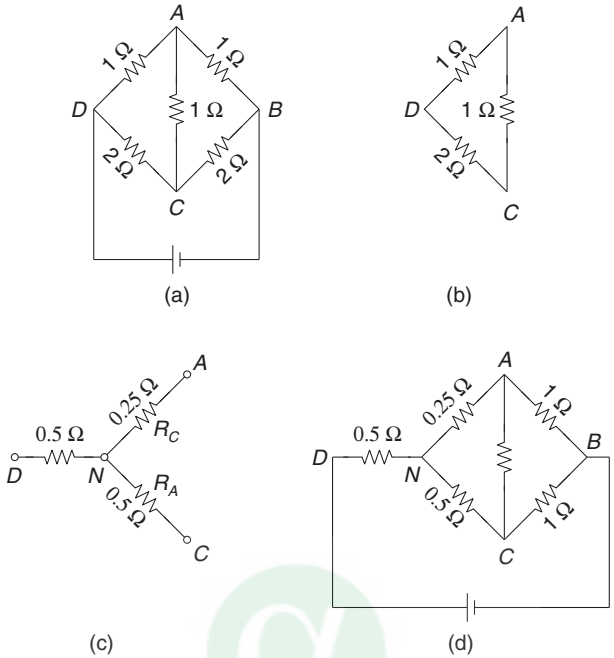


Fig. 1.47

Hence, the original network of Fig. 1.47(a) is reduced to the one shown in Fig. 1.47(d). As seen, there are two parallel paths between points *N* and *B*, one of resistance 1.25 Ω and the other of resistances 1.5 Ω. Their combined resistance is

$$= \frac{1.25 \times 1.5}{1.25 + 1.5} = \frac{15}{22}\ \Omega$$

Total resistance of the network between points *D* and *B* is

$$= 0.5 + \frac{15}{22} = \frac{13}{11}\ \Omega$$

Example 1.22 *A network of resistances is formed as follows as in Fig. 1.48(a) AB = 9 Ω; BC = 1 Ω; CA = 1.5 Ω forming a delta and AD = 6 Ω; BD = 4 Ω and CD = 3 Ω forming a star. Computer the network resistance measured between (i) A and B (ii) B and C and (iii) C and A.*

Solution The star of Fig. 1.48(a) may be converted into the equivalent delta and combined parallel with the given delta *ABC*. Using the rule given in Section 1.11 the three equivalent delta resistance of the given star becomes as shown in Fig. 1.48(b).

When combined together, the final circuit is as shown in Fig. 1.48(c).

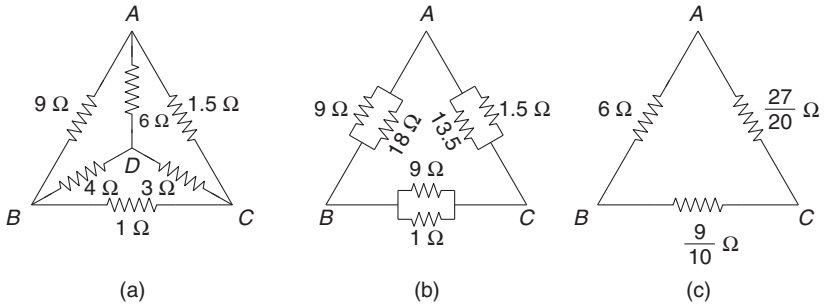


Fig. 1.48

(i) As seen, there are two parallel paths across points *A* and *B*.
   (a) one directly from *A* to *B* having a resistance of 6 Ω and
   (b) the other via *C* having a total resistance

$$= \left(\frac{27}{20} + \frac{9}{10}\right) = 2.25\ \Omega$$

$$\therefore \qquad R_{AB} = \frac{6 \times 2.25}{(6 + 2.25)} = \frac{18}{11}\ \Omega$$

(ii)
$$R_{BC} = \frac{\frac{9}{10} \times \left(6 + \frac{27}{20}\right)}{\left(\frac{9}{10} + 6 + \frac{27}{20}\right)}$$

$$= \frac{441}{550}\ \Omega$$

(iii)
$$R_{CA} = \frac{\frac{27}{20} \times \left(6 + \frac{9}{10}\right)}{\left(\frac{9}{10} + 6 + \frac{27}{20}\right)}$$

$$= \frac{621}{550}\ \Omega$$

Example 1.23 *With the help of Norton's theorem, find V_0 in the circuit shown in Fig. 1.49(a). All resistances are in ohms.*

Solution For solving this circuit, we will Nortonise the circuit to the left to the terminals 1-1′ and to the right of terminals 2-2′, as shown in Figs. 1.49(b) and (c) respectively.

The two equivalent Norton circuits can now be put back across terminals 1-1′ and 2-2′, as shown in Fig. 1.50(a).

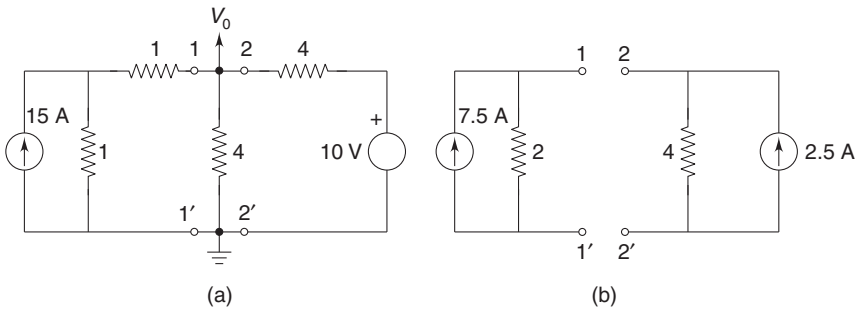


Fig. 1.49

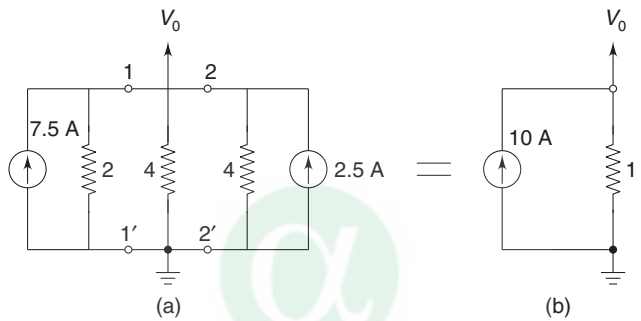


Fig. 1.50

The two current sources being in parallel, can be combined into a single source of 7.5 + 2.5 = 10 A. The three resistors are in parallel and their equivalent resistances is 2 || 4 || 4 = 1 Ω. The value of V_0 as seen from Fig. 1.50(b) is

$$V_0 = 10 \times 1 = 10 \text{ V}.$$

Example 1.24 *For the circuit shown in Fig. 1.51(a), calculate the current in the 6 Ω resistance by using Norton's theorem.*

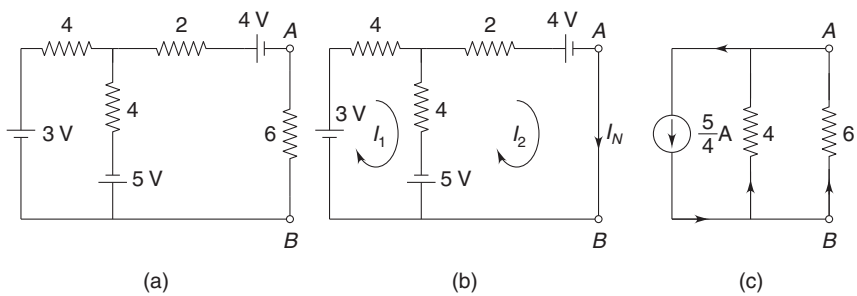


Fig. 1.51

Solution We replace the 6 Ω resistance by a short-circuit as shown in Fig. 1.51(b). Now, we have to find the current passing through the short-circuited terminals *A* and *B*. For this purpose we will use the mesh analysis by assuming mesh currents I_1 and I_2.

From mesh (i), we get

$$3 - 4\,I_1 - 4\,(I_1 - I_2) + 5 = 0$$

or
$$2\,I_1 - I_2 = 2 \qquad \text{(i)}$$

From mesh (ii), we get

$$-2\,I_2 - 4 - 5 - 4\,(I_1 - I_2) = 0$$

or
$$4\,I_1 - 6\,I_2 = 9 \qquad \text{(ii)}$$

From (i) and (ii) above, we get

$$I_2 = -5/4$$

The negative sign shows that the actual direction of flow of I_2 is opposite to the shown in Fig. 1.51(b). Hence, $I_{sh} = I_N = I_2 = -5/4$ A i.e., current flows from point *B* to *A*.

After the terminals *A* and *B* are open-circuited and the three batteries are replaced by short-circuits (since their internal resistances are zero), the internal resistance of the circuit, as viewed from these terminals is

$$R_i = R_N = 2 + 4 \parallel 4 = 4\ \Omega$$

The Norton's equivalent circuit consists of a constant current source of 5/4 A in parallel with a resistance of 4 Ω as shown in Fig. 1.51(c). When 6 Ω resistance is connected across the equivalent circuit, current through it can be found by the current-divider rule.

Current through 6 Ω resistor = 5/4 × 4/10 = 0.5 from *B* to *A*.

Example 1.25 *In the network shown in Fig. 1.52(a), find the value of R_L such that maximum possible power will be transferred to R_L. Find also the value of the maximum power and the power supplied by source under these conditions.*

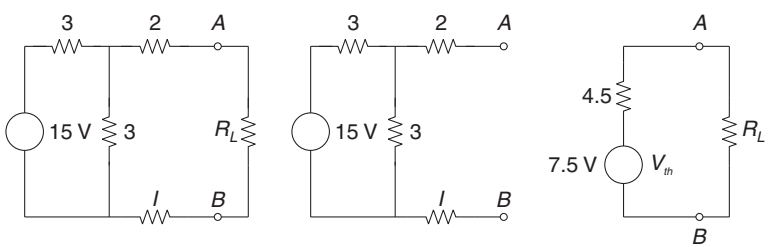


Fig. 1.52

Solution We will remove R_L and find the equivalent Thevenin's source for the circuit to the left terminals A and B. As seen from Fig. 1.52(b) V_{th} equals the drop across the vertical resistor of 3 Ω because no current flows through 2 Ω and 1 Ω resistors. Since 15 V drops across two series resistors of 3 Ω each, V_{th} = 15/2 = 7/5 V. Thevenin's resistance can be found by replacing 15 V source with a short-circuit. As seen from Fig. 1.52(b), $R_{th} = 2 + (3 \parallel 3) + 1 = 4.5$ Ω. Maximum power transfer to the load will take place when $R_L = R_{th} = 4.5$ Ω.

Maximum power drawn by

$$R_L = V_{th}^2/4 \times R_L = 7.5^2/4 \times 4.5$$
$$= 3.125 \text{ W.}$$

Since same power is developed in R_{th}, power supplied by the source

$$= 2 \times 3.125 = 6.250 \text{ W.}$$

1.16 WHEATSTONE BRIDGE

The Wheatstone bridge (Fig. 1.53) consists of four resistances – three known P, Q, R in which R is continuously variables and one unknown resistance X.

The galvanometer G has a resistance by R_g. The battery is of emf ε and assumed to have zero internal resistance. To obtain the balanced condition of the bridge, we first derive an expression for the current through the galvanometer when the bridge is slightly out of balance.

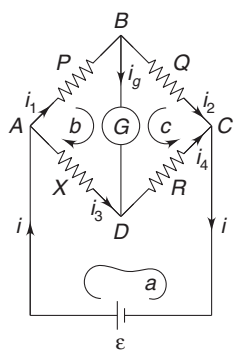


Fig. 1.53

Applying KCL, we get the following junction equations

Junction A : $\quad i - i_1 - i_3 = 0$

Junction B : $\quad i_1 - i_2 - i_g = 0$

Junction D : $\quad i_3 + i_g - i_4 = 0$

Applying KVL, we get the following loop equations

Loop a : $\quad X\, i_3 + R\, i_4 = \varepsilon$

Loop b : $\quad P\, i_1 + R_g\, i_g - X\, i_3 = 0$

Loop c : $\quad Q\, i_2 - R\, i_4 - R_g\, i_g = 0$

To find out the galvanometer current i_g, we reduce the above six equations to the following three equations:

$$X\,(i - i_1) + R\,(i - i_2) = \varepsilon$$
$$P\,i_1 + R_g\,(i_1 - i_2) - X\,(i - i_1) = 0$$
$$Q\,i_2 - R\,(i - i_2) - R_g\,(i_1 - i_2) = 0$$

or

$$i\,(X + R) - i_1\,X - i_2\,R = \varepsilon$$
$$i\,X - i_1\,(P + X + R_g) + i_2\,R_g = 0$$
$$i\,R + i_1\,R_g - i_2\,(Q + R + R_g) = 0$$

In matrix-form, this system of linear equations is written as

$$\begin{bmatrix} X + R & -X & -R \\ X & -(P + X + R_g) & R_g \\ R & R_g & -(Q + R + R_g) \end{bmatrix} \begin{bmatrix} i \\ i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} \varepsilon \\ 0 \\ 0 \end{bmatrix}$$

or in compact form as

$$AX = Y$$

where A is the coefficient matrix, X is the unknown parameter-matrix and Y is the column matrix of constants.

This matrix equation can be solved by using Cramer's rule. If $|A| \neq 0$, the system $AX = Y$ has the unique solution

$$x_1 = i = \frac{|A_1|}{|A|};$$
$$x_2 = i_1 = \frac{|A_2|}{|A|};$$
$$x_3 = i_2 = \frac{|A_3|}{|A|}$$

To find $i_g = i_1 - i_2$, we need to know only i_1 and i_2 i.e., $|A|$, $|A_2|$ and $|A_3|$. Now

$$|A| = \begin{vmatrix} X + R & -X & -R \\ X & -(P + X + R_g) & R_g \\ R & R_g & -(Q + R + R_g) \end{vmatrix}$$
$$= (X + R)\,\{(P + X + R_g)\,(Q + R + R_g) - R_g^2\} +$$
$$X\,\{-X\,(Q + R + R_g) - R\,R_g\} - R\,\{XR_g + R\,(P + X + R_g)\}$$
$$= (PQR + QRX + RXP + XPQ) + R_g\,(PX + QX + PR + QR)$$
$$= \Delta + R_g\,(P + Q)\,(X + R)$$

where $\Delta = PQR + QRX + RXP + XPQ$

$$|A_2| = \begin{vmatrix} X+R & \varepsilon & -R \\ X & 0 & R_g \\ R & 0 & -(Q+R+R_g) \end{vmatrix}$$

$$= (-1)^{1+2} \cdot \varepsilon \begin{vmatrix} X & R_g \\ R & -(Q+R+R_g) \end{vmatrix}$$

$$= \varepsilon\,[RR_g + X\,(Q+R+R_g)]$$

$$|A_3| = \begin{vmatrix} X+R & -X & \varepsilon \\ X & -(P+X+R_g) & 0 \\ R & R_g & 0 \end{vmatrix}$$

$$= (-1)^{1+3} \cdot \varepsilon \begin{vmatrix} X & -(P+X+R_g) \\ R & R_g \end{vmatrix}$$

$$= \varepsilon\,[XR_g + R\,(P+X+R_g)]$$

$\therefore$ $$i_2 = \frac{|A_3|}{|A|} = \frac{\varepsilon\,[X\,R_g + R\,(P+X+R_g)]}{\Delta + R_g\,(P+Q)\,(X+R)}$$

and $$i_1 = \frac{|A_2|}{|A|} = \frac{\varepsilon\,[R\,R_g + X\,(Q+R+R_g)]}{\Delta + R_g\,(P+Q)\,(X+R)}$$

Thus the current through the galvanometer is given by

$$i_g = i_1 - i_2 = \frac{\varepsilon\,(QX - PR)}{\Delta + R_g\,(P+Q)\,(X+R)}$$

In the balanced condition of the bridge, there is no current through the galvanometer and hence $i_g = 0$. Then, we have

$$QX = PR \quad \text{or} \quad \frac{P}{Q} = \frac{X}{R}$$

This is the condition of balance in a Wheatstone bridge. The balance condition is independent of the battery and the galvanometer.

1.17 ALTERNATING CURRENTS

When a battery is connected in series with a resistance the resulting current flows always in the same direction. This is because the battery maintains the same polarity of the output voltage. On the other hand an alternating voltage source reverses or alternates the polarity at a constant rate or frequency. For example, the current in a household electric circuit reverses direction 50 times per second. Most electronic circuits include both d.c. and a.c. voltages. The following table shows the comparison between d.c. and a.c. voltages.

d.c. voltage	a.c. voltage
Polarity fixed	Polarity reverses
Magnitude may be constant or variable	Magnitude varies periodically
Cannot be stepped up or stepped down by a transformer	Can be stepped up or down by a transformer
Easier to measure	Easier to amplify
Heating effect is the same for both d.c. and a.c. currents.	

The usual sources of a.c. voltage are the generator and oscillator.

1.18 SINUSOIDAL WAVEFORM

The basic waveform for a.c. variations is the sine wave. The output of a generator is proportional to the sine of the angle of rotation of the coil. Electronic oscillator circuits containing capacitance and inductance also produce sine wave variations. A sine wave can be derived from a circular motion. Consider a vector rotating in the anticlockwise direction with a uniform angular velocity ω as shown in Fig. 1.54. If the vertical component of the vector is plotted against time we get the time curve shown in the figure.

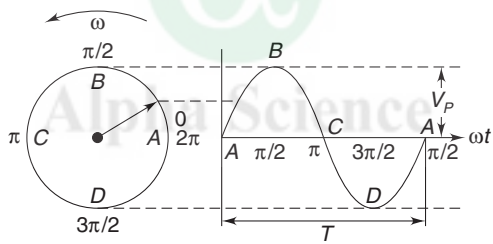


Fig. 1.54

One complete revolution of the vector is called a *cycle*. This corresponds to one complete set of changes from *A* to *D* in the waveform. The time taken for one cycle is called the *period T*. The number of cycles per second is called *frequency f*. The frequency is obviously the reciprocal of the period. One complete revolution covers 2π radian and this requires *T* second. Hence it follows that the angular frequency $\omega = 2\pi f$. If the length of the rotating vector is V_P then the instantaneous value of the vertical component at any instant of time *t* is $V_P \sin \omega t$. Here V_P is the *maximum* or *peak* value. It is the amplitude of the sine wave.

If two sinusoidal waveforms have the same frequency but pass through zero value at different instants of time then they are said to be out of *phase*. The angle between the corresponding rotating vectors is called the *phase angle*. For example the sine wave voltage v_1 in Fig. 1.55 is *leading* the sine wave voltage

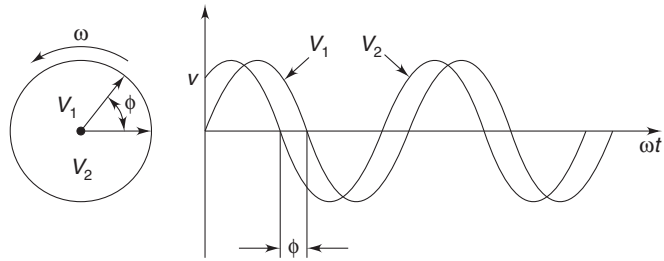


Fig. 1.55

v_2 because it passes through the zero values earlier than v_2. The phase difference between the two waves is the angle ϕ.

A sinusoidal voltage is described by its amplitude and frequency. The most general equation for a sine wave is

$$v = V_P \sin(\omega t + \phi) \tag{1.50}$$

It must be noted that capital letters are used to denote constant or d.c. values and lower case letters are used to denote time-dependent voltages and currents.

1.19 RMS VALUE

It is sometimes necessary to compare the magnitude of a sinusoidal current with a direct current. This is done by comparing the heating effect produced by the sinusoidal and direct currents. To find the heating effect of a sinusoidal current the I^2R loss in a resistor is averaged over one complete cycle. Thus the average power of a sinusoidal current is given by

$$P = \frac{1}{T}\int_0^T I^2 R dt = \frac{1}{T}\int_0^T I_P^2 R \sin^2 \omega t \, dt$$

$$= \frac{I_P^2 R}{T}\int_0^T \frac{1 + \cos 2\omega t}{2} dt$$

$$= \frac{I_P^2 R}{T}\left[\frac{t}{2} + \frac{\sin 2\omega t}{4\omega}\right]_0^T$$

$$= \frac{I_P^2 R}{2} \tag{1.51}$$

The direct current which produces the same heating effect as an alternating current is called the *effective value* of the alternating current. If therefore the effective current is I_e

$$I_e^2 R = \frac{I_P^2 R}{2} \tag{1.52}$$

$$\therefore \qquad I_e = \frac{I_P}{\sqrt{2}} \tag{1.53}$$

The effective value is usually called the *root mean square* or *rms* value. a.c. voltmeters and ammeters are almost universally calibrated in terms of rms values. This enables easy comparison with the readings of d.c. meters. Unless stated otherwise a.c. voltages and currents are characterized by their rms values.

1.20 AVERAGE VALUE

In an attempt to find the average value of a sinusoidal voltage if we consider a full cycle the average comes out to be zero. So we take half a cycle and find the average. Thus if the average value of an ac voltage is V_a.

$$\begin{aligned} V_a &= \frac{2}{T}\int_0^{T/2} V_P \sin \omega t \, dt \\ &= \frac{2}{T}\left[V_P \frac{\cos \omega t}{-\omega}\right]_0^{T/2} \\ &= \frac{2}{T}\left[V_P \cos \frac{\left(\frac{2\pi t}{T}\right)}{-\omega}\right]_0^{T/2} \\ &= \frac{2V_P}{\omega T}[1+1] \\ &= \frac{2V_P}{\left(\frac{2\pi}{T}\right)T} \times 2 = \frac{2V_P}{\pi} \end{aligned} \tag{1.54}$$

Similarly the average current is given by

$$I_a = \frac{2I_P}{\pi} \tag{1.55}$$

1.20.1 Power Factor

The current and voltage in any part of an electronic circuit may or may not be in phase. If in general the phase difference between the voltage and current in any part of a circuit is ϕ we can represent the current and voltage by the equations

$$i = I_P \sin \omega t$$

and

$$v = V_P \sin (\omega t + \phi) \tag{1.56}$$

The instantaneous power is therefore

$$p = vi = V_P I_P \sin (\omega t) \sin (\omega t + \phi) \tag{1.57}$$

Thus, the instantaneous power in any part of a circuit varies with time. It may even become negative as shown in Fig. 1.56.

The meaning of negative and positive power is understood as follows. When the power is positive in a certain part of the circuit the rest of the circuit delivers power to that part. When the power is negative that part gives up power to the rest of the circuit.

The average power is found by averaging the instantaneous power given by Eq. (1.56) over one complete cycle. Thus

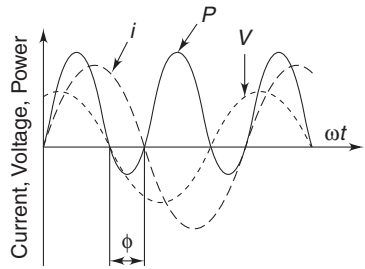


Fig. 1.56

$$P = \frac{1}{T}\int_0^T vi\, dt = \frac{V_P I_P}{T}\int_0^T \sin \omega t \sin(\omega t + \phi)\, dt$$

$$= \frac{V_P I_P}{T}\int_0^T \sin \omega t \{\sin \omega t \cos \phi + \cos \omega t \sin \phi)\, dt$$

$$= \frac{V_P I_P}{T}\int_0^T (\sin^2 \omega t \cos \phi + \sin \omega t \cos \omega t \sin \phi)\, dt$$

$$= \frac{V_P I_P}{T}\left[\int_0^T \frac{1 - \cos 2\omega t}{2} \cos \phi\, dt + \int_0^T \frac{\sin 2\omega t}{2} \sin \phi\, dt\right]$$

$$= \frac{V_P I_P \cos \phi}{T}\left[\frac{t}{2} - \frac{\sin 2\omega t}{2\omega}\right]_0^T + \frac{V_P I_P \sin \phi}{T}\left[\frac{-\cos 2\omega t}{2\omega}\right]_0^T$$

$$= \frac{V_P I_P}{2} \cos \phi = \frac{V_P}{\sqrt{2}} \frac{I_P}{\sqrt{2}} \cos \phi$$

$$= VI \cos \phi \qquad (1.58)$$

where V and I are the effective or rms values.

The product VI is called the *apparent power*. The factor $\cos \phi$ is called the *power factor*. The real power is $VI \cos \phi$. Hence

$$\text{Real power} = (\text{apparent power})\ (\text{power factor}) \qquad (1.59)$$

It must be noted that when the phase angle is 90° the power factor is zero. Then no useful electric power is delivered. It is therefore possible for the current and voltage to have large magnitudes and hence the instantaneous power may be large. But the average power will be zero if the phase angle is 90°. When the voltage and current are in phase the power factor is 1. The average power is then equal to the product of the voltage and current as in the dc circuits.

1.21 UNITS OF FREQUENCY AND PERIOD

The unit of frequency is 1 cycle per second. This unit is called the *hertz* (Hz). In electronics we often come across very large frequencies. To measure such

frequencies we use larger units such as the kilohertz (10^3 Hz), megahertz (10^6 Hz) and the gigahertz (10^9 Hz). The unit of period is the second(s). Very often we have to deal with time intervals which are much smaller than the second. For measuring such small intervals of time we use units such as the ms (10^{-3} s), μs (10^{-6} s) and ns (10^{-9} s).

1.22 PHASORS

In the analysis of ac networks it often becomes necessary to add sinusoidal voltages and currents. One method of doing this is to plot the given voltages along the same axis. Then at each point along the axis the magnitudes of the individual waves are added algebraically. These algebraic sums are then plotted along the same axis. Figure 1.57 illustrates this method in the case of two waves. This method is very laborious and has limited accuracy. A shorter method involves the use of a rotating vector.

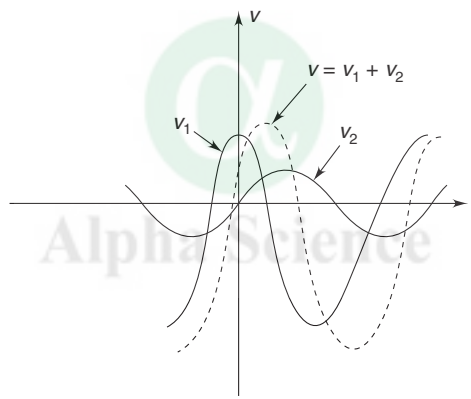


Fig. 1.57

We may recollect that in mechanics the concept of a plane vector is useful in solving problems. For example, a vector representation of forces enables the use of graphical methods for the composition and resolution of forces. In a similar way the use of *phase diagrams* enables us to get a good insight into the voltage and current relations in electronic circuits. A rotating vector that represents a sinusoidally varying quantity is called a *phasor*. Its length represents the amplitude of the quantity. It is imagined to rotate with an angular velocity equal to the angular frequency of the quantity. The instantaneous value of the quantity is represented by the projection of the phasor in a fixed direction.

To illustrate the use of phasors to add two sinusoidal voltages refer to Fig. 1.57. The two vectors *OA* and *OB* represent two sinusoidal voltages v_1 and

v_2. At $t = 0$, the voltage v_1 is along the vertical projection of v_1 is along the horizontal axis and v_2 is at right angles to this axis. At this instant the vertical projection of v_1 is 0 and that of v_2 has reached it's a maximum or peak value. Complete the parallelogram *OACB*. Then the diagonal *OC* represents the resultant of v_1 and v_2.

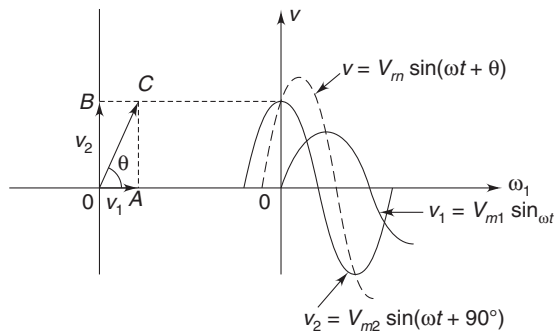


Fig. 1.58

The entire figure is now assumed to rotate in the anticlockwise direction with an angular velocity equal to the angular frequency of the two voltages. then the vertical projection of *OC* gives the instantaneous value of the resultant of the two given voltages. The waveform of this resultant is shown at the right of Fig. 1.57. The amplitude of the resultant is given by

$$v = \left(v_1^2 + v_2^2\right)^{1/2} \tag{1.60}$$

The phase angle θ is given by

$$\theta = \tan^{-1}(v_2/v_1)$$

1.23 COMPLEX NOTATION OF PHASORS

It would be convenient to have an operator which when applied to a phasor rotates it through 90° in the anticlockwise direction without altering its magnitude. Such as operator is the *j operator*. Consider a phasor *P* along the *X*-axis. If this phasor is multiplied by *j* we get the phasor *jP* along the *Y*-axis (Fig. 1.59). if the *j* operator is applied to *jP* we get the phasor j^2P along the negative *X*-axis. Thus the phasor *P* and the phasor j^2P are both of the same magnitude but opposite in direction.

$$\therefore \qquad j^2P = -P$$

or $$j^2 = -1$$

or $$j = \sqrt{-1} \tag{1.61}$$

Thus in Cartesian form the phasor P can be written as

$$P = a + jb \tag{1.62}$$

where a represents the x component and b the y component of the phasor P. Going back to Fig. 1.57 we can write

$$v = v_1 + jv_2 \tag{1.63}$$

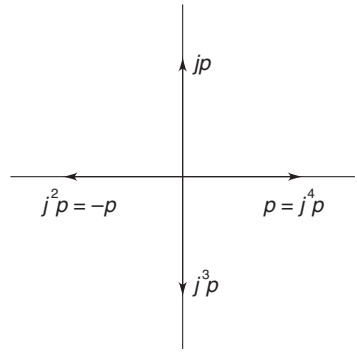


Fig. 1.59

1.24 POLAR NOTATION OF PHASORS

In polar notation the phasor P can be written in the form

$$\begin{aligned} P &= p \cos \theta + jp \sin \theta \\ &= p (\cos \theta + j \sin \theta) \end{aligned} \tag{1.64}$$

Thus, $\cos \theta + j \sin \theta$ is an operator which when applied to a phasor along the X-axis rotates it though an angle θ from its initial position in the counterclockwise direction.

In a similar way $\cos \theta - j \sin \theta$ is an operator when applied to a phasor along the X-axis turns it through an angle θ in the clockwise direction. It can be shown that the above rotation is independent of the initial position of the phasor.

The above results can be expressed in exponential form also. As $\cos \theta + j \sin \theta = e^{j\theta}$ we can write

$$P = pe^{+j\theta} \tag{1.65}$$

Example 1.26 *Two voltage sources represented by $v_1 = 10 \sin (6t)$ and $v_2 = 15 \sin (6t + 90°)$ are connected in series. Find the equivalent voltage source.*

Solution As the two voltages are 90° out of phasor one of them can be the x component and the other the y component of the resultant. Hence in complex notation the resultant can be written as

$$V = a \cos \theta + ja \sin \theta$$

where $a \cos \theta$ = peak value of $v_1 = 10$ and $a \sin \theta$ = peak value of $v_2 = 15$ and $\tan \theta = 15/10$

$\therefore$ Peak value of $V = \sqrt{10^2 + 15^2} = 18.03$

$$\theta = \tan^{-1} (15/10) = 56.31° = 56°19'$$

$\therefore$ Resultant voltage = $18.03 \sin (6t + 56°19')$

The basic elements of an a.c. circuit are resistance, inductance and capacitance. When an a.c. voltage is placed in series with a circuit element the element offers some opposite to the flow of current. This opposition depends on the nature of the circuit element and also in some cases on the frequency of the a.c.

Using the basic equations for an inductor and for a capacitor and ohm's law we first study the response of the individual basic elements to an applied sinusoidal voltage. Then we study the response of combinations of these elements to sinusoidal voltages.

1.25 A.C. VOLTAGE APPLIED TO A PURE RESISTANCE

In the range of frequencies that we come across in electronics the resistance of a resistor is for all practical purposes independent of frequency. Therefore, we can apply Ohm's law to a resistor treating its resistance as constant. Figure 1.60(a) shows a pure resistance R to which an a.c. voltage $v_R = V_m \sin(\omega t)$ is applied.

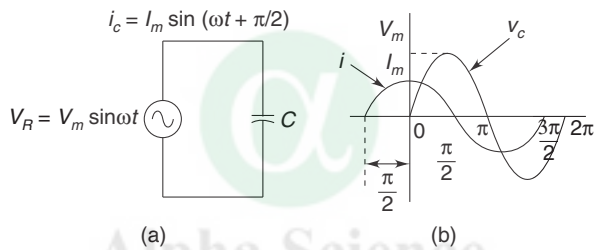


Fig. 1.60

If the current through the resistance is i_R then we have by Ohm's law

$$i_R = \frac{v_R}{R} = \frac{V_m \sin \omega t}{R}$$

$$= \left(\frac{V_m}{R}\right) \sin \omega t = I_m \sin \omega t \qquad (1.66)$$

where $$I_m = \frac{V_m}{R} \qquad (1.67)$$

For a given value of i_R

$$V_R = i_R R = (I_m \sin \omega t) R$$

$$= (I_m R) \sin \omega t = V_m \sin \omega t \qquad (1.68)$$

Figure 1.60(b) shows the plot of the voltage across a pure resistance and the current through it as functions of time. It is clear from the above that the current through a resistance is in plase with the voltage across it.

1.26 A.C. VOLTAGE APPLIED TO A PURE INDUCTANCE

Suppose a current $i_L = I_m \sin \omega t$ is flowing through an inductance L (Fig. 1.61(a)). Then the voltage across the inductance is the induced voltage in it which is given by

$$v_L = L \frac{di_L}{dt} = L \frac{d}{dt} (I_m \sin \omega t) = \omega L I_m \cos \omega t$$

$$= \omega L I_m \sin \left(\omega t + \frac{\pi}{2} \right)$$

$$= V_m \sin \left(\omega t + \frac{\pi}{2} \right) \tag{1.69}$$

where $$V_m = \omega L I_m \tag{1.70}$$

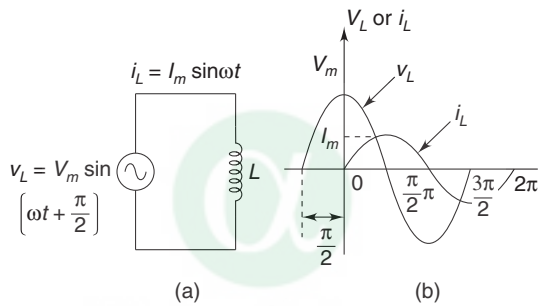


Fig. 1.61

We see from the above that the peak value of the voltage is directly dependent on the frequency of the ac and the inductance. Figure 1.61(b) shows a plot of the voltage across the current through the inductance as a function of time. Equation (1.69) and the voltage and current graphs shown in Fig. 1.6(b) show that the voltage across an inductance leads the current through it by π/2 or 90°. The quantity ωL in Eq. (1.70) is analogous to the resistance R in Eq. (1.67). This is what opposes the flow of current through the inductance. This quantity ωL is called the *reactance* of the inductance. It is denoted by X_L. Thus

$$X_L = \omega L = 2\pi f L \tag{1.71}$$

1.27 A.C. VOLTAGE APPLIED TO A CAPACITANCE

The basic equation for a capacitance is

$$q = Cv.$$

As current is rate of flow of charge we get for the current

$$i_C = \frac{dq}{dt} = C \frac{dv_C}{dt}$$

Suppose we apply an a.c. voltage $v_c = V_m \sin(\omega t)$ to a capacitor of the capacitance C. Then we get from the above equation

$$i_C = C \frac{d}{dt}(V_m \sin \omega t) = \omega C V_m \cos \omega t$$

$$= \omega C V_m \sin\left(\omega t + \frac{\pi}{2}\right)$$

$$= I_m \sin\left(\omega t + \frac{\pi}{2}\right) \tag{1.72}$$

where $$I_m = \omega C V_m \tag{1.73}$$

Figure 1.62 shows the relevant circuit and the plots of the voltage and current as functions of time.

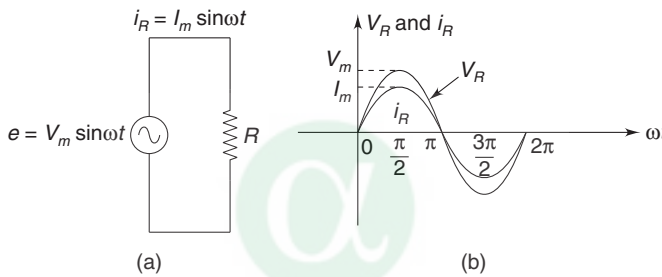


Fig. 1.62

We see from Eq. (1.72) and also from the graphs that the current in a capacitive circuit leads the voltage by $\pi/2$ or 90°. We see from Eq. (1.73) that the quantity $(1/\omega C)$ is analogous to resistance in Eq. (1.67). This is what opposes the flow of current. It is called *capacitive reactance*. It is denoted by X_C.Thus

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f C} \tag{1.74}$$

1.28 FREQUENCY RESPONSE OF THE BASIC ELEMENTS

We have seen above that the resistance of a resistor is independent of frequency in the range of frequencies that we meet with in electronics. Therefore, the frequency response curve for a resistor is a straight line parallel to the frequency axis as shown in Fig. 1.63

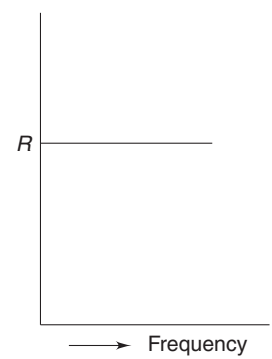


Fig. 1.63

Equation (1.71) shows that the reactance of an inductor is directly proportional to the angular frequency. The angular frequency $\omega = 2\pi f$. Hence the reactance of an

inductor is directly proportional to the frequency of the applied a.c. The frequency response curve of an inductor is therefore a straight line as shown in Fig. 1.64. Larger the inductance greater the slope of the straight line graph.

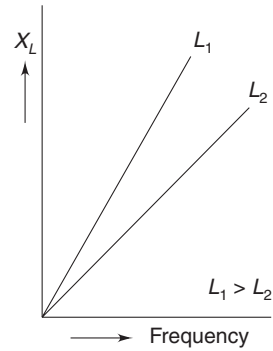


Fig. 1.64

Equation (1.74) indicates that the reactance of a capacitor is inversely proportional to the frequency of the applied a.c. Hence the frequency response curve of a capacitor is a rectangular hyperbola as shown in Fig. 1.65.

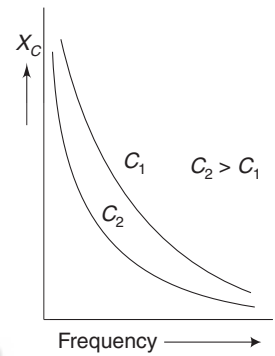


Fig. 1.65

1.29 A.C. POWER IN THE BASIC CIRCUIT ELEMENTS

We have seen earlier that the average power in an a.c. circuit is given by

$$P = VI \cos \phi$$

In the case of a resistor as the current and voltage are in phase $\phi = 0$. Hence the power dissipated in a pure resistor is

$$P = VI = I^2 R$$
$$= V^2/R \qquad (1.75)$$

In an inductor the voltage is 90° ahead of the current. Hence $\phi = 90°$. Hence the power factor $\cos \phi = 0$ and the average power = 0. Thus an ideal inductor does not dissipate any energy. If there is any dissipation of energy in an inductor it is due to the associated resistance in it.

In a capacitor the current leads the voltage by 90°. Hence here also the power factor is zero. Thus there is no power dissipation in a capacitor.

1.30 IMPEDANCE

We have seen above that the voltage across a resistance is in phase with the current, that in an inductance is 90° ahead of the current and that in a capacitance it is 90° behind the current. We can therefore treat resistance, inductive reactance and capacitive reactance as phasors and draw a phasor diagram. For any network the resistance will always be along the positive

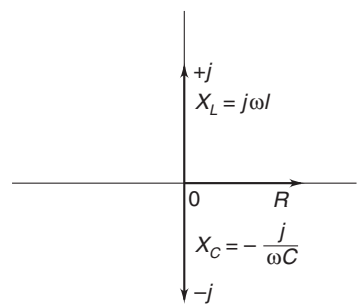


Fig. 1.66

X-axis of the phasor diagram. The inductive reactance will appear along the positive imaginary axis. The capacitive reactance will be along the negative imaginary axis (Fig. 1.66).

In an a.c. circuit we may have resistance, inductive reactance or capacitive reactance or combinations of these. Any one of these three elements or their combination in an a.c. circuit is called the *impedance* of the circuit. The impedance is a measure of the extent to which the circuit impedes or hinders the flow of current in the circuit. Figure 1.66 is called an *impedance diagram*. The symbol for impedance is *Z*. For the individual elements the impedance can be represented as follow

Resistance $Z = R = R \angle 0° = R + j0$ (1.76)

Inductance $Z = X_L = X_L \angle 90° = 0 + jX_L$ (1.77)

Capacitance $Z = X_C = X_C \angle - 90° = 0 - jX_C$ (1.78)

The impedance diagram is a very useful tool. It enables the quick analysis of circuit elements connected in series and parallel.

1.31 SERIES CONNECTION OF BASIC ELEMENTS

As in d.c. series circuits, in series a.c. circuits also the main property is that the current is the same through all the series elements. The total impedance of the series circuit is the sum of the individual impedance (Fig. 1.67).

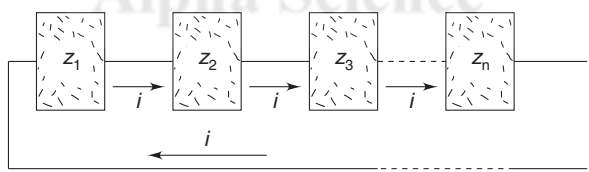


Fig. 1.67

Thus, $Z_T = Z_1 + Z_2 + Z_3 + \ldots$ (1.79)

The impedance have to be added vectorially. The method of doing this will become clear in the examples to follow.

Both reactance and impedance are measured in the same unit as resistance *viz.*, the *ohm*.

1.32 RESISTANCE AND INDUCTANCE IN SERIES

Figure 1.68 shows a resistance *R* and an inductance *L* connected in series with an a.c. voltage source. Let the voltage of the source be $v = V_m \sin \omega t$. If the instantaneous value of the current in the circuit is *i* then the back e.m.f. in

the inductance is $L \frac{di}{dt}$. The potential Ri across the resistance will be the difference between the voltae of the source and the back e.m.f.

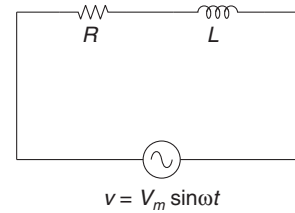


Fig. 1.68

$$\therefore \qquad Ri = V_m \sin \omega t - L \frac{di}{dt}$$

or
$$L \frac{di}{dt} + Ri = V_m \sin \omega t \qquad (1.80)$$

Let us assume that the solution of this equation is

$$i = I_m \sin (\omega t + \theta) \qquad (1.81)$$

where the amplitude l_m and the phase angel θ are constants to be determined. Differentiating Eq. (1.81) we have

$$\frac{di}{dt} = I_m \omega \cos (\omega t + \theta)$$

Substituting for i and $\frac{di}{dt}$ in Eq. (1.80) we have

$$L I_m \omega \cos (\omega t + \theta) + R I_m \sin (\omega t + \theta) = V_m \sin \omega t$$

$$\therefore \quad L I_m \omega \cos \omega t \cos \theta - L I_m \omega \sin \omega t \sin \theta + R I_m \sin \omega t \cos \theta + R I_m \cos \omega t \sin \theta = V_m \sin \omega t$$

$$\therefore \quad \cos \omega t \{L I_m \omega \cos \theta + R I_m \sin \theta\} + \sin \omega t \{R I_m \cos \theta - L I_m \omega \sin \theta\} = V_m \sin \omega t$$

This result must be true at all instants of time.

When $t = 0$ we have $\cos \omega t = 1$ and $\sin \omega t = 0$

$$\therefore \qquad L I_m \omega \cos \theta + R I_m \sin \theta = 0$$

or
$$\tan \theta = - \frac{\omega L}{R} \qquad (1.82)$$

When $\omega t = \frac{\pi}{2}$ we have $\cos \omega t = 0$ and $\sin \omega t = 1$

$$\therefore \qquad R I_m \cos \theta - L I_m \omega \sin \theta = V_m \qquad (1.83)$$

We have from Eq. (1.82)

$$\sin \theta = \frac{- \omega L}{\sqrt{R^2 + \omega^2 L^2}}$$

and
$$\cos \theta = \frac{R}{\sqrt{R^2 + \omega^2 L^2}}$$

Substituting in Eq. (1.83)

$$\left\{\frac{R^2}{\sqrt{R^2+\omega^2L^2}}+\frac{\omega^2L^2}{\sqrt{R^2+\omega^2L^2}}\right\} I_m = V_m$$

$$\therefore \quad I_m = \frac{V_m}{\sqrt{R^2+\omega^2L^2}} \tag{1.84}$$

Hence the instantaneous current is given by

$$i = \frac{V_m}{\sqrt{R^2+\omega^2L^2}} \sin(\omega t + \theta) \tag{1.85}$$

where $\theta = \tan^{-1}\left(\frac{-\omega L}{R}\right)$

The phase angle θ is obviously negative. Hence it is more convenient to write Eq. (1.85) in the form

$$i = \frac{V_m}{\sqrt{R^2+\omega^2L^2}} \sin(\omega t - \theta) \tag{1.86}$$

where $\theta = \tan^{-1}\left(\frac{\omega L}{R}\right)$

The impedance and phase angle can be calculated in a simple way by making use of an impedance diagram. Figure 1.69(a) shows the impedance or phasor diagram for the LR circuit.

The impedance can be written as

$$Z = R \angle 0° + X \angle 90°$$
$$= R + j X_L$$
$$= R + j \omega L$$

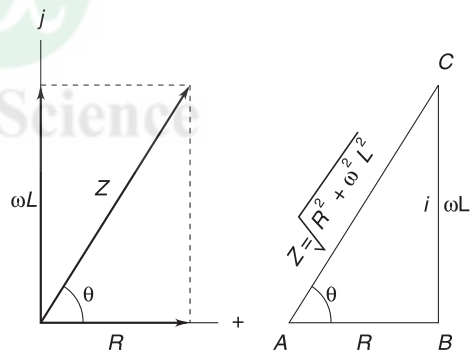


Fig. 1.69

∴ Magnitude Z of the impedance is given by

$$Z^2 = R^2 + (\omega L)^2$$
$$= R^2 + \omega^2 L^2 \tag{1.87}$$

Phase angle $\theta = \tan^{-1}\frac{\omega L}{R}$ (1.88)

If the input voltage is $v = V_m \sin \omega t$ then the current I in the circuit is given by

$$i = \frac{v}{Z} = \frac{V_m \sin(\omega t - \theta)}{\sqrt{R^2+\omega^2 L^2}} \tag{1.89}$$

Sometimes the impedance diagram is drawn as shown in Fig. 1.69(b). Such a diagram is called an *impedance triangle*.

It is clear from the above discussion that the current in an *LR* circuit lags behind the voltage or that the voltage leads the current.

1.33 RESISTANCE AND CAPACITANCE IN SERIES

Figure 1.70 shows a resistance R and a capacitance C in series with an a.c. voltage source $v = V_m \sin \omega t$. If the instantaneous value of the charge on the capacitor plates is q then the potential across the plates is q/C. The instantaneous value of the current is $i = \frac{dq}{dt}$. The potential across the capacitor is equal to the difference between the source voltage and the potential drop across the resistor. Hence

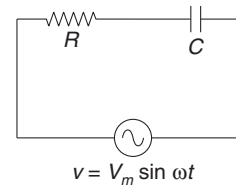


Fig. 1.70

$$\frac{q}{C} = V_m \sin (\omega t) - R \frac{dq}{dt}$$

or
$$R \frac{dq}{dt} + \frac{q}{C} = V_m \sin (\omega t) \tag{1.90}$$

Let us assume that the solution of this equation is

$$q = Q_m \cos (\omega t + \theta) \tag{1.91}$$

Then
$$\frac{dq}{dt} = - Q_m \omega \sin (\omega t + \theta)$$

Substituting for q and dq in Eq. (1.90) we get

$$- RQ_m \omega \sin (\omega t + \theta) + \frac{Q_m}{C} \cos (\omega t + \theta) = V_m \sin \omega t$$

$$\therefore \quad - RQ_m \omega \{\sin \omega t \cos \theta + \cos \omega t \sin \theta\} + \frac{Q_m}{C} \{\cos \omega t \cos \theta - \sin \omega t \sin \theta\} = V_m \sin \omega t$$

This result must be true at all instants of time.

At $t = 0$ we have $\cos \omega t = 1$ and $\sin \omega t = 0$

$$\therefore \quad - RQ_m \omega \sin \theta + \frac{Q_m}{C} \cos \theta = 0$$

$$\therefore \quad \tan \theta = \frac{1}{RC \omega} \tag{1.92}$$

When $\omega t = \pi/2$ we have $\cos \omega t = 0$ and $\sin \omega t = 1$

$$\therefore \quad - RQ_m \omega \cos \theta - \frac{Q_m}{C} \sin \theta = V_m$$

or $$-CRQ_m \omega \cos\theta - Q_m \sin\theta = CV_m \tag{1.93}$$

We have from Eq. (1.92)

$$\sin\theta = \frac{1}{\sqrt{1 + R^2 C^2 \omega^2}}$$

and $$\cos\theta = \frac{RC\omega}{\sqrt{1 + R^2 C^2 \omega^2}}$$

$$\frac{-R^2 C^2 \omega^2 Q_m}{\sqrt{1 + R^2 C^2 \omega^2}} - \frac{Q_M}{\sqrt{1 + R^2 C^2 \omega^2}} = V_m C$$

or $$-\sqrt{1 + R^2 C^2 \omega^2}\, Q_m = V_m C$$

or $$Q_m = -\frac{C}{\sqrt{1 + R^2 C^2 \omega^2}}\, V_m \tag{1.94}$$

Substituting in Eq. (1.91)

$$q = \frac{-V_m C}{\sqrt{1 + R^2 C^2 \omega^2}} \cos(\omega t + \theta) \tag{1.95}$$

Differentiating, we get

$$i = \frac{dq}{dt} = \frac{\omega C V_m}{\sqrt{1 + R^2 C^2 \omega^2}} \sin(\omega t + \theta)$$

$$= \frac{V_m}{\sqrt{R^2 + \frac{1}{\omega^2 C^2}}} \sin(\omega t + \theta) \tag{1.96}$$

or $$i = I_m \sin(\omega t + \theta) \tag{1.97}$$

where $$I_m = \frac{V_m}{\sqrt{R^2 + \frac{1}{\omega^2 C^2}}} \sin(\omega t + \theta) \tag{1.98}$$

and $$\theta = \tan^{-1}\left(\frac{1}{RC\omega}\right) \tag{1.99}$$

We see from Eq. (1.97) that the current in an *RC* circuit leads the voltage or that the voltage lags behind the current.

The impedance and the phase lag can be determined easily using the impedance diagram. Figure 1.71(a) shows the impedance diagram for a *CR* circuit.

In phasor language we can write

impedance = $$Z = R \angle 0° + X_C \angle -90° = R - jX_C$$

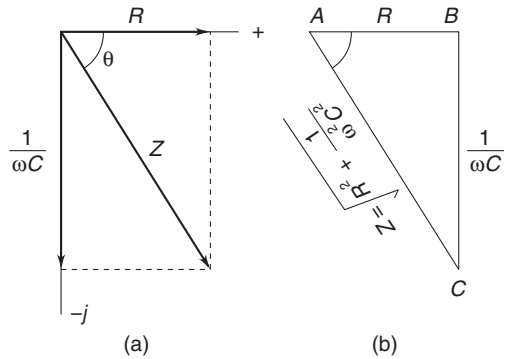


Fig. 1.71

Magnitude of the impedance

$$Z = \sqrt{R^2 + (1/\omega C)^2}$$

$$= \sqrt{R^2 + 1/\omega^2 C^2} \tag{1.100}$$

and phase angle $$\theta = -\tan^{-1} \frac{1}{\omega CR} \tag{1.101}$$

Thus if the current in the circuit is

$$i = I_m \sin \omega t \tag{1.102}$$

then the voltage will be

$$v = Zi = \sqrt{R^2 + (1/\omega^2 C^2)}\ I_m \sin (\omega t - \theta) \tag{1.103}$$

$$= V_m \sin (\omega t - \theta)$$

where θ is given by Eq. (1.101).

Conversely if the applied voltage is

$$v = I_m \sin (\omega t)$$

then $$i = I_m \sin (\omega t + \theta)$$

Figure 1.71(b) shows the impedance triangle for a *CR* circuit. From this triangle also we can obtain expressions for the impedance and the phase angle.

1.34 INDUCTANCE AND CAPACITANCE IN SERIES

Figure 1.72 shows an inductance L and a capacitance C in series with an ac voltage source $v = V_m \sin \omega t$. If the instantaneous charge on the capacitor plates is q then the equation for the circuit is

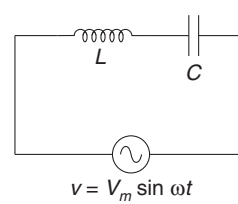


Fig. 1.72

$$L\frac{di}{dt} + \frac{q}{C} = V_m \sin \omega t$$

Since $i = \dfrac{dq}{dt}$ we get

$$L\frac{d^2q}{dt^2} + \frac{q}{C} = V_m \sin \omega t \tag{1.104}$$

Let us assume the solution of this equation to be

$$q = Q_m \cos (\omega t + \theta) \tag{1.105}$$

Differentiating twice we get

$$\frac{d^2q}{dt^2} = -\, Q_m\, \omega^2 \cos (\omega t + \theta)$$

Substituting in Eq. (1.104), we get

$$-\, L\, \omega^2\, Q_m \cos (\omega t + \theta) + \frac{Q_m}{C} \cos (\omega t + \theta) = V_m \sin \omega t \tag{1.106}$$

This equation must be true at all instants of time. Hence at $t = 0$ we have

$$(-\, Q_m\, \omega^2 L + Q_m/C) \cos \theta = 0$$

$\therefore$ $\cos \theta = 0$ or $\theta = \pi/2$

Writing $\theta = \pi/2$ in Eq. (1.106), we have

$$V_m = Q_m \left(\frac{1}{C} - \omega^2 L\right) \tag{1.107}$$

Substituting for Q_m from Eq. (1.107) in Eq. (1.105), we get

$$q = \frac{V_m}{\dfrac{1}{C} - \omega^2 L} \cos \left(\omega t + \frac{\pi}{2}\right)$$

$$= \frac{-\, V_m}{\dfrac{1}{C} - \omega^2 L} \sin \omega t$$

$$= \frac{V_m}{\omega^2 L - \dfrac{1}{C}} \sin \omega t$$

Differentiating we get

$$i = \frac{dq}{dt} = \frac{\omega\, V_m}{\omega^2 L - \dfrac{1}{C}} \cos \omega t$$

$$= \frac{V_m}{\omega L - \dfrac{1}{\omega C}} \cos \omega t \tag{1.108}$$

Writing $\dfrac{V_m}{\omega L - \dfrac{1}{\omega C}} = I_m$ we get

$$i = I_m \cos \omega t \tag{1.109}$$

Since $v = V_m \sin \omega t$ it follows that the current in an LC circuit is at all instants of time 90° out of phase with the applied voltage. The power factor of an LC circuit is therefore zero.

In phasor language we have

impedance = $Z = X_L \angle 90° + X_C \angle - 90° = j\omega L - j/\omega C$

$\therefore$ Magnitude of the impedance

$$Z = \omega L - \frac{1}{\omega C} \tag{1.110}$$

The phase angle will be + 90° if the inductive reactance is more than the capacitive reactance. Then the voltage leads the current by $\pi/2$. It will be – 90° if the capacitive reactance is more than the inductive reactance. Then the voltage lags behind the current by $\pi/2$.

1.35 RESISTANCE, INDUCTANCE AND CAPACITANCE IN SERIES

Figure 1.73 shows a resistance R, an inductance L and a capacitance C connected in series with an a.c. voltage $v = V_m \sin \omega t$. The equation for this circuit is

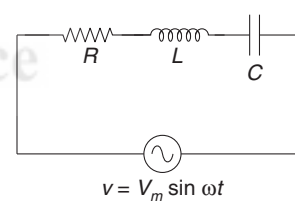


Fig. 1.73

$$L \frac{di}{dt} + iR + \frac{q}{C} = V_m \sin \omega t \tag{1.111}$$

Since $i = \dfrac{dq}{dt}$ we get

$$L \frac{d^2q}{d^2t} + R \frac{dq}{dt} + \frac{q}{C} = V_m \sin \omega t \tag{1.112}$$

Let us assume the solution of this equation to be

$$q = Q_m \cos (\omega t + \theta) \tag{1.113}$$

Then $$\frac{dq}{dt} = - \omega Q_m \sin (\omega t + \theta)$$

and $$\frac{d^2q}{d^2t} = - \omega^2 Q_m \cos (\omega t + \theta)$$

Substituting in Eq. (1.112) we get

$$-\omega^2 L Q_m \cos(\omega t + \theta) - \omega R Q_m \sin(\omega t + \theta) + \frac{Q_m}{C}\cos(\omega t + \theta) = V_m \sin \omega t \quad (1.114)$$

This equation must be true at all instants of time. Hence writing $t = 0$ we get

$$-\omega^2 L Q_m \cos\theta - \omega R Q_m \sin\theta + \frac{Q_m}{C}\cos\theta = 0$$

or
$$\tan\theta = \frac{\frac{1}{\omega C} - \omega L}{R} \quad (1.115)$$

From this we can deduce

$$\sin\theta = \frac{\frac{1}{\omega C} - \omega L}{\sqrt{R^2 + \left(\frac{1}{\omega C} - \omega L\right)^2}}$$

and

$$\cos\theta = \frac{R}{\sqrt{R^2 + \left(\frac{1}{\omega C} - \omega L\right)^2}}$$

Writing $\omega t = \pi/2$ in Eq. (1.114) and simplifying we get

$$Q_m = \frac{-V_m}{\omega\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} \quad (1.116)$$

Substituting in Eq. (1.113)

$$q = \frac{V_m}{\omega\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}\cos(\omega t + \theta) \quad (1.117)$$

Differentiating we get the instantaneous value of the current. Thus

$$i = \frac{V_m}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}\sin(\omega t + \theta) \quad (1.118)$$

Writing

$$\frac{V_m}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} = I_m \quad (1.119)$$

we get
$$i = I_m \sin(\omega t + \theta) \quad (1.120)$$

where the phase angle θ is given by Eq. (1.115).

We have seen in the case of simpler circuits that the effect of inductance is to make the voltage lead the current and that of capacitance to make the voltage lag behind the current. We see from Eq. (1.115) that if the capacitive reactance is more than the inductive reactance, θ is positive. Then the current leads the voltage. If the capacitive reactance is less than the inductive reactance, θ is negative. Then the voltage leads the current. The impedance of the circuit is Z given by

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \tag{1.121}$$

As in the earlier cases in this case also we can obtain expressions for the impedance and phase angle by using the impedance diagram (Fig. 1.74).

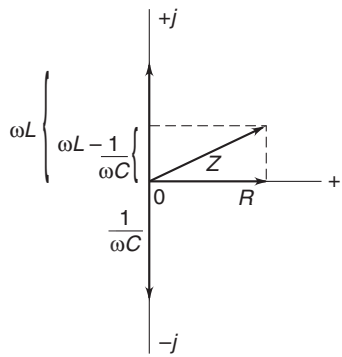


Fig. 1.74

Now

$$\begin{aligned} Z &= R \angle 0° + X_L \angle 90° \\ &\quad + X_C \angle - 90° \\ &= R + jX_L - jX_C \\ &= R + j\,(X_L - X_C) \end{aligned}$$

∴ Magnitude of the impedance

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \tag{1.122}$$

and phase angle

$$= \theta = \tan^{-1}\left(\frac{\omega L - \frac{1}{\omega C}}{R}\right)$$

The voltage leads the current by this angle.

1.36 ADMITTANCE AND SUSCEPTANCE

In dc circuits the reciprocal of resistance was called conductance G measured in the unit Siemens. When a number of conductors were connected in parallel the total conductance was equal to the sum of the conductances of the individual branches. In an ac circuit the reciprocal of impedance is called *admittance*. It is denoted by Y and is measured in the unit *Siemens*. Admittance is a measure of how well an ac circuit admits current to flow through it. For a given potential, therefore, the larger the admittance heavier the current that flows through the circuit. Figure 1.75 shows a number of impedances connected in parallel.

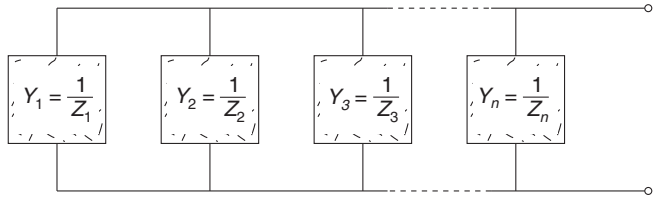


Fig. 1.75

The total admittance of this circuit is given by

$$Y = Y_1 + Y_2 + Y_3 + \ldots \tag{1.123}$$

If the impedances of the individual branches are Z_1, Z_2, Z_3, … then

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3} + \ldots \tag{1.124}$$

The reciprocal of reactance is called *susceptance*. It is also measured in Siemens. It is represents by the capital letter B. Thus for an inductor

$$Y = \frac{1}{X_L} = \frac{1}{X_L \angle 90^\circ} = \frac{1}{\omega L \angle 90^\circ} = \frac{1}{\omega L} \angle 90^\circ \tag{1.125}$$

We define susceptance as

$$B_L = \frac{1}{X_L} \tag{1.126}$$

Then

$$B_L = B_L \angle -90^\circ \tag{1.127}$$

For a capacitor

$$Y = \frac{1}{X_C} = \frac{1}{X_C \angle -90^\circ} = \frac{1}{(\omega C) \angle -90^\circ}$$

$$= \omega C \angle 90^\circ \tag{1.128}$$

If we define

$$B_C = \frac{1}{X_C} \tag{1.129}$$

then

$$B_C = B_C \angle 90^\circ \tag{1.130}$$

For an inductance an increase is frequency or inductance will result in a decrease in a susceptance and hence in admittance.

For a capacitor an increase in frequency or capacitance will result in an increase in a susceptance and hence in admittance.

The results for ac circuits may be summarized as follows:

Resistance:

$$Y = \frac{1}{R} = G = G \angle 0^\circ = G + j0 \tag{1.131}$$

Inductance:

$$Y = \frac{1}{X_L} = B_L = B_L \angle -90° = 0 - jB_L \tag{1.132}$$

Capacitance:

$$Y = \frac{1}{X_C} = B_C = B_C \angle 90° = 0 + jB_C \tag{1.133}$$

We have seen how impedance diagram enable easy analysis of series a.c. circuits. In a similar way for parallel ac circuits we can use *admittance diagrams*. Figure 1.76 shows how an admittance diagram is to be drawn. Like resistance, conductance G is on the positive real axis. Inductive susceptance B_L appears along the negative imaginary axis. Capacitive susceptance appears along the positive imaginary axis.

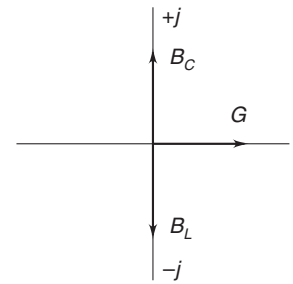


Fig. 1.76

Example 1.27 *An a.c. voltage of 300 sin ωt is applied to a resistance of 50 ohm. Using phasor algebra find the current flowing in the resistor.*

Solution We are given $v = 300 \sin \omega t$. The rms value of this voltage is $300/\sqrt{2} = 212.2$ V. Therefore in phasor language the given voltage is

$$V = 212.2 \angle 0°$$

$$\therefore \text{ Current } = I = \frac{V}{R} = \frac{212.2 \angle 0°}{50 \angle 0°} = 4.244 \angle 0°$$

$$\therefore \text{ Peak value of current } = 4.244 \angle 2 = 6$$

$$\therefore \quad i = 6 \sin \omega t.$$

Example 1.28 *A current $i = 6 \sin (\omega t + 60°)$ is flowing through a resistance of 4 ohm. Using phasor algebra find the voltage across the resistance. Draw the phasor diagram.*

Solution RMS value of the current

$$= 6/\sqrt{2} = 4.243.$$

$$\therefore \text{ In phasor form current } = 4.243 \angle 60°.$$

$$\therefore \quad V = IR$$

$$= (4.243 \angle 60°)\ (4 \angle 0°)$$

$$= 16.97 \angle 60°.$$

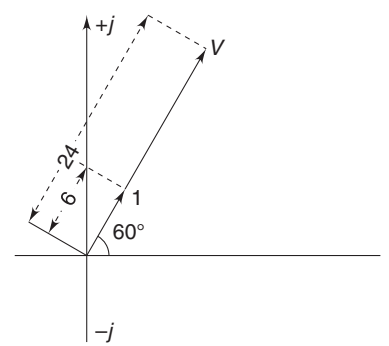


Fig. 1.77

Figure 1.77 shows the corresponding phasor diagram.

Example 1.29 *A voltage of 18 sin (ωt) is applied to an inductance of 6 ohm. Using phasor algebra find the current through the inductance.*

Solution RMS value of voltage = $18/\sqrt{2} = 12.73$. Hence in phasor language

Applied voltage $V = 12.73 \angle 0°$

Current = $I = \dfrac{V}{X_L} = \dfrac{12.73 \angle 0°}{6 \angle 90°} = 2.12 \angle -90°$

$\therefore$ $i = \sqrt{2}\ (2.12) \sin (\omega t - 90°)$

$= 3 \sin (\omega t - 90°)$

Example 1.30 *A current $i = 6 \sin (\omega t + 60°)$ is flowing through an inductance of 10 ohm. Find the voltage across the inductance. Also draw the phasor diagram for the circuit.*

Solution RMS value of current = $6/\sqrt{2} = 4.243$. Therefore in phasor language

$I = 4.243 \sin \angle 60°$

$V = IX_L = (4.243 \sin \angle 60°)\ (10 \angle 90°)$

$= 42.43 \angle 150°$

$\therefore$ $v = \sqrt{2}\ (42.43 \angle 150°) = 60 \sin (\omega t + 150°)$

Figure 1.78 shows the phasor diagram for the circuit.

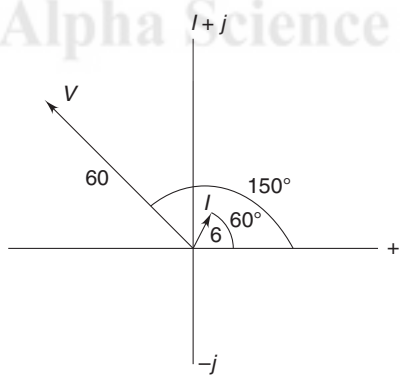


Fig. 1.78

Example 1.31 *An a.c. voltage 60 sin (ωt) is applied to a resistance of 6 ohm and an inductance of 8 ohm in series. Draw the impedance diagram. Calculate the current in the circuit and the phase angle.*

Solution In complex notation

$$Z = Z_1 + Z_2 = 6 + j8$$

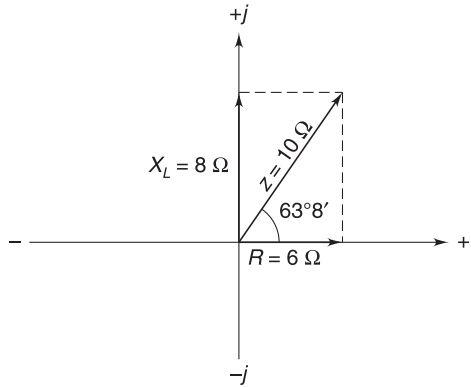


Fig. 1.79

$\therefore$ $$Z = (6^2 + 8^2) = 10 \text{ ohm}$$

$$\text{Phase angle} = \tan^{-1}(8/6) = 53°8'$$

Figure 1.79 shows the impedance diagram. The current lags behind the voltage by 53°8′. Hence the current is given by

$$i = \frac{V}{Z} \sin(\omega t - 53°8')$$

$$= \frac{60}{10} \sin(\omega t - 53°8')$$

$$= 6 \sin(\omega t - 53°8')$$

The rms value of the current is $6\sqrt{2} = 4.243$. Hence in phasor language

$$I = 4.243 \angle -53°8'.$$

Example 1.32 *A current of $i = 70 \sin(\omega t + 60°)$ is flowing in a circuit having a resistance of 3 Ω and a capacitance of 6 Ω. Find the voltage drop across the resistance and capacitance.*

Solution Impedance $Z = 3 + 6j$

$\therefore$ Magnitude of impedance

$$Z = (3^2 + 6^2)^{1/2} = 6.71 \text{ ohm}$$

Phase angle $\theta = \tan^{-1}(6/3)\ 63°26'$.

The voltage across the resistance is in phase with the current.

$\therefore$ Voltage across resistance = (resistance) (current)

$$= 3 \times 70 \sin(\omega t + 60°)$$

$$= 210 \sin(\omega t + 60°)$$

The voltage across the capacitor lags behind the current by 63°26′

$\therefore$ voltage across capacitor = $6 \times 70 \sin (\omega t + 60° - 63°26')$

$= 420 \sin (\omega t - 3°26')$

1.37 KIRCHHOFF'S LAWS FOR A.C. NETWORKS

The statements of Kirchhoff's laws are similar to those for d.c. networks except that instead of algebraic sum of currents and voltages, we take phasor or vector sums for a.c. networks.

1. *Kirchhoff's Current Law*. According to this law, in any electrical network, the phasor sum of the currents meeting at a junction is zero.

   In other words, $\Sigma I = 0$ … at a junction

   Put in another way, it simply means that in any electrical circuit the phasor sum of the currents flowing towards a junction is equal to the phasor sum of the currents going away from that junction.

2. *Kirchhoff's Voltage Law*. According to this law, the phasor sum of the voltage drops across each of the conductors in any closed path (or mesh) in a network plus the phasor sum of the e.m.fs. connected in that path is zero.

   In other words $\Sigma IR + \Sigma$ e.m.f. $= 0$ … round a mesh

Example 1.33 *Use Kirchhoff's laws to find the current flowing in each branch of the network shown in Fig. 1.80.*

Solution Let the current distribution be as shown in Fig. 1.80(b). Stating from point *A* and applying KVL to closed loop *ABEFA*, we get

$$-10(x + y) - 20x + 100 = 0$$

or
$$3x + y = 10 \qquad \text{(i)}$$

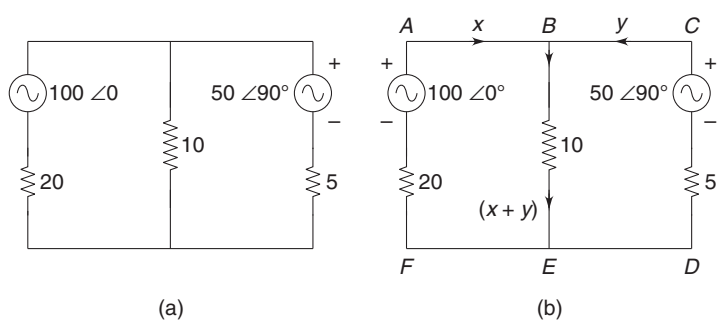


Fig. 1.80

Similarly, considering the closed loop *BCDEB* and starting from point *B*, we have

$$-50\angle 90^\circ + 5y + 10(x + y) = 0$$

or
$$2x + 3y = j10 \qquad \text{(ii)}$$

Multiplying Eq. (i) by 3 and subtracting from Eq. (ii) we get

$$7x = 30 - j10$$

or
$$x = 4.3 - j1.4 = 4.52\angle -18^\circ$$

Substituting this value of x in Eq. (i), we get

$$y = 10 - 3x = 5.95\angle 119.15^\circ$$
$$= -2.9 + j5.2$$
$$x + y = 4.3 - j1.4 - 2.9 + j5.2$$
$$= 1.4 + j3.8$$

1.38 SUPERPOSITION THEOREM (FOR A.C. NETWORKS)

As applicable to ac networks, it states as follows:

In any network made up of linear impedances and containing more than one source of e.m.f., the current flowing in any branch is the phasor sum of the currents that would flow in that branch if each source were considered separately, all other e.m.f. sources being replaced for the time being, by their respective internal impedances (if any).

Note. It may be noted that independent sources can be 'killed' i.e., removed leaving behind their internal impedances (if any) but dependent sources should not be killed.

Example 1.34 *Use Superposition theorem to find the voltage V in the network shown in Fig. 1.81.*

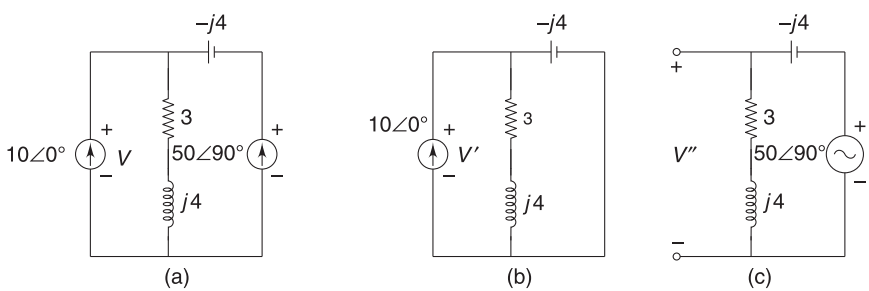


Fig. 1.81

Solution When the voltage source is killed, the circuit becomes as shown in the Fig. 1.81(b). Using current-divider rule,

$$I = 10\ \angle 0° \times \frac{-j4}{(3 + j4) - j4}$$

Now, $V' = I\ (3 + j4)$

$\therefore$ $$V' = 10\ \frac{-j4\ (3 + j4)}{3} = 53.3 - j40$$

Now, when current source is killed, the circuit becomes as shown in Fig. 1.81(c), Using the voltage-divider rule, we have

$$V'' = 50\ \angle 90° \times \frac{(3 + j4)}{(3 + j4) - j4}$$

$$= -66.7 + j50$$

$\therefore$ drop $V = V' + V''$

$$= 53.3 - j40\ (-66.7 + j50)$$

$$= -13.4 + j10 = 16.7\ \angle 143.3°\ \text{V}$$

1.39 THEVENIN'S THEOREM (FOR A.C. NETWORKS)

As applicable to a.c. networks, this theorem may be stated as follows:

The current through a load impedance Z_L connected across any two terminals A and B of a linear network is given by $V_{th}/(Z_{th} + Z_L)$ where V_{th} is the open-circuit voltage across A and B and Z_{th} is the internal impedance of the network as viewed from the open-circuited terminals A and B with all voltage sources replaced by their internal impedances (if any) and current sources by infinite impedance.

Example 1.35 *In the network shown in Fig. 1.82.*

$Z_1 = (8 + j8)\ \Omega$; $Z_2 = (8 - j8)\ \Omega$; $Z_3 = (2 + j20)$; $V = 10\ \angle 0°$ and $Z_L = j10\ \Omega$

Find the current through the load Z_L using Thevenin's theorem.

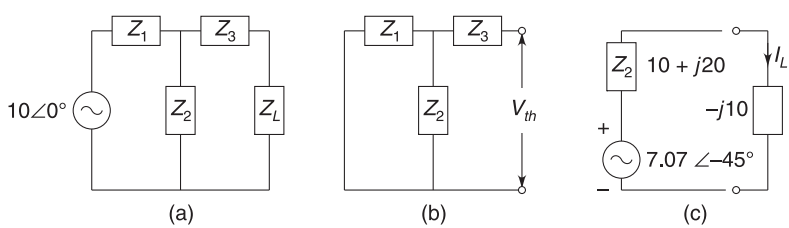


Fig. 1.82

Solution When the load impedance Z_L is removed, the circuit becomes as shown in Fig. 1.82(b). The open-circuit voltage which appears across terminals A and

B represents the Thevenin voltage V_{th}. This voltage equals the drop across Z_2 because there is no current flow through Z_3.

Current flowing through Z_1 and Z_2 is

$$I = V(Z_1 + Z_2) = 10 \angle 0° \, [(8 + j8) + (8 - j8)]$$
$$= 10 \angle 0°/16 = 0.625 \angle 0°$$
$$V_{th} = IZ_2 = 0.625 \, (8 - j8) = (5 - j5)$$
$$= 7.07 \angle - 45°$$

The Thevenin impedance Z_{th} is equal to the impedance as viewed from open terminals A and B with voltage source shorted.

$$\therefore \quad Z_{th} = Z_3 + Z_1 \parallel Z_2$$
$$= (2 + j20) + (8 + j8) \parallel (8 - j8)$$
$$= (10 + j20)$$

The equivalent Thevenin circuit is shown in Fig. 1.82(c) across which the load impedance has been reconnected. The load current is given by

$$\therefore \quad I_L = \frac{V_{th}}{Z_{th} + Z_L}$$
$$= \frac{(5 - j5)}{(10 + j20) + (- j10)} = - j/2$$

1.40 NORTON'S THEOREM (FOR A.C. NETWORKS)

As applied to ac networks, this theorem can be stated as follows:

Any two terminal active linear network containing voltage sources and impedances when viewed from its output terminals is equivalent to a constant current source and a parallel impedance. The constant current is equal to the current which would flow in a short-circuit placed across the terminals and the parallel impedance is the impedance of the network when viewed from open-circuited terminals after voltage sources have been replaced by their internal impedances (if any) and current sources by infinite impedance.

Example 1.36 *Find the Norton's equivalent of the circuit shown in Fig. 1.83. Also find the current which will flow through an impedance of (10 – j20) Ω across the terminals A and B.*

Solution As shown in Fig. 1.83(b), the terminals A and B have been short-circuited.

$$\therefore \quad I_{SC} = I_N = 20/(10 + j20) = 25/22.36 \angle 63.4°$$
$$= 1.118 \angle - 63.4°$$

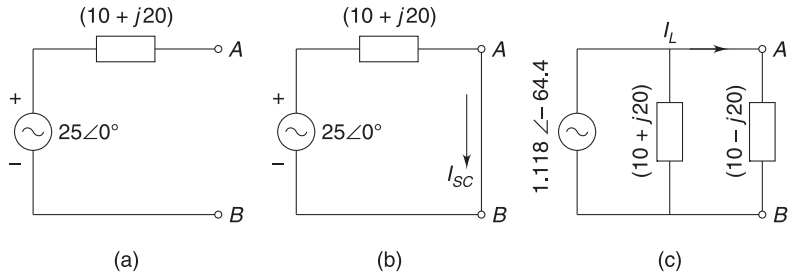


Fig. 1.83

When voltage source is replaced by a short, then the internal resistance of the circuit, as viewed from open terminals A and B, is $R_N = (10 + j20)\ \Omega$. Hence, Norton's equivalent circuit becomes as shown in Fig. 1.83(c).

When the loaf impedance of $(10 - j20)$ is applied across the terminals A and B, current through it can be found with the help of current-divider rule.

$$\therefore \quad I_L = 1.118\ \angle -63.4° \times \frac{(10 + j20)}{(10 + j20) + (10 - j20)}$$

$$= 1.25\ \text{A}$$

1.41 MAXIMUM POWER TRANSFER THEOREMS – GENERAL CASE (FOR A.C. NETWORK)

We will consider the following maximum power transfer theorems when the source has a fixed complex impedance and delivers power to a load consisting of a variable resistance or a variable complex impedance.

Case I When load consists only for a variable resistance R_L (Fig. 1.84(a)). The circuit current is

$$I = \frac{V_g}{\sqrt{(R_g + R_L)^2 + X_g^2}}$$

Power delivered to R_L is

$$P_L = \frac{V_g^2\, R_L}{(R_g + R_L)^2 + X_g^2}$$

To determine the value of R_L for maximum transfer of power, we should set the first derivative dP_L/dR_L to zero.

$$\frac{dP_L}{dR_L} = \frac{d}{dR_L} \frac{V_g^2\, R_L}{(R_g + R_L)^2 + X_g^2}$$

$$= \frac{V_g^2\, [(R_g + R_L)^2 + X_g^2] - R_L\, (2)(R_g + R_L)}{[(R_g + R_L)^2 + X_g^2]^2} = 0$$

or $$R_g^2 + 2R_g R_L + R_L^2 + X_g^2 - 2R_L R_g - 2R_L^2 = 0$$

and $$R_g^2 + X_g^2 = R_L^2$$

$\therefore$ $$R_L = \sqrt{R_g^2 + X_g^2} = |Z_g|$$

It means that with a variable pure resistive load, maximum load, maximum power is delivered across the terminals of an active network only when the load resistance is equal to the absolute value of the impedance of the active network. Such a match is called magnitude match.

Moreover, if X_g is zero, then for maximum power transfer

$$R_L = R_g$$

Case 2 Load impedance having both variable resistance and variable reactance (Fig. 1.84(b)).

The circuit current is

$$I = \frac{V_g}{\sqrt{(R_g + R_L)^2 + (X_g + X_L)^2}}$$

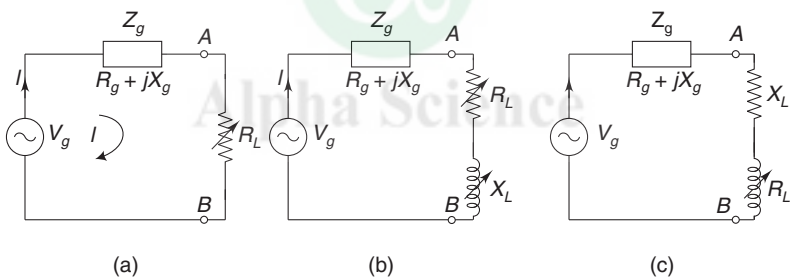


Fig. 1.84

The power delivered to the load is

$$P_L = I^2 R_L = \frac{V_g^2 R_L}{(R_g + R_L)^2 + (X_g + X_L)^2}$$

Now, if R_L is held fixed, P_L is maximum when $X_g = -X_L$. In that case

$$P_{L\max} = \frac{V_g^2 R_L}{(R_g + R_L)^2}$$

If on the other hand, R_L is variable then, as in Case 1 above, maximum power is delivered to the load when $R_L = R_g$. In that case if $R_L = R_g$ and $X_L = -X_g$, then $Z_L = Z_g$. Such a match is called conjugate match.

From the above, we come to the conclusion that in the case of a load impedance having both variable resistance and variable reactance, maximum power transfer across the terminals of the active network occurs when Z_L equals the complex conjugate of the network impedance Z_g i.e., the two impedances are conjugately matched.

Case 3 Z_L with variable resistance and fixed reactance (Fig. 1.84(c)). The equations for current I and power P_L are the same as in Case 2 above except that we will consider X_L to remain constant. When the first derivative of P_L with respect to R_L is set equal to zero, it is found that

$$R_L^2 = R_g^2 + (X_g + X_L)^2$$

and

$$R_L = |Z_g + jZ_L|$$

Since Z_g and X_L are both fixed quantities, these can be combined into a single impedance. Then with R_L variable, Case 3 is reduced to Case 1 and the maximum power transfer takes place when R_L equals the absolute value of the network impedance.

Summary

The above facts can be summarized as under:

1. When load is purely resistive and adjustable, MPT is achieved when $R_L = |Z_g| = \sqrt{R_g^2 + X_g^2}$.
2. When both load and source impedances are purely resistive (i.e., $X_L = X_g = 0$), MPT is achieved when $R_L = R_g$.
3. When R_L and X_L are both independently adjustable, MPT is achieved when $X_L = -X_g$ and $R_L = R_g$.
4. When X_L is fixed and R_L is adjustable, MPT is achieved when $R_L = \sqrt{[R_g^2 + (X_g + X_L)^2]}$

Example 1.37 *In the circuit of Fig. 1.85 which load impedance of p.f. = 0.8 lagging when connected across terminals A and B will draw the maximum power from the source. Also find the power developed in the load and the power loss in the source.*

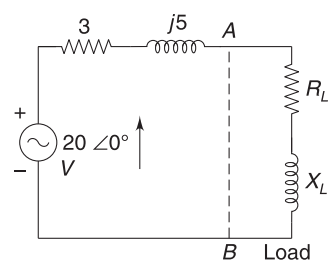


Fig. 1.85

Solution For maximum power transfer

$$|Z_L| = |Z_1| \sqrt{(3^2 + 5^2)}$$

$$= 5.83\ \Omega.$$

For p.f. 0.8, $\cos \phi = 0.8$ and $\sin \phi = 0.6$.

$\therefore$ $$R_L = Z_L \cos \phi = 5.83 \times 0.8 = 4.66\ \Omega$$

$$X_L = Z_L \sin \phi = 5.83 \times 0.6 = 3.5\ \Omega$$

Total circuit impedance

$$Z = \sqrt{[(R_1 + R_L)^2 + (X_1 + X_L)^2]}$$

$$= \sqrt{[(3 + 4.66)^2 + (5 + 3.5)^2]} = 11.44\ \Omega$$

$\therefore$ $$I = V/Z = 20/11.44 = 1.75 \text{ A.}$$

Power in the load =

$$I^2 R_L = 1.75^2 \times 4.66 = 14.3 \text{ W}$$

Powre loss in the source = $1.75^2 \times 3 = 9.2$ W

1.42 A.C. BRIDGES

Resistances can be measured by direct-current Wheatstone bridge, shown in Fig. 1.86(a) for which the condition of balance is that

$$\frac{R_1}{R_2} = \frac{R_4}{R_3}$$

or $$R_1 R_3 = R_2 R_4^{*}$$

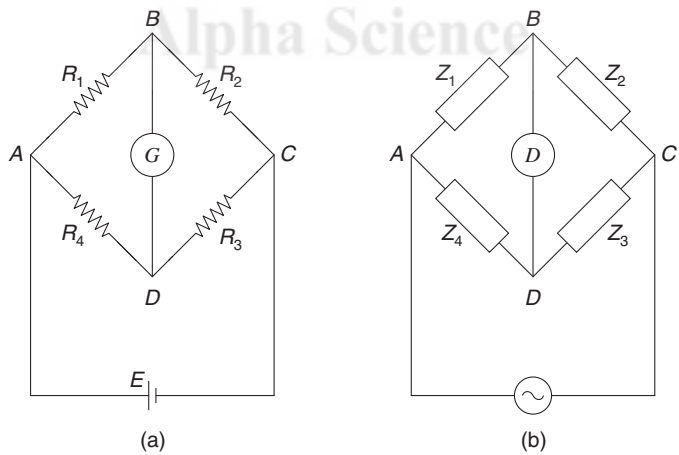


Fig. 1.86

Inductances and capacitances can also be measured by a similar four-arm bridge, as shown in Fig. 1.86(b); instead of using a source of direct current, alternating current is employed and galvanometer is replaced by a vibration

* Products of opposite arm resistances are equal.

galvanometer (for commercial frequencies or by telephone detector if frequencies are higher (500 to 2000 Hz)).

The condition for balance is the same as before but instead of resistances impedances are used i.e.,

$$Z_1/Z_2 = Z_4/Z_3$$

or $$Z_1 Z_3 = Z_2 Z_4$$

But there is one important difference i.e., not only should there be balance for the magnitudes of the impedances but also a phase balance. Writing the impedances in their polar form, the above condition becomes

$$Z_1 \angle\phi_1 \,.\, Z_3 \angle\phi_3 = Z_2 \angle\phi_2 \,.\, Z_4 \angle\phi_4$$

or $$Z_1 Z_3 \angle\phi_1 + \phi_3 = Z_2 Z_4 \angle\phi_2 + \phi_4$$

Hence, we see that, in fact, there are two balance conditions which must be satisfied simultaneously in a four-arm a.c. impedance bridge.

(i) $Z_1 Z_3 = Z_2 Z_4$ … for magnitude balance

(ii) $\phi_1 + \phi_3 = \phi_2 + \phi_4$ … for phase angle balance

Now, we will consider a few bridge circuits used for the measurement of self-inductance, capacitance and mutual inductance, choosing as examples, some bridges which are more common.

1.43 MAXWELL'S INDUCTANCE BRIDGE

The bridge circuit is used for medium inductances and can be arranged to yield results of considerable precision. As shown in Fig. 1.87 in the two arms, there are two pure resistances so that for balance relations, the phase balance depends on the remaining two arms. If a coil of an unknown impedances Z_1 is placed in one arm, then its positive phase angle ϕ_1 can be compensated for in either of the following two ways:

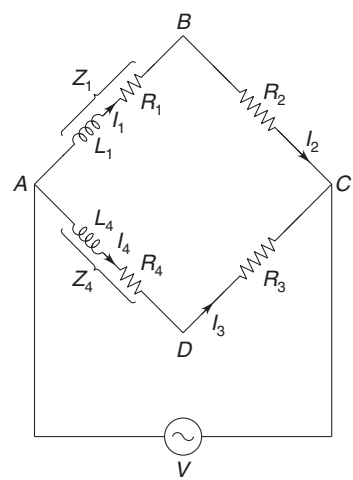


Fig. 1.87

(i) A known impedance with an equal positive phase angle may be used in either of the *adjacent* arms (so that $\phi_1 = \phi_3$ or $\phi_2 = \phi_4$), remaining two arms have zero phase angles (being pure resistances). Such a network is known as Maxwell's a.c. bridge or L_1/L_4 bridge.

(ii) Or an impedance with an equal *negative* phase angle (i.e., capacitance) may be used in *opposite* arm (so that ϕ_1 +

$\phi_3 = 0$). Such a network is known as Maxwell-Wien bridge (Fig. 1.90) or Maxwell's L/C bridge.

Hence, we conclude that an inductive impedance may be measured in terms of another inductive impedance (of equal time constant) in either *adjacent* arm (Maxwell bridge) or the unknown inductive impedance may be measured in terms of a combination of resistance and capacitance (of equal time constant) in the *opposite* arm (Maxwell-Wien bridge). It is important, however, that in each case the time constants of the two impedances must be matched.

As shown in Fig. 1.87

$$Z_1 = R_1 + jX_1 = R_1 + j\omega L_1 \quad \ldots \text{ unknown;}$$

$$Z_4 = R_4 + jX_4 = R_4 + j\omega L_4 \quad \ldots \text{ known}$$

R_2, R_3 = known pure resistance;

D = detector

The inductance L_4 is a variable self-inductance of constant resistance, its inductance being of the same order as L_1. The bridge is balanced by varying L_4 and one of the resistances R_2 or R_3.

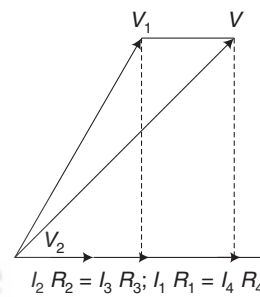


Fig. 1.88

Alternatively, R_2 and R_3 can be kept constant and the resistance of one of the other two arms can be varied by connecting an additional resistance in that arm (Fig. 1.86).

The balance condition is that

$$Z_1 Z_3 = Z_2 Z_4$$

$$\therefore \quad (R_1 + j\omega L_1) R_3 = (R_4 + j\omega L_4) R_2$$

Equating the real and imaginary parts on both sides, we have

$$R_1 R_3 = R_2 R_4$$

or $$R_1/R_4 = R_2/R_3 *^{*}$$

(i.e., products of the resistances of opposite arms are equal).

and $$\omega L_1 R_3 = \omega L_4 R_2$$

or $$L_1 = L_4 \frac{R_2}{R_3}$$

We can also write that

---

* or $\frac{L_1}{R_1} = \frac{L_4}{R_4}$ i.e., the time constants of the two coils are matched.

$$L_1 = L_4 \cdot \frac{R_1}{R_4}$$

Hence, the unknown self-inductance can be measured in terms of the known inductance L_4 and the two resistors. Resistive and reactive terms balance independently and the conditions are independent of frequency. This bridge is often used for measuring the iron losses of the transformers at audio frequency.

The balance condition is shown vectorially in Fig. 1.88. The currents I_4 and I_3 are in phase with I_1 and I_2. This is, obviously, brought about by adjusting the impedances of different branches, so that these currents lag behind the applied voltage V by the same amount. At balance, the voltage drop V_1 across branch 1 is equal to that across branch 4 and $I_3 = I_4$. Similarly, voltage drop V_2 across branch 2 is equal to that across branch 3 and $I_1 = I_2$.

Example 1.38 *The arms of an a.c. Maxwell bridge are arranged as follows: AB and BC are non-reactive resistors of 100 Ω each, DA is a standard variable reactor L_1 of resistance 32.7 Ω and CD comprises a standard variable resistor R in series with a coil of unknown impedance. Balance was obtained with L_1 = 47.8 mH and R = 1.36 Ω. Find the resistance and inductance of the coil.*

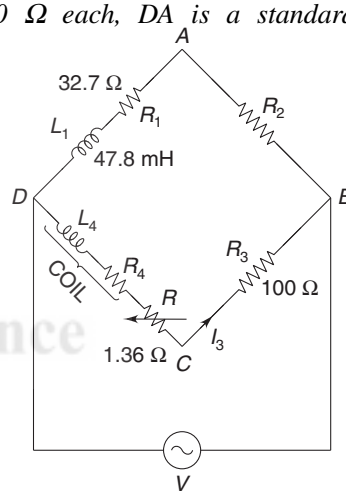


Fig. 1.89

Solution The a.c. bridge is shown in Fig. 1.89.

Since the products of the resistances of opposite arms are equal

$$\therefore \quad 32.7 \times 100 = (1.36 + R_4)100$$

$$\therefore \quad 32.7 = 1.36 + R_4$$

or $$R_4 = 32.7 - 1.36 = 31.34 \ \Omega$$

Since $$L_1 \times 100 = L_4 \times 100$$

$$\therefore \quad L_4 = L_1 = 47.8 \text{ mH}$$

or because time constants are the same, hence

$$L_1/32.7 = L_4/(31.34 + 1.36)$$

$$\therefore \quad L_4 = 47.8 \text{ mH}$$

1.44 MAXWELL-WIEN BRIDGE OR MAXWELL'S L/C BRIDGE

As referred to in Fig. 1.43, the *positive* phase angle of an inductive impedance may be compensated by the *negative* phase angle of a capacitive impedance put in the *opposite* arm. The unknown inductance then becomes known in terms of this capacitance.

Let us first find the combined impedance of arm 1.

$$\frac{1}{Z_1} = \frac{1}{R_1} + \frac{1}{-jX_C}$$

$$= \frac{1}{R_1} + \frac{j}{X_C}$$

$$= \frac{1}{R_1} + j\omega C = \frac{1 + j\omega CR_1}{R_1}$$

$$\therefore \quad Z_1 = \frac{R_1}{1 + j\omega CR_1};$$

$$Z_2 = R_2$$

$$Z_3 = R_3 + j\omega L_3$$

and $$Z_4 = R_4$$

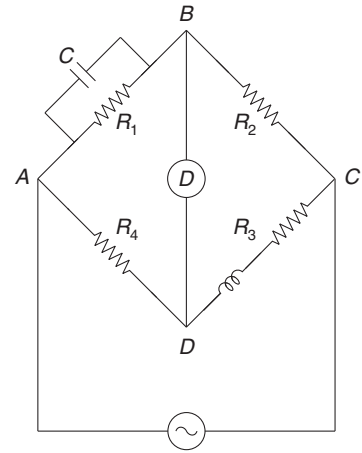


Fig. 1.90

Balance condition is

$$Z_1 Z_3 = Z_2 Z_4$$

or $$\frac{R_1 (R_3 + j\omega L_3)}{1 + j\omega CR_1} = R_2 R_4$$

or $$R_1 R_3 + j\omega L_3 R_1 = R_2 R_4 + j\omega CR_1 R_2 R_4$$

Separating the real and imaginaries, we get

$$R_1 R_3 = R_2 R_4$$

and $$L_3 R_1 = CR_1 R_2 R_4;$$

$$R_3 = \frac{R_2 R_4}{R_1}$$

and $$L_3 = CR_2 R_4$$

Example 1.39 *The arms of an a.c. Maxwell bridge are arranged as follows: AB is a non-inductive resistance of 1,000 Ω in parallel with a capacitor of capacitance 0.5 μF, BC is a non-inductive resistance of 600 Ω, CD is an inductive impedance (unknown) and DA is a non-inductive resistance of 400 Ω. If balance*

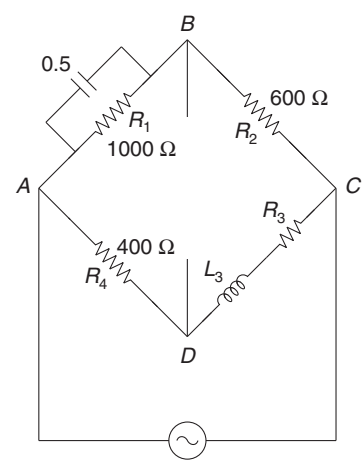


Fig. 1.91

is obtained under these conditions, find the value of the resistance and the inductance of the branch CD.

Solution The bridge is shown in Fig. 1.91. The condition of balance have already been derived in Fig. 1.44 above.

Since $R_1 R_3 = R_2 R_4$

$\therefore \quad R_3 = R_2 R_4 / R_1$

$\therefore \quad R_3 = \dfrac{600 \times 400}{1000} = 240\ \Omega$

Also $L_3 = CR_2 R_4$

$= 0.5 \times 10^{-6} \times 400 \times 600$

$= 12 \times 10^{-2} = 0.12$ H

1.45 ANDERSON BRIDGE

It is a very important and useful modification of the Maxwell-Wien bridge described in Fig. 1.44. In this method, the unknown inductance is measured in terms of a known capacitance and resistance, as shown in Fig. 1.92.

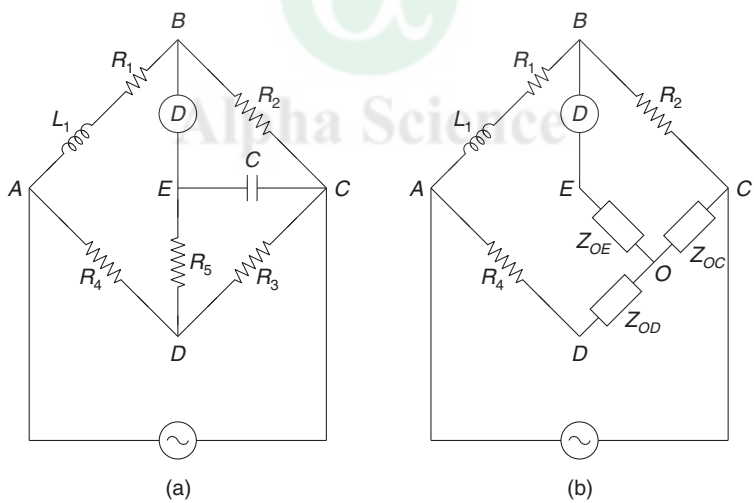


Fig. 1.92

The balance conditions for this bridge may be easily obtained by converting the mesh of impedances C, R_5 and R_3 to an equivalent star with star point O by Δ/Y transformation. As seen from Fig. 1.92(b).

$$Z_{OD} = \frac{R_3 R_5}{(R_3 + R_5 + 1/j\omega C)};$$

$$Z_{OC} = \frac{R_3/j\omega C}{(R_3 + R_5 + 1/j\omega C)} = Z_3$$

With reference to Fig. 1.92(b) it is seen that

$$Z_1 = (R_1\ 1/j\omega L_1)\ Z_2 = R_2;\ Z_3 = Z_{OC}$$

and $$Z_4 = R_4 + Z_{OD}$$

For balance $Z_1Z_3 = Z_2Z_4$

$$\therefore\ (R_1 + j\omega L_1) \times Z_{OC} = R_2\ (R_4 + Z_{OD})$$

$$\therefore\ (R_1 + j\omega L_1)\ \frac{R_3/j\omega C}{(R_3 + R_5 + 1/j\omega C)} = R_2\left(R_4 + \frac{R_3R_5}{R_3 + R_5 + 1/j\omega C}\right)$$

Further simplification leads to $R_2R_3R_4 + R_2R_4R_5 - j\ \frac{R_2R_4}{\omega C} + R_2R_3R_5 - j\ \frac{R_1R_3}{\omega C} + \frac{R_3L_1}{C}$

$$\therefore \qquad \frac{-jR_2R_4}{\omega C} = -\frac{jR_1R_3}{\omega C}$$

or $$R_1 = R_2R_4/R_3$$

Also $$\frac{R_3L_1}{C} = R_2R_3R_4 + R_2R_3R_5 - R_2R_4R_5$$

$$\therefore \qquad L_1 = CR_2\left(R_4 + R_5 + \frac{R_4R_5}{R_3}\right)$$

This is method is capable of precise measurements of inductances over a wide range of values from a few micro-henrys to several henrys and is one of the commonest and the best bridge methods.

1.46 DE SAUTY BRIDGE

With reference to Fig. 1.93, let

C_2 = capacitor whose capacitance is to be measured

C_3 = a standard capacitor

R_1, R_2 = non-inductive resistors

Balance is obtained by varying either R_1 or R_2. For balance, points B and D are at the same potential.

$$\therefore \qquad I_1\ R_1 = I_2\ R_2$$

and $$(-j/\omega C_2)\ .\ I_1 = (-j/\omega C_3)\ .\ I_2$$

Dividing one equation by the other, we get

$$\frac{R_1}{R_2} = \frac{C_2}{C_3};$$

$$C_2 = C_3 \frac{R_1}{R_2}$$

The bridge has maximum sensitivity when $C_2 = C_3$. The simplicity of this method is offset by the impossibility of obtaining a perfect balance if both the capacitors are not free from the dielectric loss. A perfect balance can only be obtained if air capacitors are used.

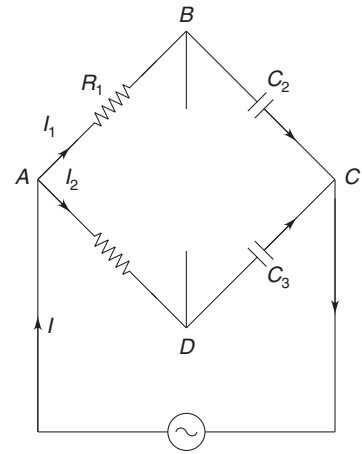


Fig. 1.93

1.47 SCHERING BRIDGE

It is one of the very important and useful methods of measuring the capacitance and dielectric loss of a capacitor. In fact, it is a device for comparing an imperfect capacitor C_2 in terms of a loss-free standard capacitor C_1 (Fig. 1.94(a)). The imperfect capacitor is represented by its equivalent loss-free capacitor C_2 in series with a resistance r (Fig. 1.94(b)).

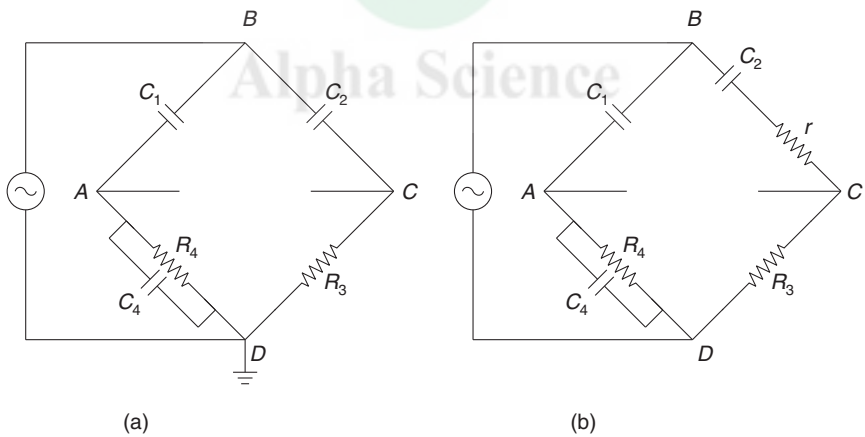


Fig. 1.94

For high voltage applications, the voltage is applied at the junction shown in the figure. The junction between arms 3 and 4 is earthed. Since capacitor impedances at lower frequencies are much higher than resistances, most of the voltage will appear across capacitors. Grounding of the junction affords safety to the operator from the high-voltage hazards while making balancing adjustment in arms 3 and 4.

Now, $$Z_1 = \frac{-j}{\omega C_1}; \quad Z_2 = r - \frac{j}{\omega C_2};$$

$$Z_3 = R_3; Z_4 = \frac{1}{(1/R_4) + j\omega C_4} = \frac{R_4}{1 + j\omega C_4 R_4}$$

For balance, $Z_1 Z_3 = Z_2 Z_4$

or $$\frac{-jR_3}{\omega C_1} = \left(r - \frac{j}{\omega C_2}\right); \left(\frac{R_4}{1 + j\omega\ C_4 R_4}\right)$$

or $$\frac{-jR_3}{\omega C_1}(1 + \omega_a C_4 R_4) = R_4\left(r - \frac{j}{\omega C_2}\right)$$

Separating the real and imaginaries, we have

$$C_2 = C_1 (R_4/R_3)$$

and $$r = R_3 (C_4/C_1).$$

The quality of a capacitor is usually expressed in terms of its phase defect angle or dielectric loss angle which is defined as the angle by *which current departs from exact quadrature from the applied voltage* i.e., the complement of the phase angle. If ϕ is the actual phase angle and δ the defect angle, then $\phi + \delta = 90°$. For small values of δ, tan δ = sin δ = cos ϕ (approximately). Tan δ is usually called the *dissipation factor* of the *R-C* circuit. For low power factors, therefore, dissipation factor is approximately equal to the power factor.

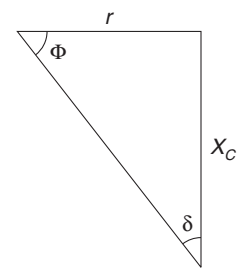


Fig. 1.95

As shown in Fig. 1.95,

Dissipation factor = power factor = tan δ

$$= \frac{r}{X_C} = \frac{r}{1/\omega C_2}$$

$$= \omega r C_2$$

Putting the value of rC_2 from above,

Dissipation factor = $\omega r C_2 = \omega C_4 R_4$

= power factor.

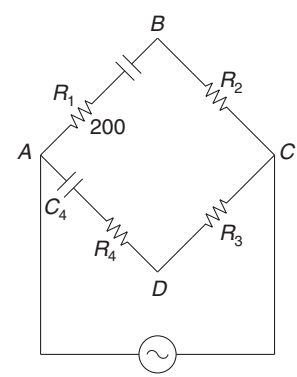


Fig. 1.96

The balance conditions may be obtained in the usual way. For balance

$$R_1 = R_2 R_4/R_3$$

and $$C_1 = C_4 (R_3/R_2)$$

1.48 WIEN PARALLEL BRIDGE

It is also a ratio bridge used mainly as the feedback network in the wide-range audio-frequency *R-C* oscillators. It may be used for measuring audio-frequencies although it is not as accurate as the modern digital frequency meters.

The bridge circuit is shown in Fig. 1.97. In the simple theory of this bridge, capacitors C_1 and C_2 are assumed to be loss-free and resistances R_1 and R_2 are separate resistors.

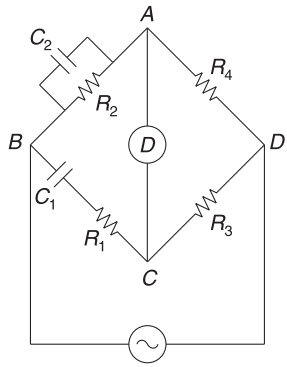


Fig. 1.97

The usual relationship for balance gives

$$R_4\left(R_1 - \frac{j}{\omega C_1}\right) = R_3\left(\frac{R_2}{1 + j\omega C_2 R_2}\right)$$

or $$R_4\left(R_1 - \frac{j}{\omega C_1}\right)(1 + j\omega C_2 R_2) = R_2 R_3$$

Separating the real and imaginary terms, we have

$$R_1 R_4 + R_2 R_4 \frac{C_2}{C_1} = R_2 R_3$$

or $$\frac{C_2}{C_1} = \frac{R_3}{R_4} = \frac{R_1}{R_2} \qquad (1.134)$$

and $$\omega C_2 R_2 R_4 - \frac{R_4}{\omega C_1} = 0$$

or $$\omega^2 = \frac{1}{R_1 R_2 C_1 C_2} \qquad (1.135)$$

or $$f = \frac{1}{2\pi\sqrt{R_1 R_2 C_1 C_2}} \text{ Hz}$$

Note. Equation (1.135) may be used to find angular frequency ω of the source if terms are known. For such purposes, it is convenient to make $C_1 = 2C_2$, $R_3 = R_4$ and $R_2 = 2R_1$. In that case, the bridge has equal ratio arms so that Eq. (1.134) will always be satisfied. The bridge is balanced simultaneously by adjusting R_2 and R_1 (though maintaining $R_2 = 2R_1$). Then, as seen from Eq. (1.135) above

$$\omega^2 = 1/(R_1 \,.\, 2R_1 \,.\, 2C_2 \,.\, C_2)$$

or $$\omega = 1/(2R_1 C_2)$$

Example 1.40 *The arms of a four-arm bridge ABCD, supplied with a sinusoidal voltage, have the following values:*

AB : 200 ohm resistance in parallel with 1 μF capacitor; BC : 400 ohm resistance; CD : 1000 ohms resistance and DA : resistance R in series with a 2 μF capacitor.

Determine (i) the value of R and (ii) the supply frequency at which the bridge will be balanced.

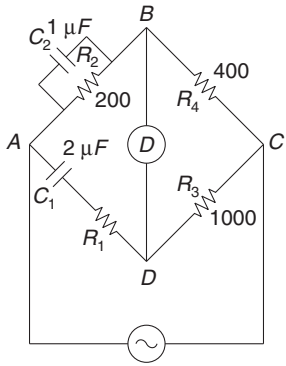


Fig. 1.98

Solution The bridge circuit is shown in Fig. 1.98.

(i) As discussed in Fig. 1.48, for balance we have

$$\frac{C_2}{C_1} = \frac{R_3}{R_4} - \frac{R_1}{R_2}$$

or
$$\frac{2}{1} = \frac{1000}{4000} - \frac{R_1}{200}$$

$\therefore$
$$R_1 = 200 \times 0.5 = 100\ \Omega$$

(ii) The frequency at which bridge is balanced is given by

$$f = \frac{1}{2\pi \sqrt{R_1\, R_2\, C_1\, C_2}}\ \text{Hz}$$

$$= \frac{10^6}{2\pi \sqrt{100 \times 200 \times 1 \times 2}}$$

$$= 796\ \text{Hz}$$

Questions

1.1 What do you understand by the terms 'node', 'mesh', branch', 'linear' and 'bilateral' as applied to networks.

1.2 State and explain Kirchhoff's Voltage Law and Kirchhoff's Current Law.

1.3 Distinguish between *T* and Π networks. Outline the modalities of converting a Π network into a *T* network.

1.4 Transform a *T* network into a Π network.

1.5 State and explain the Superposition theorem.

1.6 Write short note on Thevenin's theorem.

1.7 State and explain Norton's theorem.

1.8 State and prove the maximum power transfer theorem. What is the maximum power developed?

1.9 Write a short note on Wheatstone bridge.

1.10 What is a phasor? Explain its polar notation.

1.11 With the help of an example each, show how Kirchhoff's voltage and current laws can be applied to a.c. circuits.

1.12 Write a note on:

(a) Superposition theorem
(b) Thevenin's theorem
(c) Norton's theorem
(d) Maximum power transfer theorem

as applied to ac networks.

1.13 Write notes on:

(a) Anderson Bridge
(b) De Sauty Bridge
(c) Schering Bridge
(d) Wien Bridge

2 Semiconductor Diodes

2.1 SEMICONDUCTORS

We are familiar with conductors and insulators. *Conducting materials* (e.g. silver, copper, aluminium and gold etc.) are good conductors of electricity and are characterized by a large electrical conductivity and small electrical resistivity. Insulating material (e.g., porcelain, glass, quartz, rubber etc.) are characterized by poor electrical conductivity and are used to block electric current from flowing where it is not to pass. Yet, there is another group of materials (e.g., germanium, silicon, selenium) which are neither good conductors nor good insulators. At room temperature, such materials have conductivities considerably lower than that of conductors and much higher than those of insulators. Such materials are called the *semiconductors*. Their resistivities lie in a range from 10^{-4} Ω-m to about 0.5 Ω-m. (The resistivities of conductors and insulators are, respectively, of the order of 10^{-8} Ω-m and 10^{12} Ω-m). An advantageous property of semiconductors is that *their resistance depends largely on various factors and therefore can be controlled*.

Unlike conductors (metals), semiconductors have a negative temperature coefficient of resistance (i.e. their resistance decreases with the increase in temperature). As a result, the semiconductors behave like insulators at very low temperature but act as conductors at high temperatures.

The resistivity of semiconductors depends on the magnitudes of the electric field in the semiconductor, consequently, the current in it is not proportional to the voltage, as with conductors that obey Ohm's law, but increases by far more than the voltage i.e., semiconductors are *non-linear resistors*.

The semiconductors have another important property – the conductivity can be altered considerably by adding minute amounts of certain other elements called the *impurities*. (This process is known as *doping*). Thus introduction of certain impurities may even change the mechanism of electrical conduction in the semiconductor and certain physical phenomena begin to occur at the junction between sections of the semiconductors having different impurities. This enables us to fabricate important semiconductor devices such as diodes, transistor, SCRs, UJTs, varistors and thermistors.

2.2 FORMATION OF ENERGY BANDS IN SOLIDS

Most metals and semiconductors are crystalline in structure and a crystal is made up of a space array of atoms (a regular repetition in three dimensions of some fundamental structural unit). For a gaseous element, the electronic energy levels are the same as for a single free atom because the individual atoms are well far apart and have negligible influence on each other. However, in a crystal, the individual atoms are so closely packed that the resulting energy levels are modified because of interaction between the atoms. When atoms form a crystal, the energy levels of the inner-shell electrons are not appreciably affected. However, the energy levels of the outer shell electrons are altered considerably as these electrons are shared by adjacent atoms in the crystal. As a result, the coupling between the outer shell electrons of the atoms results in a band of closely spaced energy states instead of the widely separated energy levels of the isolated atoms. Thus, in a solid, a large number of atoms is involved, and their interaction with one another leads to a splitting of each of their energy levels e.g., for a solid having N atoms, each of the levels splits into closely spaced N levels (called a qhasi-continuous energy band) i.e., energy bands are formed as depicted in Fig. 2.1.

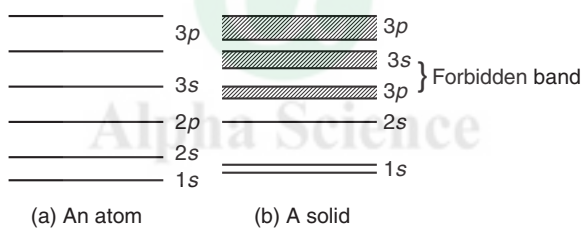


Fig. 2.1 *Allowed energy levels for (a) bound electrons in an atom (b) electrons in a solid*

The bands formed may overlap in some materials whereas they may not overlap in some other materials. The space between the bands are called forbidden *energy gaps*, since electrons in the solid may not have these energies. The energy of the highest filled level is called the *Fermi energy* (E_F). At OK all states up to E_F are full and all above E_F are empty. Because of the Pauli exclusion principle, each s-band of a crystal, consisting of N atoms, has space for $2N$ electrons (i.e., for two electrons per atom). If the highest filled s-band of a crystal is occupied by two electrons per atom (i.e., completely filled), then the electrons cannot drift within the crystal on application of an external electric field. Solids in which the highest filled band is completely occupied by electrons are therefore *insulators*.

2.3 ENERGY BANDS IN DIAMOND

Consider the case of diamond. The isolated carbon atoms have the electronic configuration $1s^2$, $2s^2$, $2p^6$. However, in the diamond crystal, as a result of the

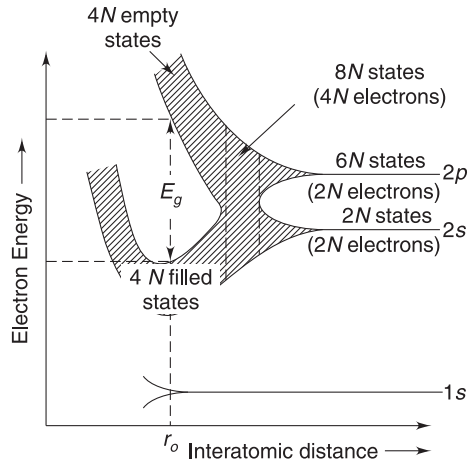


Fig. 2.2 *Development of energy bands in a diamond crystal*

formation of sp^3 hybrid (a mixture of 2*s*- and 2*p*-wave functions with covalent tetrahedral bonding), there is modification of the *s*- and *p*-levels (Fig. 2.2).

As the interatomic spacing is decreased, the 6*N* states of the upper band and the 2*N* states of the lower band merge to give a wider energy band with 8*N* states and 4*N* electrons. With still further reduction in interatomic separation (i.e., at the equilibrium separation), the 2*N* states in the 2*s* subshell and the filled 2*N* states in the 2*p* subshell come together and spread into a band of 4*N* states, all of which are occupied by electrons. The remaining 4*N* empty states in the 2*p* subshell spread into another band, which is separated from the completely filled lower band by a forbidden energy region (*energy gap* = E_g). The completely filled lower band contains the valence electrons and is known as a *valence band*. The upper band is termed as the conduction band because the excitation of electrons from the lower valence band into this band may cause electrical conduction.

For diamond, the sp^3 hybrid stems from the 2*s* and $2p^3$ atomic states, for silicon from the 3*s* and $3p^3$ states, whereas for germanium from the 4*s* and $4p^3$ states. E_g = 5.5 eV for Ge (at 0 K) and 1.2 eV for Si (at 0 K). Diamond is an insulator whereas Ge and Si are *semiconductors*.

2.4 ENERGY BANDS IN METALS, INSULATORS AND SEMICONDUCTORS

Within any given material, there are two distinct energy bands in which electrons may exist. These are the *valence band* and the *conduction band* and are separated by an energy gap in which no electron can normally exist (known as the *forbidden gap* (E_G)).

The energy band of interest is the highest energy band which are occupied by bound electrons. Electrons in the valence band are usually in normal orbit around the nucleus and normally electrons do not have energies to orbit in the forbidden region. If a sufficient amount of energy is given to an electron in the valence band ($\geq E_G$), the electron is free from the atomic structure and goes into the conduction band. Such free electrons can more readily under the influence of an external electric field.

In case of solids, the forbidden energy gap may be large small or non-existent. The distinction between conductors, insulators and semiconductors is largely concerned with the relative widths of the forbidden energy gaps. As illustrated in Fig. 2.3 (a), (b) and (c), there is a wide forbidden gap (> 5 eV) for insulators, narrow forbidden gap (~0.5 to 3 eV) in case of semiconductors and no forbidden gap in case of conductors (metals). In case of insulators, there is practically no electron in the conduction band and the valence band is filled i.e., all the permissible energy levels in the valence band are occupied by the electrons and these are bound electrons. Thus, an electron in a completely filled band cannot contribute to an electric current. Furthermore, for an insulator, the forbidden gap between valence band and conduction band is so large (~5 eV) that at normal temperature (27°C), the energy which can be supplied to an electron from an applied electric field is too small to carry the electron from the filled band to the vacant (conduction) band). Since the electron cannot acquire externally applied (electrical) energy, conduction is impossible. But, at very high temperatures, or under high external electric field, the electrons can jump the gap resulting in breakdown of the insulating material.

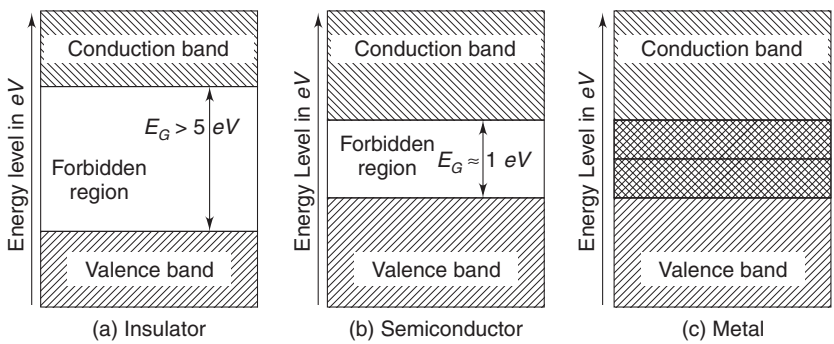


Fig. 2.3 *Energy band structure*

For semiconductors, at absolute zero temperature (0° K or – 273.15°C), the valence band is completely full and the conduction band is completely empty. No current can flow at absolute zero and semiconductor behaves as an insulator. However, the band gap E_G is small (~1 eV) and an electron can be lifted from the

valence band to the conductor band by imparting some energy to it (This energy must be greater than or equal to E_G). The energy imparted by heat at room temperature makes some electrons jump from valence band to the conduction band. This semiconductors are capable of conducting some electric current even at normal ambient temperature.

In case of conductors (metals), there is no forbidden gap, the valence and conduction bands overlap each other. An electron in the conduction band experiences almost negligible nuclear attraction as it does not belong to any particular atom. Due to availability of vacant states in the conduction band, s sight potential difference across the conductor causes the free electrons to constitute an electric current.

2.5 INTRINSIC SEMICONDUCTORS

Semiconductors may be classified into two types *viz.*, intrinsic (or pure) semiconductors and extrinsic (or impure) semiconductors.

An *intrinsic semiconductor* is one which is made of material in its extremely pure form. A semiconductor is not truly intrinsic unless its impurity content is less than one part in 100 million parts of semiconductor. (It is to be noted here that the addition of one part impurity (of the proper type) per million in a wafer of silicon material (a semiconductor) can change that material from a relatively poor conductor to a good conductor of electricity)

There are many semiconductor materials such as Ge (germanium), Si (silicon), grey crystalline tin (Sn), Se (selenium), Te (tellurium), B (boron) etc. available, but Si and Ge are the two most widely used materials in electronics. The reason is that the energy required to break the covalent bonds (E_G) i.e., the energy required to release an electron from their valence bands is very small (1.12 eV for Si and 0.72 eV for Ge). Both elements have the same crystal structure and similar characteristics (both belong to group IV of the periodic table). When an electron jumps from valence band to the conduction band, it leaves behind a hole in the valence band, which acts as a positive charge carrier. Thus, under the influence of an electric field, both electrons (in the conduction band) and holes (in the valence band) take part in conduction.

An intrinsic semiconductor is one in which the concentration of charge carriers is the characteristic of the material itself and charge carriers are equally divided between electrons and holes. Ge and Si are typical examples and both crystallize in diamond structure.

Consider a pure crystal of Si containing no impurities and no lattice defects. Figure 2.4 shows a two-dimensional representation of such a crystal. In the two dimensional array, each Si atom is surrounded by four other Si atoms and bonded

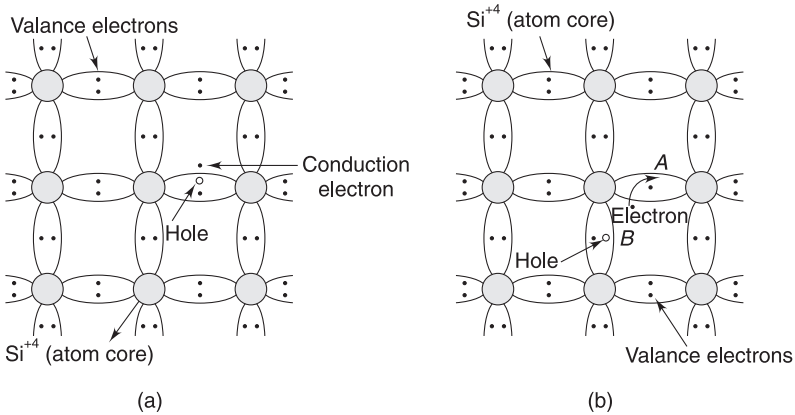


Fig. 2.4 *A two-dimensional schematic representation of silicon crystal showing the atom cores and valence electrons (a) thermal generation of an electron-hole pair and (b) motion of a hole from location A to B*

by covalent bonds, each (bond) formed by sharing of two (outermost) electrons, one from each atom in a pair.

In the figure, the circles represent the atom – cores with four positive charges and dots represent the valence electrons. At 0°K all the electrons are lightly bound in the bonds, but at higher temperatures some of the bound electrons require sufficient energy leaving the bond and more freely inside the crystal. The Fig. 2.4(a) illustrates the formation of an electron – hole pair by thermal generation.

In Fig. 2.4(b), as a result of thermal motion, an electron in a covalent bond at B adjacent to the hole jumps into the original vacant site A, thus transferring the location of the hole to the site (B) from where the electron came (A). In this way, the motion of the hole can be regarded as the transfer of ionization from one atom to next by the jumping of the bound electron from a covalent bond to a neighbouring vacant site. This process does not affect the motion of a free electron. Thus conduction electrons and holes in the valence band can move freely independent of each other.

If p and n represents the hole and electron concentrations respectively, then

$$p = n = n_i$$

For an intrinsic semiconductor, the pair – concentration (n_i) varies with temperature and is different for different semiconductors. (There is a recombination of electron – hole at the same rate at which they are generated. The generation rate increases with the rise in temperature, and a new carrier concentration n_i is established such that the higher recombination rate just balances generation).

2.6 EXTRINSIC SEMICONDUCTORS

The characteristics of semiconductor materials can be altered significantly by the addition of certain impurity atoms into the intrinsic materials. These impurities, although added to nearly 1 part in 10 million, can alter the band structure sufficiently to totally change the electrical properties of the material.

A semiconductor material that has been subjected to the doping process is called an extrinsic semiconductor.

There are two extrinsic types of semiconductors which are of immense importance to semiconductor device fabrication: *N*-type and *P*-type.

2.6.1 *N*-type Semiconductor

The *N*-type semiconductor is created by introducing those impurity elements that have five valence electrons (substitutional impurity) such as phosphorous, arsenic or antinomy. (These belong to group V of the periodic table). The effect of this addition is indicated in Fig. 2.5. There are still four covalent bonds. However, there is an extra electron per impurity atom available for conduction. This fifth electron which is unassociated with any covalent bond is loosely bound to its parent (As) atom and is relatively free to move within the newly formed *n*-type material. The substituted impurity atom thus donates a relatively free-electron to the structure consequently, diffused impurities with 5 valence electrons are called *donor atom*.

The effect of the doping process can be described by the use of energy-band diagram (Fig. 2.5(b)). A discrete energy level called the *donor level* appears in the forbidden band near the conduction band edge with an E_g significantly less than that of the intrinsic material. As a result, free electrons have less difficulty in absorbing a sufficient thermal energy to move into the conduction band, at room temperature. (Electrons are majority carriers and holes the minority carriers).

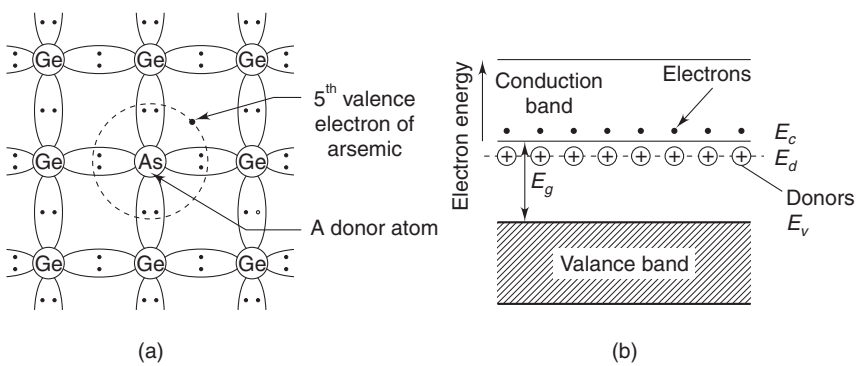


Fig. 2.5 *(a) n-type semiconductor. Note that although free carriers have been established in the material, still it is electricity neutral (since As is As^{+5}). (b) Energy band diagram of n-type semiconductor (creation of a donor level)*

2.6.2 *P*-type Semiconductor

The *P*-type material is formed by doping a pure Ge or Si crystal with impurity atoms having three valence electrons. The elements used for this purose are boron, gallium or indium. (These belong to Group III of the periodic table). The effect of this substitution (for 1 atom of Ge) is indicated in Fig. 2.6.

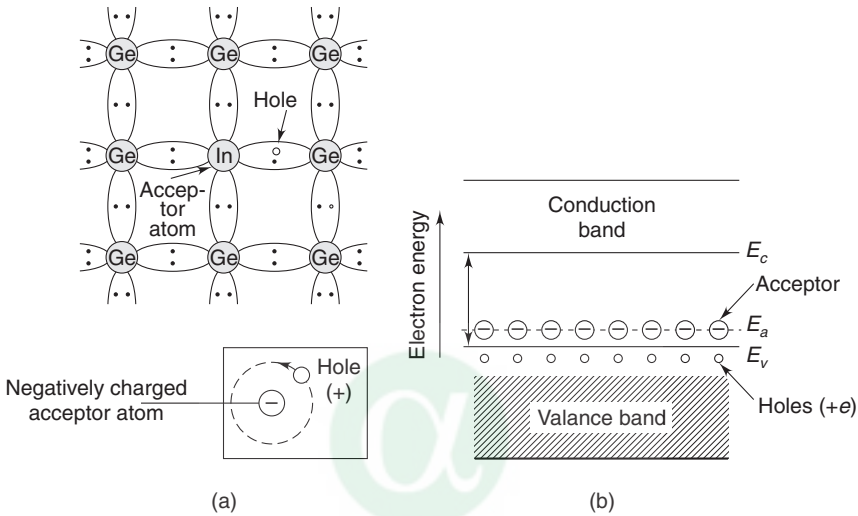


Fig. 2.6 *(a) P-type semiconductor. Although free carriers (holes) have been established, the material as a whole is electrically neutral, because 'In' which has three valence electrons acquires one negative charge, contributing a positive charged hole (b) energy band diagram of p-type semiconductor*

There is now an insufficient number of electrons to complete the covalent bonds. The resulting vacancy is called a hole and is represented by a small circle. Since the resulting vacancy will readily accept a free electron, the diffused impurities with three valence electrons. (In behaves as In^{+3}) are called *acceptor atoms* (one electron is missing). A discrete energy level called the *acceptor-level* appear in the forbidden energy gap near the valence band edge. In this case, holes are the majority carriers and electrons, the minority carriers.

2.7 MASS ACTION LAW

The concentration of free electrons and holes is always equal in an intrinsic semiconductor. When *N*-type impurity is added to an intrinsic semiconductor, the concentration of free electrons is increased (or the concentration of holes is reduced below the intrinsic value). Similarly, addition of *P*-type impurity results in reduction of free electrons below the intrinsic value. Theoretically, it has

been found that under thermal equilibrium, the product of concentration of free electrons and concentration of holes is contant (= n_i^2) i.e.,

$$np = n_i^2 \tag{2.1}$$

and is independent of the amount of doping by donor or acceptor impurities. Here n_i is the intrinsic concentration and is a function of temperature. For an intrinsic semiconductor,

$$n = p = n_i$$

This law is very important as it enables us to determine minority carries concentration. According to this law, the addition of impurities to an intrinsic semiconductor increases the concentration of one type of carriers, which consequently become majority carriers and simultaneously decreases the concentration of the other carriers which, as a result, become minority carriers.

2.8 CHARGE DENSITIES IN AN EXTRINSIC SEMICONDUCTORS

Equation $np = n_i^2$ gives relationship between the electron and hole concentrations n and p respectively.

If N_D is the concentration of donor atoms, there will be N_D immobile positive charges per unit volume contributed by donor ions as the donor atoms are ionized. The positive charge density will be $N_D + p$. (where p is the hole density). Similarly as the acceptor atoms are ionized at room temperature, there will be N_A immobile negative charges per unit volume and total negative charge density will be $N_A + n$, where N_A is the concentration of acceptor atoms (and n is the electron density). Since the semiconductors is electrically neutral, the magnitude of positive charge density will be equal to the magnitude of negative charge density i.e.,

$$N_D + p = N_A + n \tag{2.2}$$

In N-type semiconductors, there is no acceptor doping i.e., $N_A = 0$. Also, the number of electrons is much greater than the number of holes i.e., $n >> p$. So

$$n \approx N_D \tag{2.3}$$

i.e., in N-type semiconductors, the free electron concentration is approximately equal to the density of donor atoms.

Now, the concentration of holes in N-type semiconductors will be given by the equation

$$p = \frac{n_i^2}{n} = \frac{n_i^2}{N_D} \tag{2.4}$$

Similarly, in case of *P*-type semiconductors

$$p \approx N_A \tag{2.5}$$

and

$$n = \frac{n_i^2}{N_A} \tag{2.6}$$

2.9 DRIFT AND DIFFUSION CURRENTS

The two basic processes which cause electrons and holes to move in a semiconductor are: (i) *drift*, which is the movement caused by electric fields, and (ii) *diffusion*, which is the flow caused by variations in the concentration, i.e., concentration gradients. Such gradients can be caused by a non-homogeneous doping distribution, or by injection of a quantity of electrons or holes into a region.

2.9.1 Drift Current

When a steady electric field E volts/metre is applied to a metal, the electrons move to the positive terminal of the applied voltage. In their way, they continually collide with the atoms and rebound in a random fashion, as illustrated in Fig. 2.7. Each collision being inelastic i.e., the electrons lose some kinetic energy. After the collision, the electrons are accelerated and gain certain component of velocity in the direction opposite to that of applied electric field ($-E$) and lose their energy at the next collision. Thus, the applied electric field does not stop collisions and random motion but makes the electrons to drift towards the positive terminal. Consequently, the electrons gain average drift velocity v in the direction opposite to that of the applied electric field. It is observed that the drift velocity v is proportional to the applied electric field E,

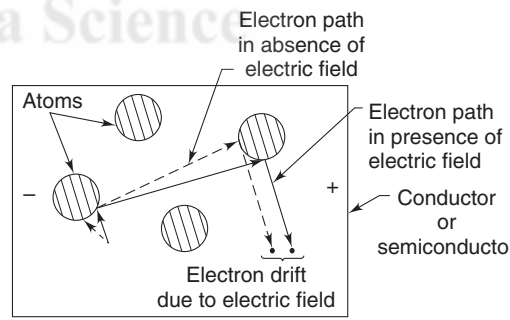


Fig. 2.7

i.e., $$v \propto E \quad \text{or} \quad v = \mu_e E \tag{2.7}$$

where μ_e is called the electron *mobility*[*] and is expressed in m^2/V-s.

This steady-state drift velocity is superimposed on the random motion of free electrons caused by thermal agitation. This steady flow of electrons in one direction caused by the applied electric field constitutes an electric current, called the *drift current*.

[*] Mobility is defined as the steady-state drift velocity per unit electric field.

2.9.2 Diffusion Current

A second type of current, called the *diffusion current*, also exists in a semiconductor. The diffusion of charge carriers is as a result of a gradient of carrier concentration (i.e. the difference of carrier concentration from one region to another). In this case concentrations of charge carriers (either electrons or holes) tend to distribute themselves uniformly throughout the semiconductor crystal. This movement continues until all the carriers are evenly distributed throughout the material. This type of movement of charge carriers is called the *diffusion current*.

Diffusion current occurs when charge carriers diffuse from a point of concentration, to spread uniformly throughout the volume of a piece of material, as illustrated in Fig. 2.8.

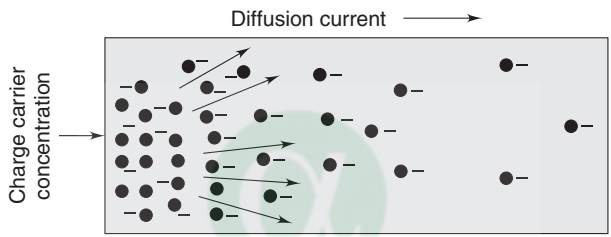


Fig. 2.8 *Flow of particles from a region of high concentration to a region of low concentration*

Both drift and diffusion occur simultaneously in a semiconductor device.

2.10 FERMI LEVEL IN AN INTRINSIC SEMICONDUCTOR

Electrons and holes in a semiconductor are Fermions and obey Fermi-Dirac statistics. The Fermi-Dirac energy distribution function enables us to find the fraction of electrons having energy E (eV) which are occupied under condition of thermal equilibrium. This Fermi-Dirac probability function is given by

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}}$$

where E_F is the Fermi energy (i.e. energy of the Fermi level) k is the Boltzmann constant ($k = 8.62 \times 10^{-5}$ eV/K) and T is the absolute temperature of the semiconductor. For $E = E_F$,

$$f(E)\big|_{E = E_F} = \frac{1}{1 + e^{o/kT}} = \frac{1}{1 + 1} = \frac{1}{2}$$

Thus, the Fermi level has the probability of 50% of being occupied by an electron.

In case of a semiconductor, the Fermi level is a reference energy level. *It is the energy level at which the probability of finding an electron n energy units above it (in the conduction band) is equal to the probability of finding a hole (in the valence band) n energy units below it in an intrinsic semiconductor.*

Let at any temperature T°K,

number of electrons in the conduction band = n_c

number of electrons in the valence band = n_v

and total number of electrons $n = n_c + n_v$

Let (for make of reference), the zero-energy level be taken at the top of the valence band ($E_v = 0$) and energy at the bottom of the conductive band $E_C = E_G$ (Fig. 2.9).

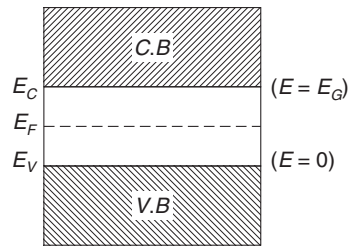


Fig. 2.9

Then, number of electrons in the conductive band

$$n_c = n\, f\,(E_G)$$

where $f\,(E_G)$ is the probability of an electron having energy E_G.

$$\therefore \qquad f\,(E_G) = \frac{1}{1 + e^{(E_G - E_F)/kT}}$$

and

$$n_c = \frac{n}{1 + e^{(E_G - E_F)/kT}} \tag{2.8}$$

The number of electrons in the valence band

$$n_v = n\, f\,(o)$$

$$f\,(o) = \frac{1}{1 + e^{(o - E_F)/kT}} = \frac{1}{1 + e^{-E_F/kT}}$$

$$\therefore \qquad n_v = \frac{n}{1 + e^{-E_F/kT}} \tag{2.9}$$

Now, total number of electrons in both bands

$$n = n_c + n_v$$

or

$$n = \frac{n}{1 + e^{(E_G - E_F)/kT}} + \frac{n}{1 + e^{-E_F/kT}}$$

or

$$1 - \frac{1}{1 + e^{-E_F/kT}} = \frac{1}{1 + e^{(E_G - E_F)/kT}}$$

or

$$E_F = \frac{1}{2} E_G.$$

i.e., in an intrinsic semiconductor, the Fermi level lies midway between the conduction and valence bands.

2.11 FERMI LEVEL IN AN EXTRINSIC SEMICONDUCTOR

The equation for $f(E)$ specifying the fraction of all states at energy E (electron volts) occupied under conditions of thermal equilibrium is called the *Fermi-Dirac probability* function and is given as

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}} \tag{2.10}$$

where k is Boltzmann constant in eV/K, T is temperature in K and E_F is the *Fermi level* or *characteristic energy* for crystal in eV.

Concentrations of free electrons n and concentration of holes p are given as

$$n = n_C \, e^{-(E_C - E_F)/kT} \tag{2.11}$$

and

$$p = n_V \, e^{-(E_F - E_V)/kT} \tag{2.12}$$

where n_C is the number of electrons in conduction band, n_V is the number of electrons in the valence band, E_C is the lowest energy in the conduction band in eV and E_V is the maximum energy of valence band in eV.

The only parameter that changes with the addition of impurity in Eqs. (2.11) and (2.12) is the Fermi level.

Fermi Level in N-type Extrinsic Semiconductor In case of an intrinsic semiconductor, E_F lies in the middle of the energy gap, indicating equal concentrations of free electrons and holes.

When a donor type impurity is added to the crystal, then, at a given temperature and assuming all donor atoms are ionized, the first N_D states in the conduction band will be filled. Hence it will be more difficult for the electrons from the valence band to bridge the energy gap by the thermal agitation. Consequently the number of electron-hole pairs thermally generated for that temperature will be decreased. Since the Fermi level is a measure of the probability of occupancy of the allowed energy states, it is clear that E_F must be more closer to the conduction band to indicate that many of the energy states in that band are filled by the donor electrons, and fewer holes exist in the valence band. This situation is illustrated in Fig. 2.10, for an *N*-type extrinsic semiconductor.

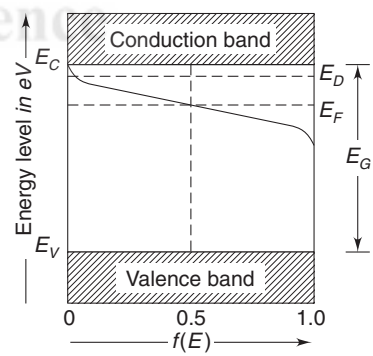


Fig. 2.10 *Position of fermi level in a n-type semiconductor*

The same kind of argument leads to the conclusion that E_F must move from the centre of the forbidden gap closer to the valence band for a *P-type extrinsic semiconductor* (Fig. 2.11).

Exact position of the Fermi level in an *N*-type extrinsic semiconductor can be determined by substituting $n = N_D$ in Eq. (2.11) and thus we have

$$N_D = n_C \, e^{-(E_C - E_F)/kT} \qquad (2.13)$$

Taking logarithm we have,

$$\log_e N_D = \log_e n_C - \frac{(E_C - E_F)}{kT}$$

Solving the above equation we have,

$$E_F = E_C - kT \log_e \frac{n_C}{N_D} \qquad (2.14)$$

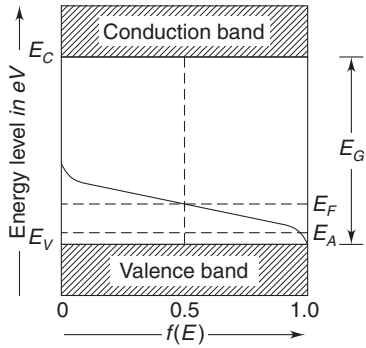


Fig. 2.11 *Position of fermi level in an p-type semiconductor*

Similarly for the *P*-type extrinsic semiconductor,

$$E_F = E_V + kT \log_e \frac{n_V}{N_A} \qquad (2.15)$$

2.12 CURRENT IN A SEMICONDUCTOR (CONDUCTIVITY AND MOBILITY)

As already mentioned in Section 2.9, conduction through a semiconductor is due to two mechanisms *viz.*, drift and diffusion.

The movement of charge carriers (electrons and holes) under the influence of an applied electric field E is termed as the *drift current*.

The conductivity σ_e of the semiconductor due to electrons in the conduction band is given as

$$\sigma_e = ne\mu_e \qquad (2.16)$$

where n is the number of electrons per unit volume of the conductor (i.e., electron density), e is the electronic charge and μ_e is the electron mobility. Similarly, conductivity due to holes

$$\sigma_h = pe\mu_h \qquad (2.17)$$

where p is the number of holes per unit volume and μ_h is the hole mobility.

Since in semiconductors the conduction is by free electrons as well as by holes, total conductivity is

$$\sigma = \sigma_e + \sigma_h = e\,(n\,\mu_e + p\,\mu_h) \qquad (2.18)$$

Current Density $\quad J = \sigma E \qquad (2.19)$

In case of *N*-type semiconductor, hole concentration is negligible and electron concentration $n_n = N_D$ (density of donor atoms). So conductivity

$$\sigma_n = e\, N_D\, \mu_e \tag{2.20}$$

Similarly, in case of *P*-type semiconductor, electron concentration is negligible and hole concentration

$$p_h = N_A \quad \text{(density of acceptor atoms)}$$

and conductivity

$$\sigma_h = e\, N_A\, \mu_h \quad (\because \text{hole charge} = +e) \tag{2.21}$$

Total current density due to drift of electron and holes on application of electric field E

$$\begin{aligned} J &= \text{Current density due to hole drift } J_h \\ &\quad + \text{current density due to electron drift } J_e \\ &= Ee\,(p\mu_h + n\mu_e) \end{aligned} \tag{2.22}$$

Now, the carrier currents are also due to concentration gradients in the doped material which leads to *diffusion of carriers* from high concentration region to low concentration region. The current due to this process is called the diffusion current (Fig. 2.12).

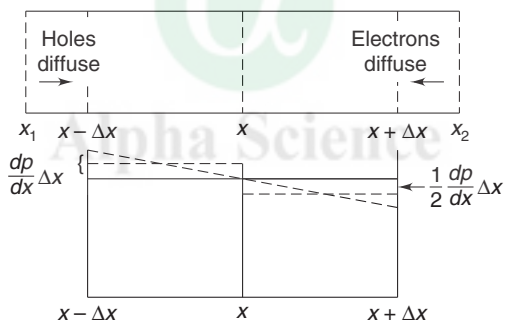


Fig. 2.12

Let us assume that at reference plane x, the density of holes is p per m^3. Let the carriers have a mean free path Δx and mean free time Δt between collisions. Assuming uniform density gradient of holes as $-dp/dx$, the average density of holes in the region $x - \Delta x$ and x is

$$= \frac{1}{2}\left[p + p - \frac{dp}{dx}\Delta x\right] = p - \frac{1}{2}\frac{dp}{dx}\Delta x \tag{2.23}$$

Flow of holes to the right per unit time

$$p_1 = \frac{1}{2}\left[p - \frac{1}{2}\frac{dp}{dx}\Delta x\right]\frac{\Delta x}{\Delta t} \tag{2.24}$$

Left directed hole flow across reference plane x per unit time

$$p_2 = \frac{1}{2}\left[p + \frac{1}{2}\frac{dp}{dx}\Delta x\right]\frac{\Delta x}{\Delta t} \tag{2.25}$$

The net rate of flow of holes is to the right and is given as

$$\begin{aligned} p &= p_1 - p_2 \\ &= \frac{1}{2}\left[p - \frac{1}{2}\frac{dp}{dx}\Delta x\right]\frac{\Delta x}{\Delta t} - \frac{1}{2}\left[p + \frac{1}{2}\frac{dp}{dx}\Delta x\right]\frac{\Delta x}{\Delta t} \\ &= -\frac{1}{2}\Delta x\,\frac{\Delta x}{\Delta t}\,\frac{dp}{dx} \equiv -D_h\frac{dp}{dx} \end{aligned} \tag{2.26}$$

where $D_h = \frac{1}{2}\Delta x\,\frac{\Delta x}{\Delta t} = \frac{1}{2}\frac{\Delta x^2}{\Delta t}$ m^2/s is called the *diffusion constant due to holes*.

The current density due to diffusion of holes is given as

$$J_h = -e\,D_h\frac{dp}{dx} \tag{2.27}$$

Similarly, current density due to diffusion of electrons is given as

$$J_e = e\,D_e\frac{dn}{dx} \tag{2.28}$$

Total Current

It is possible that a potential gradient and a concentration gradient may exist simultaneously within a semiconductor. In such a case, the total current is the sum of drift current due to potential gradient and the diffusion current due to charge carrier concentration gradient. Thus,

$$J_h = E\,e\,p\,\mu_h - e\,D_h\frac{dp}{dx} \tag{2.29}$$

$$J_e = E\,e\,n\,\mu_e + e\,D_e\frac{dn}{dx} \tag{2.30}$$

The various constants for silicon and germanium are given in Table 2.1.

Table 2.1 *Properties of Si and Ge*

Property	Silicon	Germanium
Atomic number	14	32
Atomic weight	28.1	72.6
Density	2,330 kg/m^3	5,320 kg/m^3
Dielectric constant	12	16

Contd..

Contd..

Atoms/m^3	5.0×10^{28}	4.4×10^{28}
E_{G0} at 0K	1.21 eV	0.785 eV
E_G at 300K	1.12 eV	0.72 eV
Intrinsic concentration n_i at 300 K	$1.5 \times 10^{16}/m^3$	$2.5 \times 10^{19}/m^3$
Intrinsic resistivity ρ_i at 300 K	2,300 Ω-m	0.45 Ω-m
Electron mobility, μ_e	0.13 m^2/V-s	0.38 m^2/V-s
Hole mobility, μ_h	0.05 m^2/V-s	0.18 m^2/V-s
Diffusion constant, D_e	0.0034 m^2/s	0.0099 m^2/s
Diffusion constant, D_h	0.0013 m^2/s	0.0047 m^2/s

2.13 EINSTEIN RELATIONSHIP

Diffusion constant D and mobility μ are statistical thermodynamics phenomena and are, therefore, not independent of each other. The diffusion constants D_h and D_e, and mobility μ_h and μ_e are related by Einstein's equation:

$$\frac{D_h}{\mu_h} = \frac{D_e}{\mu_e} = V_T$$

where V_T is the *volt-equivalent of temperature*.

2.14 THE CONTINUITY EQUATION

The continuity equation is one of the most important and fundamental equations in fluid mechanics and electrostatics. In the case of flow of an incompressible fluid, the continuity equation states that the rates of flow of an incompressible fluid entering into and coming out from an elementary volume must be equal provided there is no source or sink of the fluid in that volume. So far as the charge carriers (e.g. free electrons and holes) in a semiconductor are concerned, the situation is slightly more complicated due to the following three processes. Firstly, electrons and positive holes are being continuously generated due to thermal excitation and, at the same time, they recombine with each other after an average time T known as the life time of the charge carriers. Secondly, the charge carriers in a semiconductor can move with a drift velocity when they come under the influence of an electric field. Thirdly free electrons and holes can also move when there is a concentration gradient along the length of the sample. These three processes may even occur simultaneously. Thus, in order to describe the behaviour of excess carriers in a semiconductor, the effects of these three processes must be combined to form the continuity equation which is extremely important in the study of properties of semiconductor diodes and transistors.

(a) The Generation and Recombination of Free Electrons and Holes

Free electrons and holes are generated by thermal excitation and, in equilibrium, they recombine at the same rate as they are generated so that their density remains constant. But in the non-equilibrium condition, the rate of increase of positive hole density $\frac{dp}{dt}$ with time is equal to the difference between the rate of generation due to thermal excitation $\frac{p_o}{T_p}$ and the rate of recombination $\frac{p}{T_p}$ provided only these two mechanisms are present. It may, thus, be expressed by

$$\frac{dp}{dt} = \frac{p_o - p}{T_p} \tag{2.31}$$

where p_o is the equilibrium density of holes in the semiconductor, p is the density of holes at any time t and T_p is the lifetime of holes.

By a similar argument,

$$\frac{dn}{dt} = \frac{n_o - n}{T_n} \tag{2.32}$$

where n_o is the density of the free electrons in the equilibrium condition n is the density of free electrons in the semiconductor, at any time t and T_n is the lifetime of electrons.

These two equations account for the increase of density of charge carriers due to generation and recombination.

(b) The Drift of Charge Carriers in a Uniform Field of *F* Volt/cm Applied along the Sample

When an electric field F is applied along the X-axis of the semiconductor sample containing free carriers such as free electrons and holes, the holes move in the X-direction with an average drift velocity V_{dp} where

$$V_{dp} = \mu_p F \tag{2.33}$$

and μ_p is the mobility of holes in the semiconductor. This movement of holes constitutes a current density J_{1p} in the X-direction; the hole current density J_{1p} is given by

$$J_{1p} = pe\ V_{dp} = pe\ \mu_p F \tag{2.34}$$

where p is the density of holes in the semiconductor and e is the electronic charge. Similarly, the electron current density J_{1n} is given by

$$J_{1n} = ne\ \mu_n F \tag{2.35}$$

where n is the density of free electrons in the semiconductor and μ_n is the mobility of free electrons. The stream of electrons move opposite to the X-direction, while the current due to electrons flows in the X-direction.

(c) The Diffusion of Charge Carriers in a Semiconductor Sample having Concentration Gradient along its Length

The charge carriers (free electrons or holes) can move in a semiconductor sample due to diffusion and their motion is independent of the electric field. Diffusion occurs whenever there is a non-uniform density of charge carriers. Thus, the charge carriers such as free electrons or holes move from a region of higher concentration to a region of lower concentration. Suppose the hole density is one region in a semiconductor sample is increased by some external means. The excess of holes produced will move from the region of higher density of holes to a region of lower density so that the density of holes in the material becomes more uniform in the sample. Diffusion is the flow of charge carriers in the direction of decreasing carrier concentration regardless of the nature of the electric charge on the carriers.

The number of holes N_p diffusing per unit area per second in the X-direction is proportional to the negative gradient of hole density in the X-direction $\left(-\frac{dp}{dz}\right)$.

Thus

$$N_p = - D_p \frac{dp}{dz} \tag{2.36}$$

where D_P is the diffusion constant for holes in the semiconductor sample.

The current density due to diffusion of holes J_{2p} is

$$J_{2p} = eN_p = - eD_p \frac{dp}{dx} \tag{2.37}$$

The total current density J_p due to the movement of holes can be found by adding the two components of the hole current densities due to the drift and diffusion of holes. Thus, J_p is given by

$$J_p = J_{1p} + J_{2p} = pe\mu_p F - eD_p \frac{dp}{dz} \tag{2.38}$$

The number of free electrons N_e diffusing per unit area per second in the X-direction is proportional to the negative gradient of electron density in the X-direction $\left(-\frac{dn}{dz}\right)$. Thus

$$N_e = - D_n \frac{dn}{dx} \tag{2.39}$$

where D_n is the diffusion constant for free electrons in the semiconductor sample. The current J_{2n} due to diffusion of free electrons is

$$J_{2n} = - eD_n \left(-\frac{dn}{dx}\right) = eD_n \frac{dn}{dx} \tag{2.40}$$

since the free electrons carry negative charge.

When the drift and diffusion current densities of free electrons are considered, the total electron current J_n is given by

$$J_n = J_{1n} + J_{2n} = ne\mu_n F + eD_n \frac{dn}{dx} \tag{2.41}$$

(d) Derivation of the Continuity Equation

Figure 2.13 shows a volume element in a semiconductor. In this three-dimensional volume element the area in the *Y-Z* plane is equal to unity ($dy.dz = 1$) and the length of the volume element in the *X*-direction is dx. The total volume of the volume element $dV = dxdydz = dx$ (since $dy.dz = 1$). We shall derive the continuity equation for the positive holes.

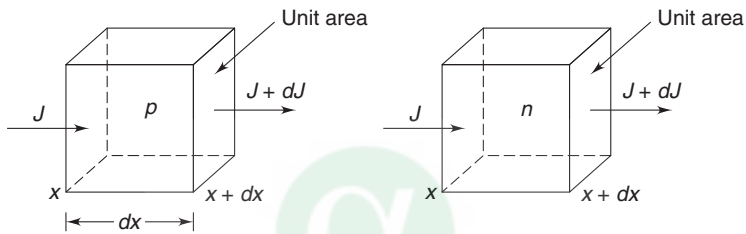


Fig. 2.13 *Volume element used in the derivation of the continuity equation for holes*

Suppose the average density of holes in the volume element at any time t is p per unit volume. Considering only the one-dimensional case, we assume that the density of hole current due to both drift and diffusion of holes, entering into the volume element by the left hand face is J_p and the current density leaving the volume element through the right hand face is $(J_p + dJ_p)$. The difference between these two current densities indicates that an amount dJ_p of hole current is lost in the volume element. In other words, $\frac{dJ_p}{e}$ number of positive holes are being removed from the volume element every second by the hole current flow on account of drift and diffusion. $\frac{dp}{dt}$ denotes the rate of increase of positive hole density due to drift and diffusion. Thus, the increase of positive holes in the volume element may be obtained by multiplying $\frac{dp}{dt}$ by the volume of the element, dx. The rate of increase of the total number of positive holes in the volume element is equal to $\frac{dp}{dt}.dx$ where dx is the volume of the element. If the net current flow is responsible for this change in the number of holes contained in the volume element, we may write

$$\frac{dp}{dt}.dx = -\frac{dJ_p}{e}$$

or
$$\frac{dp}{dt} = -\frac{1}{e}\frac{dJ_p}{dx} \tag{2.42}$$

This equation takes account of the drift and diffusion currents. When the generation and recombination processes are also simultaneously considered, the continuity equation for the holes becomes

$$\frac{\partial p}{\partial t} = \frac{p_o - p}{T_p} - \frac{1}{e}\frac{dJ_p}{dx} \tag{2.43}$$

On substitution of the value of J_p from Eq. (2.38) in Eq. (2.43) the continuity equation for the holes becomes

$$\frac{\partial p}{\partial t} = \frac{p_o - p}{T_p} - \mu_p \frac{\partial (pF)}{dx} + D_D \frac{d^2 p}{dx^2} \tag{2.44}$$

In order to derive the continuity equation for free electrons an elementary volume as in Fig. 2.13 is considered so that an electric current J_n enters through the left hand face and $(J_n + dJ_n)$ leaves the right hand face of the volume element, where J_n is the current density due to the drift and diffusion of free electrons. Using the argument as before, we can sho that

$$\frac{\partial n}{\partial t} = \frac{1}{e}\frac{dJ_n}{dx} \tag{2.45}$$

In this case we find that there is a positive sign before $\left(\frac{1}{e}\right)$, because the free electron carries a negative charge and the direction of flow of the conventional current is opposite to the direction of the flow of electrons.

If we now consider the change in the density of free electrons due to generation and recombination and substitute the value of J_n from Eq. (2.41) in Eq. (2.45), the continuity equation for free electrons may be written as

$$\frac{\partial n}{\partial t} = \frac{n_o - n}{T_n} + \mu_n \frac{\partial (nF)}{\partial x} + D_n \frac{d^2 n}{dx^2} \tag{2.46}$$

Equations (2.44) and (2.46) are the continuity equations for holes and free electrons respectively and in deriving these equations all the three processes, generation and recombination, drift and diffusion, of the charge carriers have been considered.

2.15 SATURATION OF DRIFT VELOCITY

Current density due to drift in presence of an electric field E is given as

$$J = \sigma E \text{ (Ohm's law)} \tag{2.47}$$

where the conductivity

$$\sigma = en\,\mu_e$$

e is the electronic charge, n the electron density and μ_e is the electron mobility, given by

$$\left(\frac{ne^2\tau}{m} \equiv \sigma\right) \mu_e = \frac{e\tau}{m_e} \tag{2.48}$$

τ represents the mean time between scattering events and is called the mean free time.

In the above equation, it has been assumed that Ohm's law is valid in the carrier drift process, i.e., it has been assumed that the drift current is proportional to the electric field and that, the proportionality constant σ is not a function of the field. This assumption is valid over a wide range of E. However, large electric fields ($> 10^3$ V/cm) can cause the drift velocity and therefore the current

$$J = - en\, v_d \tag{2.49}$$

to exhibit a sublinear dependence on the electric field. This dependence of σ upon E is an example of a *hot carrier effect*, which implies that the carrier drift velocity v_d is comparable to the thermal velocity v_{th}.

In many cases, an upper limit is reached for the carrier drift velocity in a high field. This limit occurs near the mean thermal velocity ($\approx 10^7$ cm/s) and represents the point at which added energy imparted by the field is transferred to the lattice rather than increasing the carrier velocity. The result of this scattering limited velocity is a fairly constant current at high field. This behaviour is typical of Si, Ge and some other semiconductors (Fig. 2.14).

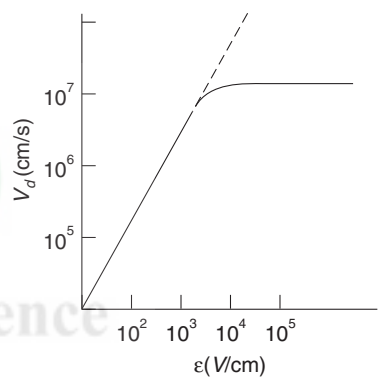


Fig. 2.14 *Saturation of electron drift velocity at high electric fields for Si*

However, there are other important effects in some other semiconductors, for example, there is a decrease in electron velocity at high fields for GaAs (and certain other materials) which results in negative conductivity.

2.16 FABRICATION OF P-N JUNCTIONS

Most semiconductors devices contain one or more than one junction between *P*-type and *N*-type material. These *P-N* junctions are fundamental to the performance of variety of functions such as rectification, amplification, switching and other operations in electronic circuits. In this chapter, we will restrict ourselves to a (single) *P-N* junction. Before investigating its electrical properties, we shall discuss how *P-N* junctions are made. Without going into the detailed manufacturing techniques, we will consider some of the more basic methods

of forming junctions. The type of junction we seek is a change from *N*-type to *P*-type material within a single crystal.* Therefore, we may change the dopant from donors to acceptors during the growth of the crystal. Alternatively, we may introduce impurities of one type into regions of a crystal which was grown with higher doping of the opposite type. In either case, there are no gross changes in the lattice structure of the crystal. There are basically following four methods employed for fabrication of *P-N* junctions.

1. Grown junctions
2. Alloyed junctions
3. Diffused junctions
4. Epitaxial grown or planar diffused diode

2.16.1 Grown Junction Diode

Diodes of this type are formed during the crystal pulling process. *P*- and *N*-type impurities can be alternately added to the molten semiconductor material in the crucible, which results in a *P-N* junction, as illustrated in Fig. 2.15 when crystal is pulled. After slicing, the larger area device can then be cut into a large number (say in thousands) of smaller area semiconductor diodes. Though such diodes, because of larger area, are capable of handling large currents but larger area also introduces more capacitive effects, which are undesirable. Such diodes are used for low frequencies.

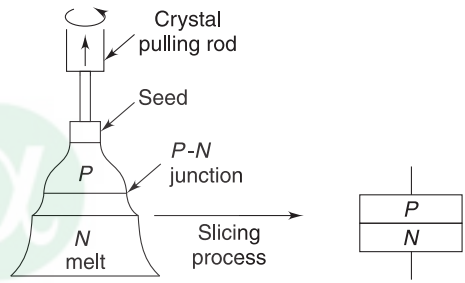


Fig. 2.15

2.16.2 Alloy Type or Fused Junction Diode

Such a diode is formed by first placing a *P*-type impurity (a tiny pellet of aluminium or some other *P*-type impurity) into the surface of an *N*-type crystal and heating the two until liquefaction occurs where the two materials meet. An alloy will result that on cooling will give a *P-N* junction at the boundary of the alloy substrate.

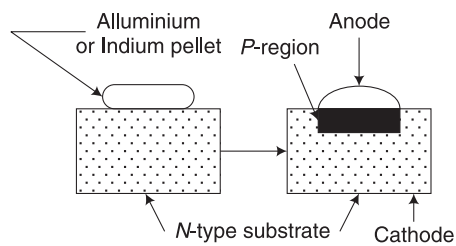


Fig. 2.16

* A crystal is a solid having a structure formed by a regular three dimensional periodic array of atoms in space. The repeat unit is called a unit cell which is characterized by a *basis* (i.e. an atom or a group of atoms).

Similarly, an *N*-type impurity may be placed into the surface of a *P*-type crystal and the two are heated until liquefaction occurs. The process is illustrated in Fig. 2.16.

Alloy type diodes have a high current rating and large PIV (peak inverse voltage) rating. The junction capacitance is also large, due to the large junction area.

2.16.3 Diffused Junction Diode

Diffusion is a process by which a heavy concentration of particles diffuse into a surrounding region of lower concentration. The main difference between the diffusion and alloy processes is the fact that liquefaction is not reached in the diffusion process. In the diffusion process heat is applied only to increase the activity of elements involved.

For formation of such diodes, either solid or gaseous diffusion process can be employed.

The process of solid diffusion starts with formation of layer of an acceptor impurity on an *N*-type substrate and heating the two until the impurity diffuses into the substrate to form the *P*-type layer, as illustrated in Fig. 2.17(a). A large *P-N* junction is divided into parts by cutting process. Metallic contacts are made for connecting anode and cathode leads.

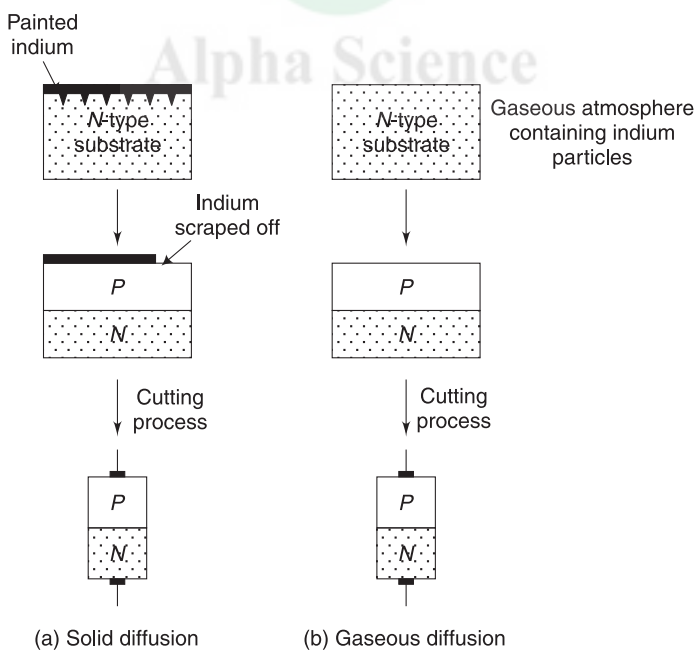


Fig. 2.17 *Diffusion process*

In the process of gaseous diffusion instead of layer formation of an acceptor impurity, an *N*-type substrate is placed in a gaseous atmosphere of acceptor impurities and then heated. The impurity diffuses into the substrate to form *P*-type layer on the *N*-type substrate. The process is illustrated in Fig. 2.17(b).

Though, the diffusion process requires more time than the alloy process but it is relatively inexpensive, and can be very accurately controlled. The diffusion technique leads itself to the simultaneous fabrication of many hundreds of diodes on one small disc of semiconductor material and is most commonly used in the manufacture of semiconductor diodes. This technique is also used in the production of transistors and ICs (integrated circuits).

2.16.4 Epitaxial Growth or Planar Diffused Diode

The term 'epitaxial' is derived from the Latin terms *epi* meaning 'upon' and *taxis* meaning 'arrangement'.

To construct an epitaxially grown diode, a very thin (single crystal) layer of semiconductor material (silicon or germanium) is grown on a heavily doped substrate (base) of the same material. This complete structure then forms the *N*-region on which *P*-region is diffused. SiO_2 layer is thermally grown on the top surface, photo-etched and then aluminium contact is made to the *P*-region. A metallic layer at the bottom of the substrate forms the cathode to which lead is attached. (Fig. 2.18)

This process is usually employed in the fabrication of IC chips.

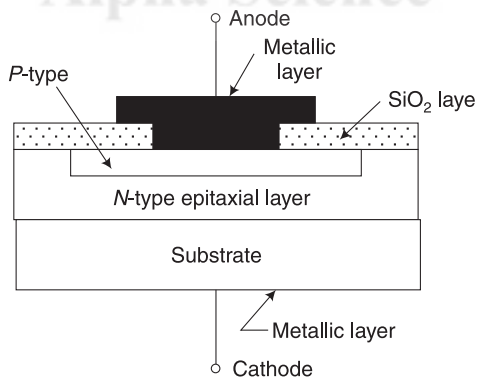


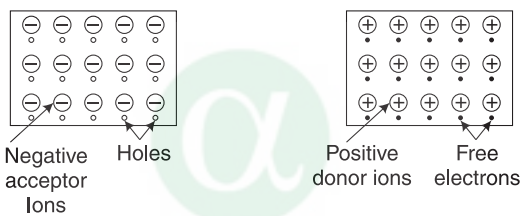
Fig. 2.18 *Epitaxially grown or planar diffused diode*

2.17 BASIC STRUCTURE OF *P-N* JUNCTION

Most semiconductor devices employ one or more *P-N* junctions. The *P-N* junction is the control element for the performance of all semiconductor devices such as rectifiers, amplifiers, switching devices, linear and digital integrated circuits. The

P-N junction is produced by placing a layer of *P*-type semiconductor next to the layer of *N*-type semiconductor. The interface separating the *N* and *P*-regions is referred to as the *metallurgical junction*.

Figure 2.19 represents two blocks of semiconductor material, one *P*-type, and the other *N*-type. The *P*-type semiconductor block has mobile holes (shown by small circles) and the same number of fixed negative acceptor ions (shown by encircled minus sign). Similarly, the *N*-type semiconductor block has mobile or free electrons (shown by dots) and the same number of fixed donor positive ions. Normally, the holes, which are the majority charge carriers in *P*-type of material, are uniformly distributed throughout the volume of that material. Similarly, the electrons, which are the majority charge carriers in *N*-type of material, are uniformly distributed throughout the volume of that material. Each region is electrically neutral because each of them carries equal positive and negative charges.



(a) *P*-type semiconductor material (b) *N*-type semiconductor material

Fig. 2.19

On the formation of *P-N* junction some of the holes from *P*-type material tend to diffuse across the boundary into *N*-type material and some of the free electrons similarly diffuse into the *P*-type material, as illustrated in Fig. 2.20. This happens due to density gradient (as concentration of holes is higher on *P*-side than on *N*-side and concentration of electrons is higher on *N*-side than that on *P*-side). This process is known as *diffusion*.

The doping profile of an ideal uniformly doped *P-N* junction is depicted in Fig. 2.21.

As a result of the displacement of the charges, an electric field appears across the junction. Equilibrium is established when the field becomes large enough to restrain the process of diffusion. The general shape of the charge distribution may be as illustrated in Fig. 2.20(b). The electric charges are confined to the neighbourhood of the junction, and consists of immobile ions.

We see that the free electrons crossing the junction create negative ions on the *P*-side by giving some atoms one more electron than their total number of

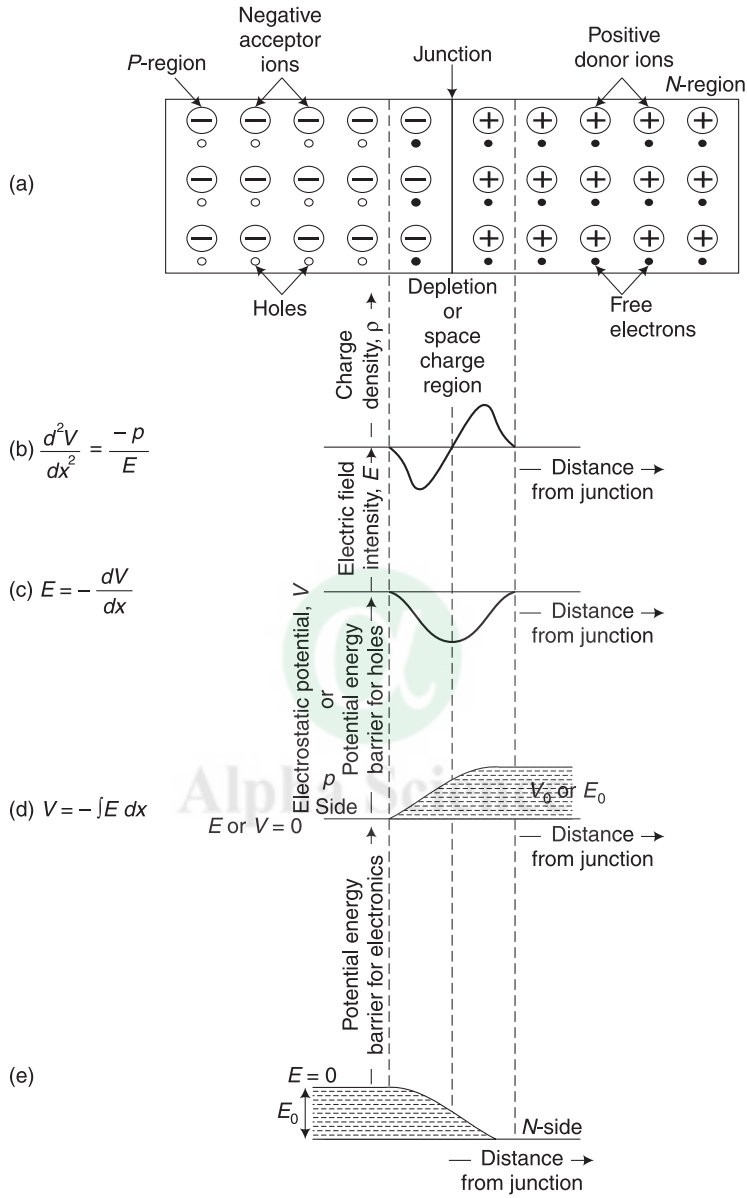


Fig. 2.20 *A schematic diagram of a P-N junction, including charge density, electric field intensity, and potential-energy barriers at the junction (not to scale)*

protons. The electrons also leave positive ions (atoms with one fewer electron than the number of protons) behind them on the *N*-side. As negative ions are created on the *P*-side of the junction, the *P*-side acquires a negative potential, as

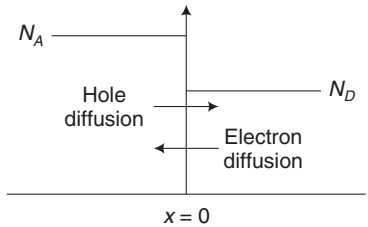


Fig. 2.21 *Doping profile of an ideal uniformly doped P-N junction*

illustrated in Fig. 2.20(e). Similarly, the positive ions are created on the *N*-side and the *N*-side acquires a positive potential as illustrated in Fig. 2.20(d). The negative potential on the *P*-side prevents the migration of any more electrons from the *N*-type material to the *P*-type material. Similarly, the positive potential on the *N*-side prevents any further migration of holes across the boundary. Thus, the initial diffusion of charge carriers creates a barrier potential at the junction.

The region around the junction is completely ionized. As a result, there are no free electrons on the *N*-side, nor holes on the *P*-side. Since the region around the junction is depleted of mobile charges it is called the *depletion region*, the *space-charge region*, or the *transition region*. The thickness of the depletion region (or layer) is of order of 1 micron (10^{-6} m).

The electric field intensity in the neighbourhood of the junction is illustrated in Fig. 2.20(c). The electrostatic potential variation in the depletion region is shown in Fig. 2.20(d), and is the negative integral of the function *E* of Fig. 2.20(c). The variation constitutes a potential-energy barrier against the further diffusion of holes across the barrier. The form of potential-energy barrier against the flow of electrons from the *N*-side across the junction is shown in Fig. 2.20(e). It is similar to that shown in Fig. 2.20(d), except that it is inverted, since the charge on electron is negative.

2.18 BUILT IN POTENTIAL BARRIER

We consider here the case of a step junction in thermal equilibrium where no currents exist when no external bias (voltage) is applied. Then, the Fermi energy level is constant throughout the specimen. The valence and conduction bands energies bend while going across the space charge region, because the relative positions of these bands with respect to the Fermi energy changes between *P*- and *N*-regions. (Fig. 2.22)

We have seen how a potential barrier develops at the metallurgical junction in a *P-N* diode in article 2.17. This potential barrier is known as the *built in potential barrier*. Electrons in the conduction band of the *N*-region face this

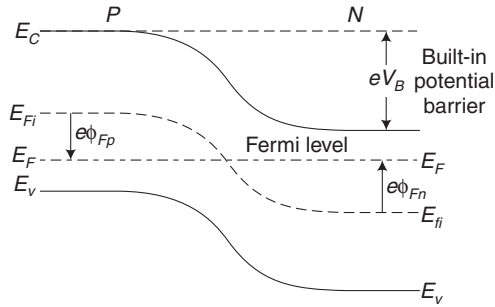


Fig. 2.22 *Energy band diagram of a P-N junction in thermal equilibrium with zero applied bias*

barrier potential V_B when moving into the conduction band of the *P*-region. The built in potential barrier maintains equilibrium between majority carrier electrons in the *N*-region and minority carrier electrons in the *P*-region. It also maintains equilibrium between majority carries holes in the *P*-region and minority carrier holes in the *N*-region. As this condition pertains to equilibrium, no current is caused to flow. The intrinsic Fermi-level is equidistant from the conduction band edge through the junction.

In order to derive the expression for V_B, we first consider *N*- and *P*-semiconductors separately. Figure 2.23(a) shows the position of the Fermi level and (b) the density of carriers in an *N*-type semiconductor.

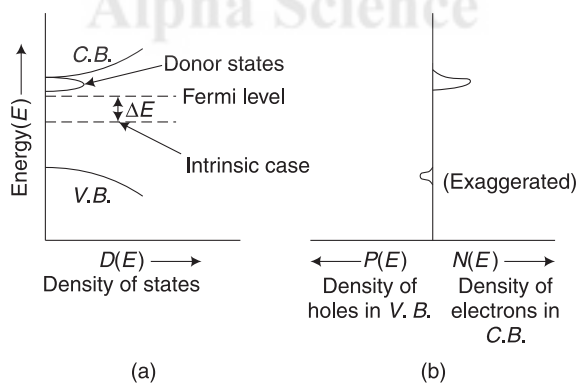


Fig. 2.23 *Fermi level and carrier density in N-type semiconductor (a) shows the position of Fermi level and (b) carrier density in P-type semiconductor*

Density of states A function that gives the number of available electron states per unit volume in an energy interval between E and $E + dE$ is designated by $D(E)dE$, where $D(E)$ is known as the density of states function and represents the number of states per unit volume per electron volt.

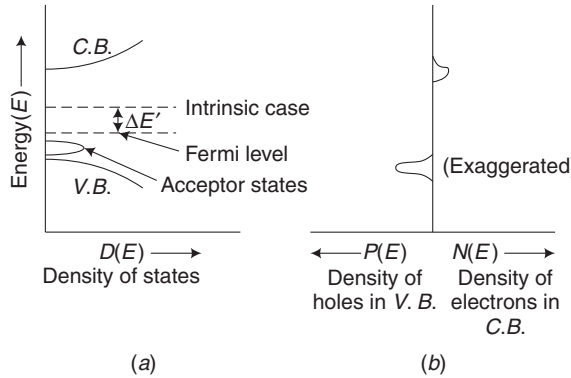


Fig. 2.24 *Fermi level and carrier concentration in P-type semiconductor*

The Fermi-level is at an energy ΔE (above the centre of the gap) near the conduction band. The total number of electrons in the conduction band is given by

$$n = n_i \exp\left(\frac{\Delta E}{kT}\right) \tag{2.50}$$

where n_i is the carrier density for the intrinsic case.

The Fermi-level is at an energy $\Delta E'$ below the centre of the gap. The total number of holes in the valence band is given by

$$p = n_i \exp\frac{\Delta E'}{kT} \tag{2.51}$$

When the junction is formed, due to the barrier voltage, conduction band on P-side is higher by an amount eV_B from the conduction band on N-side i.e., if E_p and E_n are the energy levels of the bottom of the conduction bands on P-side and N-side respectively, then

$$E_p - E_n = eV_B \tag{2.52}$$

Further, the Fermi levels on the two sides coincide as shown in Fig. 2.25.

$$n_n = N_d = n_i \exp\left(\frac{\Delta E}{kT}\right)$$

and

$$n_p = \frac{n_i^2}{N_a} = n_i \exp\left(-\frac{\Delta E'}{kT}\right) \tag{2.53}$$

where N_d and N_a are the donor and acceptor ion concentrations on N- and P-sides respectively.

$$\Delta E + \Delta E' = eV_B \tag{2.54}$$

Thus,

$$\frac{n_n}{n_p} = \exp\left(\frac{\Delta E + \Delta E'}{kT}\right) = \exp\left(\frac{eV_B}{kT}\right) \tag{2.55}$$

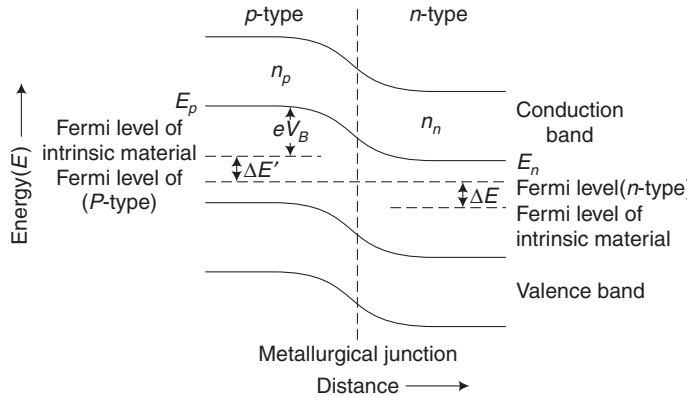


Fig. 2.25 *Energy band diagram for P-N function*

or

$$V_B = \frac{kT}{e} \log_e \left(\frac{n_n}{n_p}\right) = \frac{kT}{e} \log_e \left(\frac{N_d\, N_a}{n_i^2}\right) \tag{2.56}$$

The quantity (kT/e) is called *thermal voltage*.

Example 2.1 *A sample of Ge is made p-type by adding acceptor atoms at a rate of one atom per 4×10^8 Ge atoms. $n_i = 2.5 \times 10^{19}/m^3$ at 300 K and all the acceptor atoms are ionized at 300 K. The density of germanium is 4.4×10^{28} atoms/m^3. Compare the density of electrons with intrinsic charge carriers.*

Solution For 4×10^8 Ge atoms, one acceptor atom is added. For 4.4×10^{28} Ge atoms, the acceptor atoms added will be 1.1×10^{20}.

Now

$$n_i^2 = np, \quad n = n_p \quad \text{and} \quad p = N_a$$

$$\therefore \quad n_p = \frac{n_i^2}{N_a} = \frac{(2.5 \times 10^{19})^2}{1.1 \times 10^{20}} = 5.7 \times 10^{18}$$

$$\Rightarrow \quad \frac{n_p}{n_i} = \frac{5.7 \times 10^{18}}{2.5 \times 10^{19}} = 0.22$$

Figure 2.27 shows the plots of (b) space charge, (c) electric field and (d) barrier potential at an abrupt *P-N* junction.

2.19 TRANSITION (OR SPACE-CHARGE) CAPACITANCE

When a *P-N* junction is reverse biased, the depletion region acts like an insulator or dielectric material while the *P*- and *N*-type regions on either side have a low resistance and act as the plates. The *P-N* junction may be considered a parallel plate capacitor. The junction capacitance is termed as space charge capacitance or *transition capacitance* or *depletion region capacitance* and is denoted by C_T.

As mentioned earlier, a reverse bias causes majority carriers to move away from the junction, thereby uncovering more immobile charges. So the thickness W of the depletion layer increases with the increase in reverse bias voltage. This increase in uncovered charge with applied voltage may be considered a capacitive effect. The incremental capacitance C_T may be defined as

$$C_T = \left|\frac{dQ}{dV}\right| \tag{2.57}$$

where dQ is the increase in charge due to increase in voltage, dN.

But $$i = \frac{dQ}{dt} = C_T \frac{dV}{dt} \tag{2.58}$$

Since the depletion region increases with the increase in reverse-bias potential, the resulting transition capacitance decreased as shown in Fig. 2.26.

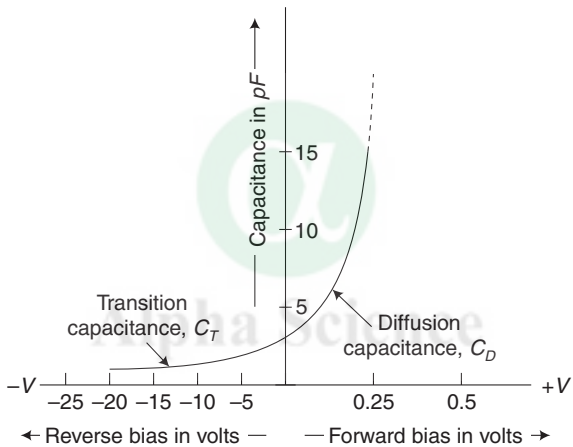


Fig. 2.26

The fact that the capacitance depends on the applied reverse bias, but application in a number of electronic systems.

Step Graded Junction A junction is said to be step graded if there is an abrupt change from acceptor ion concentration on the *P*-side to donor ion concentration on the *N*-side. Such a junction gets formed in alloyed junction or fused junction diode. Usually the acceptor density N_A and the donor density N_D are kept unequal. Figure 2.27 shows the charge density as a function of distance from an alloy junction in which acceptor impurity density is assumed to be much larger than the donor concentration. Since the net charge is zero,

$$eN_AW_p = eN_DW_n \tag{2.59}$$

When N_A is much larger than N_D, the $W_p << W_n$ and $W_n \approx W$.

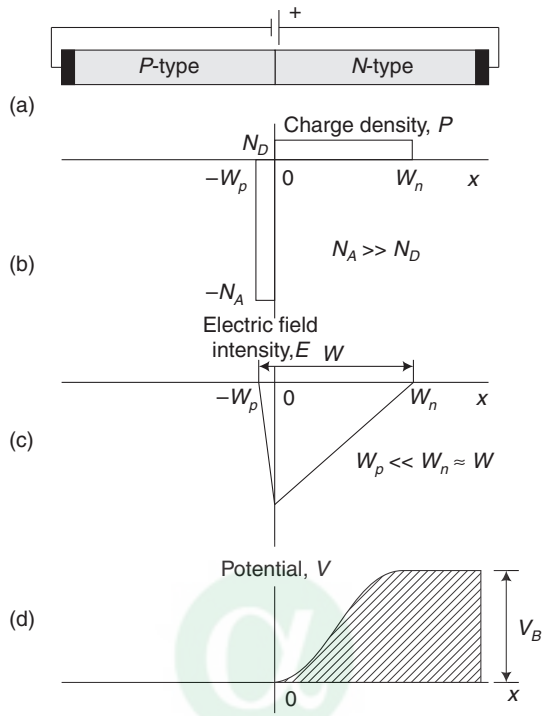


Fig. 2.27 *Charge density, field intensity and potential variation with distance x*

The relationship between potential and charge density is given by the Poisson's equation,

$$\frac{d^2V}{dx^2} = \frac{-eN_D}{\varepsilon} \tag{2.60}$$

where ε is the permittivity of the semiconductor and is given as $\varepsilon = \varepsilon_r\varepsilon_0$ where ε_r is the relative permittivity of the dielectric and ε_0 is the permittivity of free space.

The electric lines of flux start on the positive donor ions and terminate on the negative acceptor ions. So there are no flux lines to the right of the boundary $x = W_n$ in Fig. 2.27 and field intensity $\varepsilon = \frac{dV}{dx} = 0$

at $$x = W_n \approx W.$$

Integrating Eq. (2.60) subject to these boundary conditions, we have

$$\frac{dV}{dx} = \frac{-eN_D}{\varepsilon}(x - W) = -\varepsilon \tag{2.61}$$

Neglecting the small voltage drop across W_p, we may arbitrarily choose $V = 0$ at $x = 0$.

Integrating Eq. (2.61) subject to these conditions

$$V = \frac{-eN_D}{2\varepsilon}(x^2 - 2Wx) \tag{2.62}$$

The linear variation in field intensity and the quadratic dependence of potential upon distance are plotted in Figs. 2.27(c) and (d).

At $x = W$, $V = V_B$, the barrier height. Thus

$$V_B = \frac{eN_D W^2}{2\varepsilon} \tag{2.63}$$

Here $V_B = V_0 - V$ where V is the negative number for an applied reverse bias and V_0 is the contact potential.

Hence, the width of depletion layer increases with applied reverse bias and varies as $\sqrt{V_B}$.

The total charge of N-type material with area of junction A is given by

$$Q = eN_D WA$$

The transition capacitance C_T, given by Eq. (2.57), is

$$C_T = \left|\frac{dQ}{dV}\right| = eN_D A \left|\frac{dW}{dV_B}\right|$$

From Eq. (2.63), $\left|\frac{dW}{dV_B}\right| = \frac{\varepsilon}{sN_D W}$, and therefore,

$$C_T = \frac{\varepsilon A}{W} \tag{2.64}$$

where ε is the absolute permittivity of the semiconductor medium, A is the cross-sectional area of the junction and W is the width of the depletion layer and is given as

$$W^2 = \left[\frac{2\varepsilon V_B}{e}\right]\left[\frac{1}{N_A} + \frac{1}{N_D}\right] \tag{2.65}$$

or

$$W = \sqrt{\frac{2\varepsilon V_B}{eN_D}} \qquad \text{when } N_A >> N_D \tag{2.66}$$

So

$$C_T = \varepsilon A \sqrt{\frac{eN_D}{2\varepsilon V_B}} = A\sqrt{\frac{N_D}{V_B}} \cdot \sqrt{\frac{e\varepsilon}{2}} \tag{2.67}$$

Thus in a step graded junction, C_T is inversely proportional to the square root of the junction voltage V_B where V_B is given as

$$V_B = V_0 - V_R \tag{2.68}$$

where V_R is a negative number indicating the applied reverse bias and V_0 is the contact potential.

2.20 A *PN*-JUNCTION UNDER FORWARD AND REVERSE BIAS (DIODE CHARACTERISTICS)

When a voltage is applied to a *P*-*N* semiconductor junction, the voltage current characteristics is found to be non-linear. If the *P*-type material is made positive with respect to the *N*-type, the junction is said to be forward biased. In this case the device resistance is low and a large (forward) current flows readily across the *P*-*N* junction. When the *N*-type is made positive with respect to *P*-type, the junction is said to be reverse biased and its resistance is high. (Fig. 2.28)

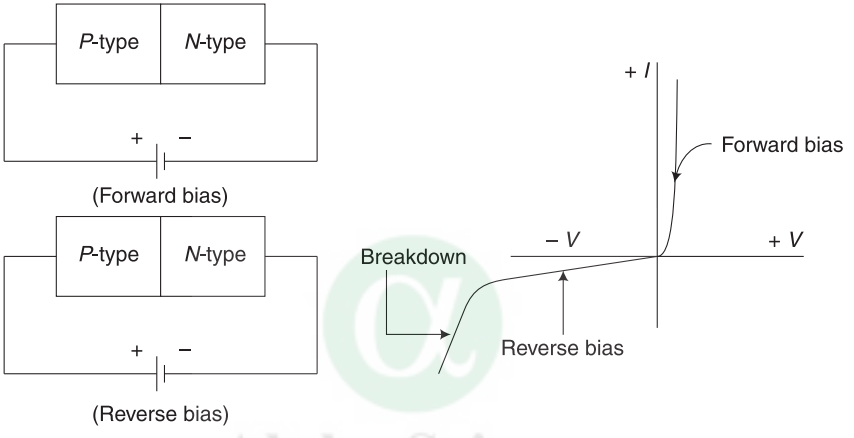


Fig. 2.28 *(a) A P-N-junction forward and reverse biased, (b) its I-V characteristics*

In case of forward bias, many electrons from the negative terminal of battery enter the *N*-side. The life story of a single electron is as under:

1. After leaving the negative terminal, it enters the right end of the crystal.
2. It travels through the *N*-region as a free electron.
3. Near the junction it recombines and becomes a valence electron.
4. It travels through the *P*-region as a valence electron.
5. After leaving the left end of the crystal, it flows into the positive source terminal.

Thus, forward bias produces a large forward current.

In case of reverse bias, conduction band electrons and valence band holes move away from the junction, the depletion layer gets wider, until its potential difference equals the supply voltage. Under this condition, free electrons and holes stop moving. However, because of thermal generation, a small current due to minority carriers exists in the circuit. This reverse current caused by minority carriers is called the saturation current (because we cannot get more of this current by increasing the reverse voltage unless we increase the temperature).

2.21 THE ENERGY BAND DIAGRAMS

When only a small current is flowing in a typical *P-N*-junction, there is negligible voltage drop in the bulk material itself, and all the applied voltage is dropped across the junction. The presence of an applied voltage across the junction adds to (for reverse bias) or subtracts from (for forward bias) the barrier voltage V_B.

Under the forward bias V, when an electron passes across the junction from the *N*- to the *P*-type material, it is moving in the same direction as the force due to the applied electric field. As a result, the potential energy in the *P*-type material is less than what would have been in an unbiased junction by an amount V electron volts. Thus, the conduction and valence bands on the *P*-side are lowered by this value. (Figs. 2.29 and 2.30)

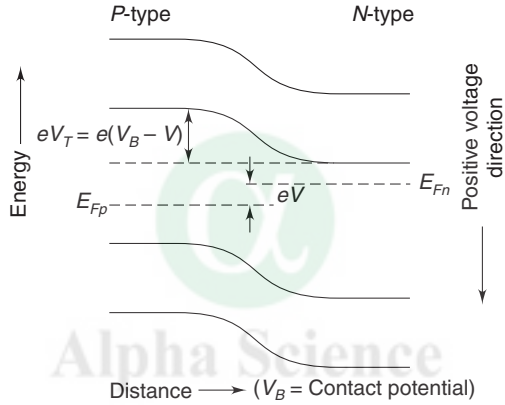


Fig. 2.29 *Energy-band diagram for a forward biased P-N-junction*

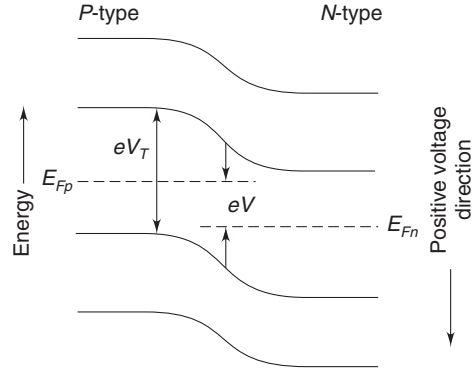


Fig. 2.30 *Energy band diagram for a reverse biased P-N-junction*

2.22 CURRENT FLOW MECHANISM: THE RECTIFIER EQUATION FOR A FORWARD BIASED DIODE

Consider Fig. 2.31, which shows the energy bands of a *P-N*-junction. In the diagram, a forward bias (V volts)has been applied. As a result, the Fermi level on n side E_{fn} is higher than that on the P-side by the value eV.

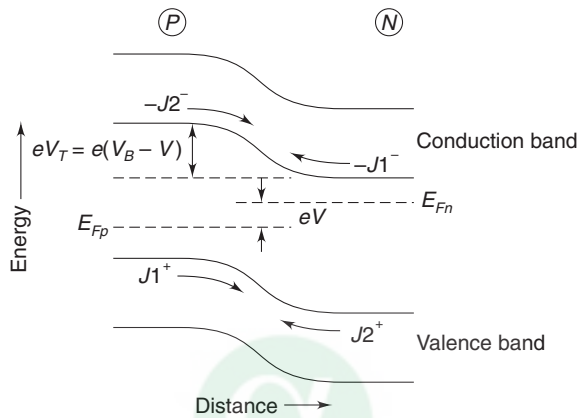


Fig. 2.31 *Energy band diagram of a P-N-junction, forward biased by V-volts. Superscript + refers to hole current and – to electron current*

There are four possible components of current flowing across the junction (Fig. 2.31). (In the figure, suffixes 1 and 2 represent the majority and minority carriers respectively). In the valence band, holes can flow to the right or to the left (indicated by J_1^+ and J_2^+). In the conduction band, electrons can flow to the right or to the left. Due to the negative charge, a flow of electrons to the left produces a conventional current to the right therefore arrows in conduction band indicate the direction of flow of electrons and a negative signs is placed before the current density.

Thus current density J_1^- is flowing to the right and current density J_2^- is flowing to the left. This mean that the current densities J_1^- and J_1^+ add together in the right hand direction and current densities J_2^- and J_2^+ add together in the left hand direction. The net current flowing to the right is

$$J = (J_1^- + J_1^+) - (J_2^- + J_2^+) \tag{2.69}$$

Here J_2^- is the minority carrier electron current drifting from P-side to N-side. It is independent of the applied voltage V (provided $V < V_B$). It is known as saturation current density. Similarly J_2^+ is minority carrier hole current drifting from N-side to P-side (also saturation current).

In the conduction band, J_1^- is the majority carrier electron current diffusing from N-side to P-side. Provided the current is not excessive, the only restriction

to flow of this current is the barrier potential $V_T = V_B - V$. For potential of height V_T, the number of electrons which are available source of current is proportional to exp $(- eV_T/kT)$.

Thus

$$J_1^- = C_1 \exp\left(-\frac{eV_T}{kT}\right) \tag{2.70}$$

where C_1 is a constant to be determined. Since holes carry a positive sign, holes flowing downward in the figure are effectively flowing up the barrier. The majority carrier hole current diffusing from *P*-side to *N*-side is given by

$$J_1^+ = C_2 \exp\left(-\frac{eV_T}{kT}\right) \tag{2.71}$$

where C_2 is another constant.

Since there is no net current flow across the junction when applied $V = 0$ (or $V_T = V_B$) then,

$$J_1^- - J_2^- = 0 \quad \text{(net electron current in conduction band).}$$

and

$$J_1^+ - J_2^+ = 0 \quad \text{(net hole current in valence band).} \tag{2.72}$$

(because the current flow in one band is independent of flow in the other).

$$\Rightarrow \qquad C_1 \exp\left(-\frac{eV_B}{kT}\right) - J_2^- = 0$$

or

$$C_1 = J_2^- \exp\left(\frac{eV_B}{kT}\right) \tag{2.73}$$

Similarly,

$$C_2 = J_2^+ \exp\left(\frac{eV_B}{kT}\right) \tag{2.74}$$

$$\therefore \qquad J_1^- = J_2^- \exp\left[\frac{e(V_B - V_T)}{kT}\right] \tag{2.75}$$

and

$$J_1^+ = J_2^+ \exp\left[\frac{e(V_B - V_T)}{kT}\right] \tag{2.76}$$

$$\Rightarrow \qquad J = (J_1^+ + J_1^-) - (J_2^+ + J_2^-)$$

or

$$J = (J_2^- + J_2^+)\left[\exp\left(\frac{eV}{kT}\right) - 1\right] \tag{2.77}$$

or

$$J = J_2\left\{\exp\left(\frac{eV}{kT}\right) - 1\right\} \tag{2.78}$$

where $J_2 = J_2^- + J_2^+$ is the sum of two saturation current densities. This equation is known as the rectifier equation and is plotted in Fig. 2.32.

If forward bias $V >> \frac{4kT}{e}$ (kT/e = 0.0259 V at room temperature), then

$$\exp\left(\frac{eV}{kT}\right) >> 1$$

and

$$J \approx J_2 \exp\left(\frac{eV}{kT}\right)$$

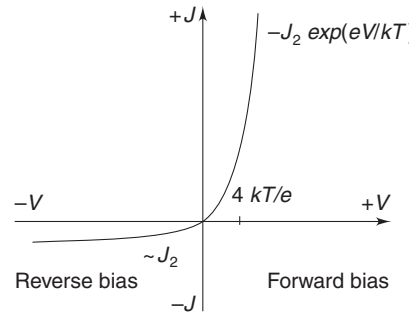


Fig. 2.32 *V-J characteristics curve of a P-N-diode*

When a reverse bias is applied such that $-V \geq 4kT/e$, then $\exp(eV/kT) << 1$ (exponential term approaches zero), and

$$J \approx -J_2, \tag{2.79}$$

that is, current at reverse bias remains constant at the saturation current.

2.23 APPLICATION OF THE CONTINUITY EQUATIONS TO THE ABRUPT *P-N* JUNCTION WITH CONSTANT CURRENT

Let us apply the continuity equation for holes to the case of a forward biased *P-N* junction where the *P*-region is positive with respect to the *N*-region by E_f volts. Suppose the injected hole density into the *N*-region at a distance x from the junction is p_n. We shall investigate the variation of the injected hole density with distance x from the junction. We can make the following simplifying assumptions.

(a) The current flow due to the injected holes in the *N*-region is simply due to the diffusion of holes. The hole component of the current due to the drift of the hole on account of the applied field is absent, because the voltage drop across the bulk material is practically zero and almost the entire applied voltage drops across the junction.

Thus
$$\frac{\partial(pF)}{dx} = 0 \tag{2.80}$$

(b) The continuity equation may easily be solved in the case of a *P-N* junction with a constant current flowing across the junction. If the applied forward bias voltage is constant (d.c.) the total current density of a *P-N* junction diode is constant and the electrons and hole densities in the two regions are also constant with respect to time. Thus, for a *P-N* junction operating with a constant current it may be assumed that

$$\frac{\partial p}{dt} = \frac{\partial n}{dt} = 0 \tag{2.81}$$

Under the above two restricted conditions, the continuity equation becomes

$$\frac{\partial^2 p}{dx^2} = \frac{p - p_o}{D_p T_p} = \frac{p - p_o}{L_p^2} \quad \text{for holes} \tag{2.82}$$

where $(D_p T_p)^{1/2} = L_p$ = diffusion length of holes in the *N*-region.

and $$\frac{d^2 n}{dx^2} = \frac{n - n_o}{D_n T_n} = \frac{n - n_o}{L_p^2} \text{ for the free electrons} \tag{2.83}$$

where $(D_n T_n)^{1/2} = L_n$ = diffusion length of free electrons in the *P*-region.

Let us consider the solution of the continuity equation for holes. Eq. (2.82) in *N*-type material where holes are the minority carriers and the additional subscript "*n*" is put with *p* to indicate the hole density in the *N*-type material so that p_n is the density of holes at a distance *x* from the edge of the depletion layer and p_{no} is the equilibrium hole density in the *N*-type material. Thus, the continuity equation for holes in an *N*-type material is

$$\frac{d^2 p_n}{dx^2} = \frac{p_n - p_{no}}{L_p^2} \tag{2.84}$$

The solution of the continuity equation for holes (Eq. 2.84) is of the form

$$p_n - p_{no} = A \exp\left(\frac{-x}{L_p}\right) + B \exp\left(\frac{x}{L_p}\right) \tag{2.85}$$

where *A* and *B* are arbitrary constants and are to be determined by the boundary conditions. At a larger distance from the junction the hole density should approach the equilibrium value p_{no} in the bulk material. By a large distance we mean that the distance should be at least five times larger than L_p. Hence when *x* tends to be infinite,

$$p_n = p_{no}$$

and so $$0 = B \exp\left(\frac{x}{L_p}\right)$$

giving $$B = 0 \tag{2.86}$$

The solution is, therefore,

$$p_n - p_{no} = A \exp\left(\frac{-x}{L_p}\right) \tag{2.87}$$

The value of the constant A is to be determined by assuming a second boundary condition. If we assume that p_n^* is the density of holes at the edge of the depletion layers near the N-region, the density p_n^* is far greater than the equilibrium density p_{no} in the case of a forward P-N junction where hole injection takes place from the P-region to the N-region even when the forward biased voltage is greater than 0-1 volt. In the reverse biased case where the reverse biased voltage is greater than 0.1 volt, the value of p_n^* is practically zero.

At $x = X_2$, $\qquad p_n = p_n^*$.

On substitution of the above values in Eq. (2.87)

$$p_n^* - p_{no} = A \exp\left(\frac{-X_2}{L_p}\right)$$

or
$$A = (p_n^* - p_{no}) \exp \frac{X_2}{L_p} \tag{2.88}$$

Thus Eq. (2.87) becomes

$$p_n^* - p_{no} = (p_n^* - p_{no}) \exp -\left(\frac{x - X_2}{L_D}\right) \tag{2.89}$$

where $(p_n^* - p_{no})$ is the value of the excess holes injected from the P-region to the N-region at the edge of the depletion layer near the N-region where $x = X_2$ (Fig. 2.33) and p_n – is its value at a distance x from the depletion layer in the N-region. When the distance $(x - X_2)$ is equal to the diffusion length L_p, the density of injected holes becomes 37% (1/ε) of its value at the edge of the depletion layer due to recombination. The square of the diffusion length L_p of holes in the N-region is equal to the square root of the product of the diffusion constant D_p of holes and the life time of holes in the N-region (T_p).

Using a similar argument, assuming the boundary condition that

at $x = -X_1$, $\qquad n_p = n_p^*$

and when $x = -\infty$ $\qquad n_p = n_{po}$

and solving Eq. (2.83) we can show that the excess of electron density in the P-region due to injection from the N-region is

$$n_p - n_{po} = (n_p^* - n_{po}) \exp\left(\frac{X_1 + x}{L_n}\right) \tag{2.90}$$

where n_p^* is the electron density at the edge of the depletion layer near the P-region, n_{po} is the equilibrium electron density in the P-region so that $(n_p^* - n_{po})$ is the number of excess electrons injected from the N-region at the edge of the depletion layer near the P-region. $L_n = (D_n\, T_n)^{1/2}$ is the diffusion length of electrons in the P-region so that at a distance L_n from the edge of the depletion layer near the P-region the injected electron density becomes 37% of its value at the edge of the depletion layer ($z = -X_1$).

2.24 CHARGE CARRIER DISTRIBUTION AND CURRENT DENSITY IN A FORWARD BIASED *P-N* JUNCTION

If p_{po} is the equilibrium density of holes in the *P*-region, p_{no} is the equilibrium density of holes in the *N*-region of an unbiased *P-N* junction and V_B is the barrier potential developed across the junction, these three quantities are related by equation (as can be shown from Eq. (2.91)).

$$p_{no} = p_{po} \exp - \left(\frac{eV_B}{kT}\right) \tag{2.91}$$

where k is the Boltzmann constant and T is the absolute temperature.

If the *P-N* junction is forward biased by E_f volts, the net potential barrier will decrease and become equal to $(V_B - E_f)$. Due to this reduced potential barrier a considerable number of electrons will be injected into the *P*-region, while a large number of holes will be injected into the *N*-region. Naturally the density of the holes at the edge of the depletion layer near the *N*-region is so adjusted that the modified potential barrier $V_B - E_f$, the hole density in the *P*-region, p_p, and the increased hole density at the edge of the depletion layer in the *N*-region due to injection, p_n* are related as

Thus,
$$p_n^* = p_p \exp\left[\frac{-e\,(V_B - E_f)}{kT}\right]$$

$$= p_p \exp\left(\frac{-eV_B}{kT}\right) \exp\left(\frac{eE_f}{kT}\right)$$

$$= p_{no} \exp\left(\frac{eE_f}{kT}\right) \tag{2.92}$$

On substitution of the value of p_n* from Eq. (2.81) in Eq. (2.89) of Section 2.23, we get

$$p_n - p_{no} = \left[p_{no} \exp\left(\frac{eE_f}{kT}\right) - p_{no}\right] \exp - \frac{(x - X_2)}{L_p}$$

$$= \left[p_{no} \exp\left(\frac{eE_f}{kT}\right) - 1\right] \exp - \frac{(x - X_2)}{L_p} \tag{2.93}$$

Considering the equilibrium density of electrons in the *N*-region and the *P*-region and the potential barrier developed, we can show in a similar way that,

$$n_p^* = n_n \exp - \frac{e\,(V_B - E_f)}{kT}$$

$$= n_{po} \exp - \frac{eE_f}{kT} \tag{2.94}$$

where n_n is the density of free electrons in the *N*-region, n_p* is the density of electrons due to injection in the *P*-region near the depletion layer and E_f is the forward bias voltage. On substitution of the value of n_p* from Eq. (2.94) in Eq. (2.90) of Section 2.23, we get

$$n_p - n_{po} = n_{po}\left[\exp\frac{eE_f}{kT} - 1\right]\exp\frac{(X_1 + x)}{L_n} \tag{2.95}$$

The distribution of free electrons and hole densities along the length of the *P-N* junction biased in the forward direction is shown in Fig. 2.33. The distribution curves are plotted with the help of Eq. (2.93) and Eq. (2.95). At this stage it is worthwhile to have some idea of the values of the diffusion lengths. For example, in germanium of 00 microsecond carrier life time, the diffusion length for electrons is about 0.1 cm, while for holes it is about 0.07 cm. The corresponding values for silicon are even smaller due to the smaller values of diffusion coefficients and typically low carrier lifetimes in this material.

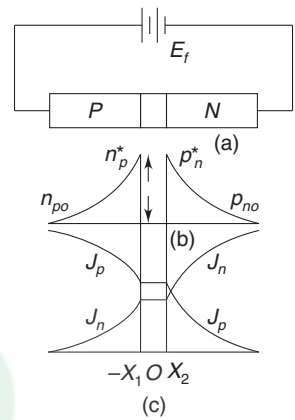


Fig. 2.33 *(a) A forward biased P-N junction with (b) distribution of minority carriers and (c) distribution of the current densities along the length of the P-N junction*

It is already mentioned that almost the entire current flow in a forward biased *P-N* junction is due to the diffusion of free electrons from the *N*-region to the *P*-region and the diffusion of holes from the *P*-region to the *N*-region. The current density J_{2p}, due to the diffusion of holes is given by

$$J_{2D} = -eD_p\frac{d\,(p_n - p_{no})}{dx}$$

$$= \frac{eD_p}{L_p}p_{no}\left(\exp\frac{eE_f}{kT} - 1\right)\exp -\frac{(x - X_2)}{L_D} \tag{2.96}$$

This component of current density at the edge of the depletion layer near the *N*-region where x is equal to X_2, is given by

$$J_{2p}\big|_{x = X_2} = J_{2po} = \frac{eD_p\,p_{no}}{L_p}\left(\exp\frac{eE_f}{kT} - 1\right) \tag{2.97}$$

This component of electric current, J_{2po}, flows from the left to the right (Fig. 2.33). The current density J_{2n}, due to the diffusion of free electrons from the *N*-region to the *P*-region, is given by

$$J_{2n} = eD_n \frac{d\,(n_p - n_{po})}{dx}$$

$$= -\frac{eD_n}{L_n} n_{po} \left(\exp \frac{eE_f}{kT} - 1\right) \exp\,(X_1 + x)/L_n \qquad (2.98)$$

This current density at the edge of the depletion layer near the *P*-region, where x is equal to $-X_1$, may be given by

$$J_{2nx} = -X_1 = J_{2no} = \frac{eD_n\, n_{po}}{L_n} \left(\exp \frac{eE_t}{kT} - 1\right) \qquad (2.99)$$

The current density J_{2no} owes its origin to the diffusion of free electrons which flow from the right to the left and thus the conventional current J_{2no} flows from left to the right (Fig. 2.33). Thus the total current flow due to diffusion of both holes and free electrons at the junction J_{2o} can be shown to be

$$J_{2o} = J_{2po} + J_{2no} = \left(\frac{eD_p\, p_{no}}{L_p} + \frac{eD_n\, n_{po}}{L_n}\right) \left(\exp \frac{eE_f}{kT} - 1\right) \qquad (2.100)$$

This is the junction current density in the forward biased *P-N* junction. The value of $\frac{kT}{e}$ at the room temperature (300 K) is equal to $\frac{1}{40}$ volt and even when the applied forward biased voltage is greater than 0.1 volt, $\exp \frac{eE_f}{kT}$ is much greater than 1. Therefore, in this condition the forward biased current density at the junction, J_{2o}, is given by

$$J_{2o} = \left[\frac{eD_p\, p_{no}}{L_p} + \frac{eD_n\, n_{po}}{L_n}\right] \exp \left(\frac{eE_f}{kT}\right) \qquad (2.101)$$

The ratio of electron and hole components of the current densities in the forward biased *P-N* junction is given by

$$\frac{J_{2no}}{J_{2po}} = \frac{D_n\, L_p\, n_{po}}{D_p\, L_n\, p_{no}} \qquad (2.102)$$

Since, according to the Einstein relation, $\frac{D_n}{D_p} = \frac{\mu_n}{\mu_p}$, where μ_n and μ_p are the electron mobilities.

Suppose n_{po} and p_{no} are the equilibrium electron and hole densities in the *P*-region and the *N*-region respectively, while n_{no} and p_{po} are the corresponding

equilibrium electron and hole densities in the *N*-region and *P*-region respectively.

Then $$n_{po}\, p_{po} = n_{no}\, p_{no} = n_i^2 \tag{2.103}$$

where n_i is the equilibrium density of electron-hole pairs in the intrinsic semiconductor. With the help of Eq. (2.103) and Einstein relation, it can be shown that

$$\frac{J_{2no}}{J_{2po}} = \frac{\mu_n\, n_{po}\, L_p}{\mu_p\, p_{no}\, L_n} = \frac{\mu_n\, n_{no}\, L_p}{\mu_p\, p_{po}\, L_n} = \frac{\sigma_n\, L_p}{\sigma_p\, L_n} \tag{2.104}$$

Thus, the ratio of electron and hole currents in a forward biased *P-N* junction is determined by the ratio $\frac{L_p}{L_n}$ (which may be slightly less than unity) and the ratio between the conductivity σ_n of the *N*-region and the conductivity σ_p of the *P*-region. Suppose the conductivity σ_p of the *P*-region is 50 times larger than that of the *N*-region σ_n, then the electron current is less than 2% of the total current, while the hole current is more than 98% of the total *P-N* junction current. This idea is useful in understanding the transistor action.

2.25 JUNCTION BREAKDOWN

When an ordinary *P-N* junction diode is reverse biased, normally only very small reverse saturation current flows. This current is due to movement of minority carriers. It is almost independent of the voltage applied. However, if the reverse bias is increased, a point is reached when the junction breaks down and the reverse current increases abruptly, as shown in Fig. 2.34. This current could be large enough to destroy the junction. If the reverse current is limited by means of a suitable series resistor, the power dissipation at the junction will not be excessive, and the device may be operated continuously in its breakdown region to its normal (reverse saturation) level. It is found that for a suitably designed diode, the breakdown voltage is very stable over a wide range of reverse currents. This quality gives the breakdown diode many useful applications as a voltage reference source.

The critical value of the voltage, at which the breakdown of a *P-N* junction diode occurs is called the breakdown voltage. The breakdown voltage depends on the width of the depletion region, which in turn, depends on the doping level. The junction offers almost zero resistance at the breakdown point.

There are two mechanisms by which breakdown point at a reverse biased *P-N* junction: avalanche breakdown and zener breakdown.

The minority carriers, under reverse biased conditions, flowing through the junction acquire a kinetic energy which increases with the increase in reverse

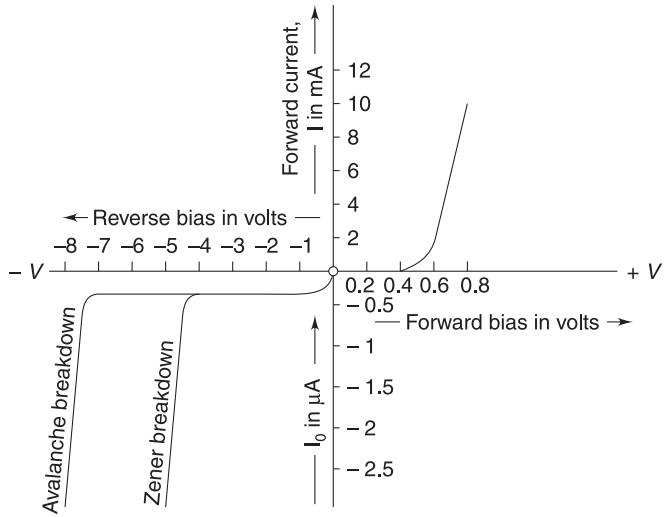


Fig. 2.34

voltage. At a sufficiently high reverse voltage (say 5 V or more), the kinetic energy of minority carriers becomes so large that they knockout electrons from the covalent bonds of the semiconductor material. As a result of collision, the liberated electrons in turn liberate more electrons and the current become very large leading to the breakdown of the crystal structure itself. This phenomenon is called the *avalanche breakdown*. The breakdown region is the 'knee' of the characteristic curve. Now the current is not controlled by the junction voltage but rather by the external circuit.

Under a very high reverse voltage, the depletion region expands and the potential barrier increases leading to a very high electric field across the junction. The electric field will break some of the covalent bonds of the semiconductor atoms leading to a large number of free minority carriers, which suddenly increase the reverse current. This is called the *zener effect*. The breakdown occurs at a particular and constant value of reverse voltage called the breakdown voltage. It is found that zener breakdown occurs at electric field intensity of about 3×10^7 V/m.

Either of the two (zener breakdown or avalanche breakdown) may occur independently, or both of these may occur simuntaneously. Diode junctions that breakdown below 5 V are caused by zener effect. Junctions that experience breakdown above 5 V are caused by avalanche effect. Junctions that breakdown around 5 V are usually caused by combination of two effects. The zener breakdown occurs in heavily doped junctions (*P*-type semiconductor moderately doped and *N*-type heavily doped), which produce narrow depletion layers. The

avalanche breakdown occurs in lightly doped junctions, which produce wide depletion layers.

With the increase in junction temperature zener breakdown voltage is reduced while the avalanche breakdown voltage increases. The zener diodes have a negative temperature coefficient while avalanche diodes have a positive temperature coefficient. Diodes that have breakdown voltages around 5 V have zero temperature coefficient.

The breakdown phenomenon is reversible and harmless so long as the safe operating temperature is maintained.

Despite the fact that avalanche and zener constitute two types of breakdown diode, the name zener is commonly applied to both types of diodes.

2.26 THE IDEAL DIODE

We have seen that a diode has a very important property. It permits only unidirectional conduction. It conducts well in the forward direction and poorly in the reverse direction. It would have been ideal if a diode acted as a perfect conductor (with zero voltage across it) when forward biased, and as a perfect insulator (with no current through it) when reverse biased. The *V-I* characteristics of such as ideal diode would be as shown in Fig. 2.35(a). An ideal diode acts like an automatic switch. When the current tries to flow in the forward direction, the switch is closed. On the other hand, when the current tries to flow the other way (against the direction of the diode arrow) the switch is open.

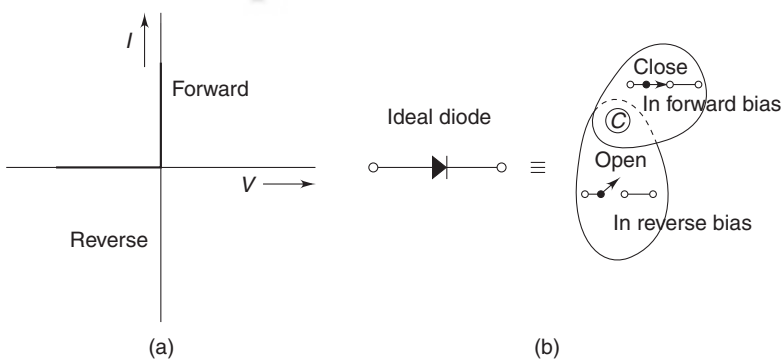


Fig. 2.35 *(a) Ideal-diode characteristics; (b) Switch analogy*

2.27 STATIC AND DYNAMIC RESISTANCE OF A DIODE

No diode can act as an ideal diode. An actual diode does not behave as a perfect conductor when forward biased, and as a perfect insulator when reverse biased. It does not offer zero resistance when forward biased. Also its reverse-resistance, through very large, is not infinite.

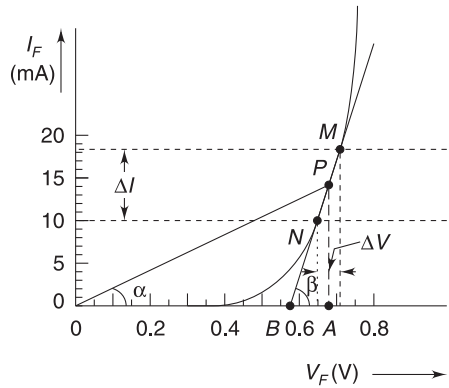


Fig. 2.36 *Calculation of static and dynamic resistance of a diode*

Figure 2.36 shows the forward characteristics of a typical silicon diode. This diode may be connected in a d.c. circuit. When forward biased, it offers a definite resistance in the circuit. This resistance is known as the d.c. or static resistance (R_F) of the diode. It is simply the ratio of the d.c. voltage across the diode to the d.c. current flowing through it. For instance, if the d.c. voltage across the diode is 0.7 V, the current through it can be found from Fig. 2.36. The operating point of the diode is at point *P*, and the corresponding current can be read as 14 mA. The static resistance of the diode at this operating point will be given as

$$R_F = \frac{OA}{AP} = \frac{0.7 \text{ V}}{14 \text{ mA}} = 50 \ \Omega$$

In general, the static resistance is given by the cotangent of the angle α. That is

$$R_F = \frac{OA}{AP} = \cot \alpha \tag{2.105}$$

If the characteristic is linear, this ratio *OA*/*AP* will be a constant quantity. But, in case the characteristic is nonlinear, the dc resistance will vary with the point of measurement.

In addition to 14 mA of d.c. current, small ac current may be superimposed in the circuit. The resistance offered by the diode to this a.c. signal is called its dynamic or ac resistance. The ac resistance of a diode, at a particular d.c. voltage, is equal to the reciprocal of the slope of the characteristic at that point, i.e.,

$$r_f = \frac{\text{change in voltage}}{\text{resulting change in current}} = \frac{\Delta V}{\Delta I} \tag{2.106}$$

[**Note:** The Greek letter Δ(delta) means "a change of", wherever it appears in formulae. So, ΔI is a change in current. Generally, it indicates a small-scale (or incremental) change].

We can calculate the ac resistance of a diode as follows:

Around the operating point P, take two points M and N very near to it, as shown in Fig. 2.36. These two points will then indicate incremental changes in voltage and current. The dynamic resistance is related to the slope of the line MN and is calculated as follows.

$$r_f = \frac{\Delta V}{\Delta I} = \frac{(0.73 - 0.66)\text{ V}}{(17.5 - 10)\text{ mA}}$$

$$= \frac{0.07\text{ V}}{7.5\text{ mA}} = 9.46\ \Omega \tag{2.107}$$

The smaller the incremental changed ΔV and ΔI, the closer is the above result to the exact value of the dynamic resistance. For making these incremental values smaller, the points M and N have to be closer. It then becomes difficult to read the voltage and current values accurately from the graph. We can circumvent this difficulty if we remember that as ΔV becomes smaller and smaller, the slope of the line MN becomes the same as that of the tangent to the curve at point P. In this alternative approach, we first draw a tangent to the curve at point P. This tangent meets the x-axis at point B (see Fig. 2.36). The dynamic resistance of the diode is then given as

$$r_f = \frac{BA}{AP} = \cot\beta \tag{2.108}$$

From the graph, we can calculate the dynamic resistance as

$$r_f = \frac{BA}{AP} = \frac{(0.7 - 0.57)\text{ V}}{14\text{ mA}}$$

$$= \frac{0.13\text{ V}}{14\text{ mA}} = 0.3\ \Omega \tag{2.109}$$

This may be seen to be almost the same as the value obtained earlier.

Now look at the reverse characteristic of the PN-junction diode (Fig. 2.32). We find that ever for a large reverse voltage (but below breakdown) the current is very small. The reverse current may be 1 μA at a voltage of 5 V. Then the static resistance of the diode is

$$R_R = \frac{5\text{ V}}{1\ \mu A} = 5\ \text{M}\Omega \tag{2.110}$$

This is sufficiently high. It is much higher than the forward resistance R_F. Since the diode curve in the reverse bias is almost horizontal, its dynamic resistance r_r will be extremely high in this region of operation.

Example 2.2 *Why is Si preferred over Ge in the manufacture of semiconductor devices?*

Solution The silicon semiconductor devices have, in general, higher PIV and current ratings and wider temperature range than germanium semiconductor devices and therefore, silicon is preferred over germanium in the manufacture of semiconductor devices.

Example 2.3 *What is a step-graded junction?*

Solution A junction is said to be step-graded if there is an abrupt change from acceptor ion concentration on the *P*-side to donor ion concentration on the *N*-side such as alloyed or fused junction.

Example 2.4 *What is a linear graded junction?*

Solution A junction is said to be linearly graded if the charge concentration varies gradually with the distance in its transition region such as a grown junction.

2.28 LOAD LINE CONCEPT

Kirchhoff's voltage law is applicable both to linear and nonlinear circuits, so we may write (Fig. 2.37)

$$V = V_D + V_{out}$$
$$= V_D + IR_L \qquad (2.111)$$

where I is the circuit current

or $$I = \frac{V}{R_L} - \frac{V_D}{R_L} \qquad (2.112)$$

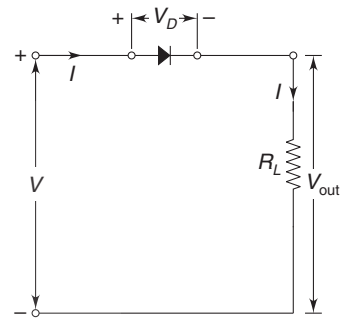


Fig. 2.37 *Basic diode circuit*

The above equation gives a linear relation between the diode or circuit current I and the diode voltage V_D for a given input voltage V and load resistance R_L. This equation is referred to as the circuit load line, and is usually plotted on graph with the current I as the vertical axis and the voltage V_D as the horizontal axis. (Fig. 2.38). From Eq. (2.112), we see that for current $I = 0$, $V_D = $ V which is the horizontal axis intercept. Also from this equation, if $V_D = 0$, then current $I = \frac{V}{R_L}$ which is the vertical axis intercept. The load line can be drawn between these two points. The slope of the load line is determined by R_L, the negative value of the slope is equal to $\frac{1}{R_L}$.

The intersection of load line with the static characteristic of diode provides the operating point, sometimes referred to as quiescent or *Q point*. The d.c. load line illustrates all d.c. conditions that could exist within the circuit for given supply voltage V and load resistance R_L.

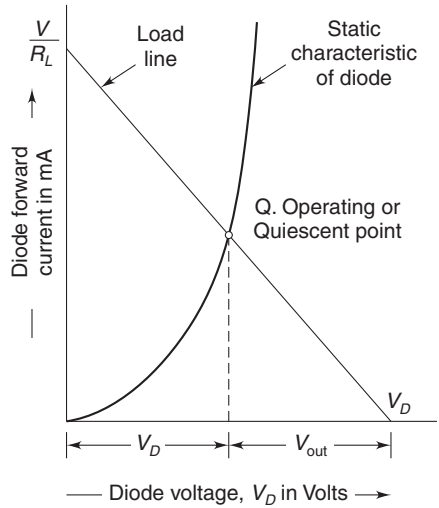


Fig. 2.38 *The diode and load line characteristics for the circuit given in Fig. 2.37*

2.29 DYNAMIC CHARACTERISTICS

Now let a variable voltage v_{in} be applied to a diode circuit, depicted in Fig. 2.39. A plot of current i vs input voltage v_{in} is called the dynamic characteristic. The curve is drawn for different values of input voltage v_{in}.

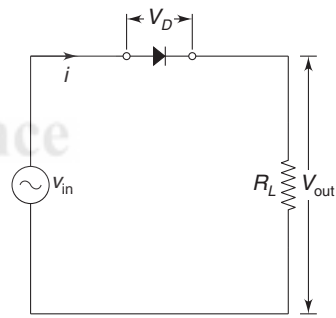


Fig. 2.39 *Basic diode circuit*

Figure 2.40 depicts the load lines for two voltages v_{in} and v'_{in}. It is observed from Fig. 2.40 that the load lines drawn for the two input voltages v_{in} and v'_{in} are parallel to each other. This is because slope of the load line does not vary with the change in input voltage since R_L is constant. The load lines intersect the static characteristics at points A and A' respectively. Now, vertical lines are drawn from v_{in} and v'_{in} and horizontal lines are drawn from points A and A' respectively. A curve drawn passing through $B'BO$ is known as *dynamic characteristic*.

2.30 TRANSFER CHARACTERISTIC

The curve that relates the output voltage v_{out} to the input voltage v_{in} of any circuit is known as the *transfer* or *transmission characteristic*. Figure 2.41 illustrates the graphical method for determination of output voltage waveform for a given input voltage waveform (say, sinusoidal waveform) using transfer characteristic.

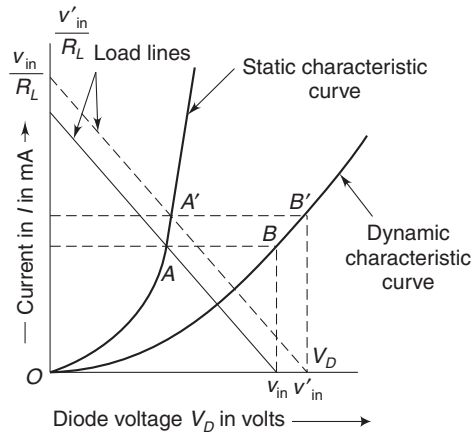


Fig. 2.40 *Dynamic characteristic*

The input signal waveform is drawn with its time axis vertically downward, so that the voltage axis is horizontal. The different input voltages are at different time instants t_1, t_2, t_3 etc., and corresponding vertical lines are drawn to meet the transfer curve. Now horizontal lines are drawn from the intersection points on transfer curve to provide the output voltage corresponding to time instants t_1, t_2, t_3 etc. Thus output voltage waveform can be obtained by joining these points. From Fig. 2.41, it can be visualized that (i) the output voltage is distorted at the start due to nonlinear transfer characteristic and (ii) for $v_{in} < V_r$, the diode acts as a clipper and a portion of the input signal does not appear at the output.

2.31 CLIPPERS

There are a variety of diode networks called *clippers* that have the ability to "clip" off a portion of the input signal without distorting the remaining part of the alternating waveform.

There are two general categories of clippers: *series* and *parallel*. The series configuration is defined as one where the diode is in series with the load, while the parallel variety has the diode in a branch parallel to the load.

2.31.1 Series

The response of the series configuration of Fig. 2.42(a) to a variety of altering waveforms is provided in Fig. 2.42. Although first introduced as a half-wave rectifier (for sinusoidal waveforms), there are no boundaries on the type of signals that can be applied to a clipper. The addition of a d.c. supply such as shown in Fig. 2.43 can have a pronounced effect on the output of a clipper.

There is no general procedure for analyzing networks such as the type in Fig. 2.43, but there are a few thoughts to keep in mind as you work toward a solution.

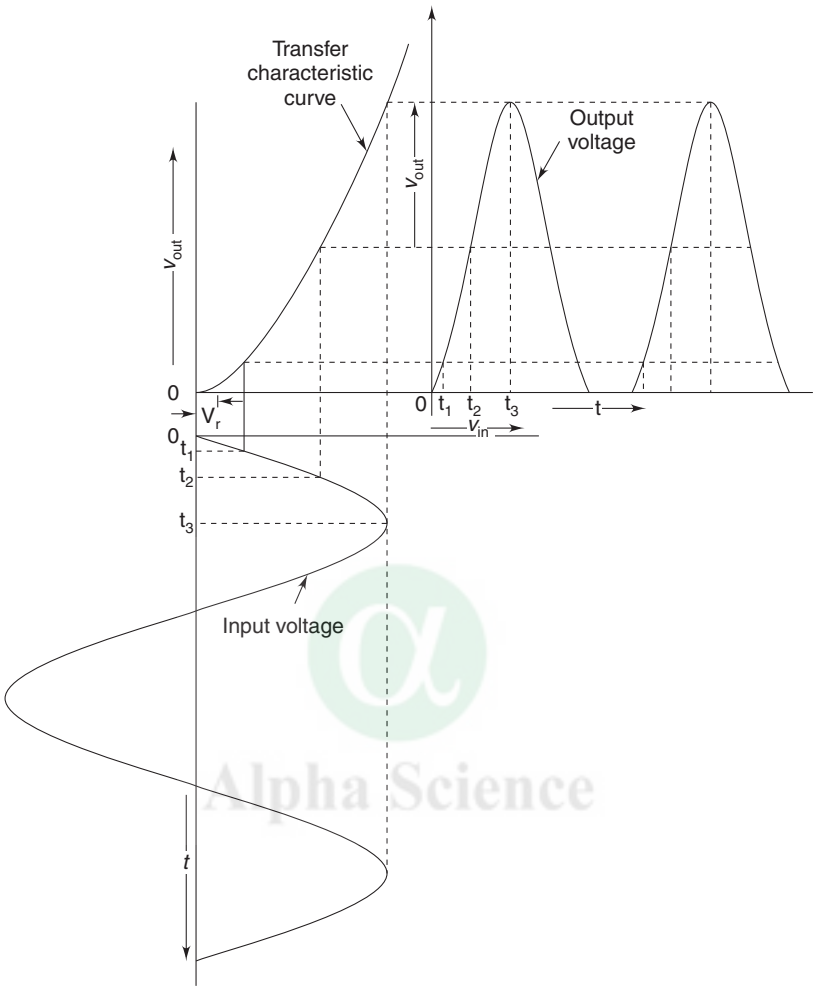


Fig. 2.41 *Graphical method for obtaining output voltage waveform using transfer curve*

1. *Make a mental sketch of the response of the network based on the direction of the diode and the applied voltage levels.*

For the network of Fig. 2.43, the direction of the diode suggests that the signal v_i must be positive to turn it on. The d.c. supply further requires that the voltage v_i be greater than V volts to turn the diode on. The negative region of the input signal is "pressuring" the diode into the "off" state, supported further by the d.c. supply. In general, therefore, we can be quite sure that the diode is an open circuit ("off" state) for the negative region of the input signal.

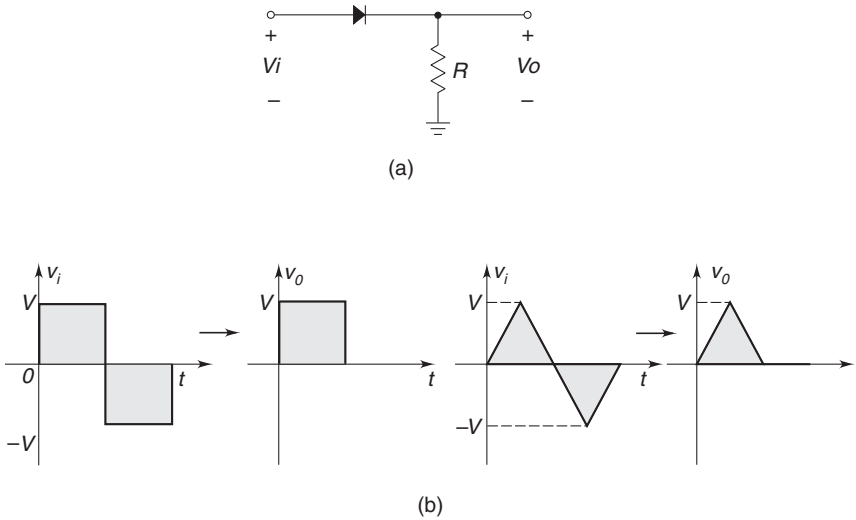


Fig. 2.42 *Series clipper*

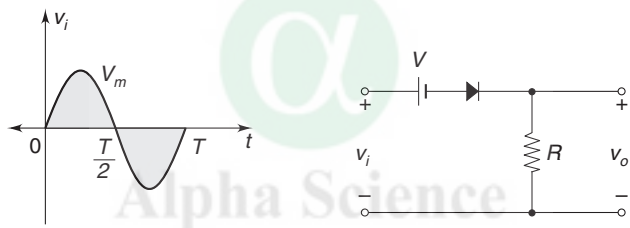


Fig. 2.43 *Series clipper with a d.c. supply*

2. *Determination the applied voltage (transition voltage) that will cause a change in state for the diode.*

For the ideal diode the transition between states will occur at the point on the characteristics where $v_d = 0$ V and $i_d = 0$ A. Applying the condition $i_d = 0$ at $v_d = 0$ to the network of Fig. 2.43 will result in the configuration of Fig. 2.44 where it is recognized that the level of v_i that will cause a transition in state is

$$v_i = V \tag{2.113}$$

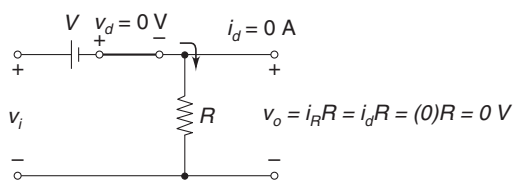


Fig. 2.44 *Determining the transition level for the circuit of Fig. 2.68*

For an input voltage greater than V volts the diode is in the short-circuit state, while for input voltages less than V volts it is in the open-circuit or "off" state.

3. *Be continually aware of the defined terminals and polarity of v_o.*

When the diode is in the short-circuit state, such as shown in Fig. 2.45, the output voltage v_o can be determined by applying Kirchhoff's voltage law in the clockwise direction:

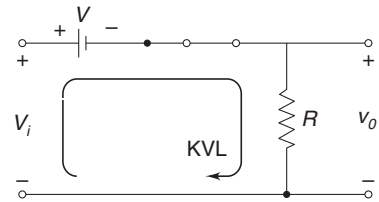


Fig. 2.45 *Determining v_o*

$$v_i - V - v_o = 0 \quad \text{(CW direction)}$$

and
$$v_o = v_i - V \qquad (2.114)$$

4. *It can be helped to sketch the input signal above the output and determine the output at instantaneous values of the input.*

It is then possible that the output voltage can be sketched from the resulting data points of v_o as demonstrated in Fig. 2.46. Keep in mind that at an instantaneous value of v_i the input can be treated as a dc supply of that value and the corresponding dc value (the instantaneous value) of the output determined. For instance, at $v_i = V_m$ for the network of Fig. 2.43, the network to be analyzed appears in Fig. 2.47. For $V_m > V$ the diode is in the short-circuit state and $v_o = V_m - V$, as shown in Fig. 2.46.

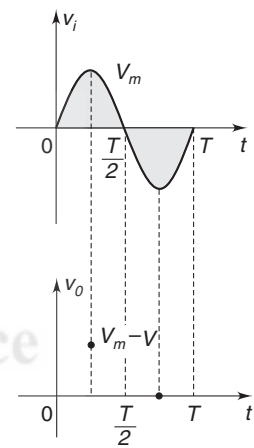


Fig. 2.46 *Determining levels of v_o*

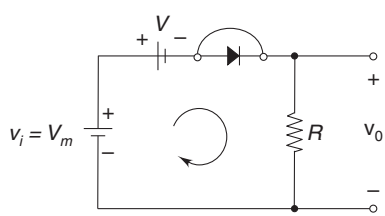


Fig. 2.47 *Determining v_o when $v_i = V_m$*

At $v_i = V$ the diodes change state and at $v_i = -V_m$, $v_o = 0$ V, and the complete curve for v_o can be sketched as shown in Fig. 2.48.

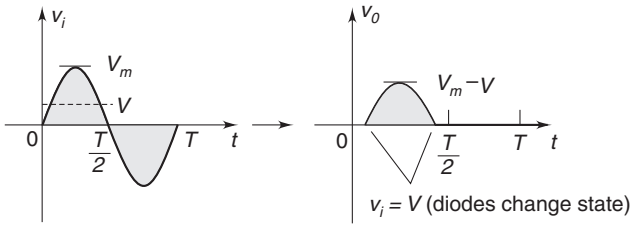


Fig. 2.48 *Sketching v_o*

Example 2.5 *Determine the output waveform for the network of Fig. 2.49.*

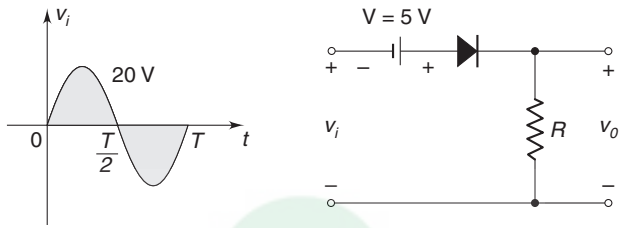


Fig. 2.49 *Series clipper for Example 2.5*

Solution Experience suggests that the diode will be in the "on" state for the positive region of v_i — expecially when we note the aiding effect of V = 5 V. The network will then appear as shown in Fig. 2.50 and $v_o = v_i + 5$ V. Substituting $i_d = 0$ at $v_d = 0$ for the transition levels, we obtain the network of Fig. 2.51 and $v_i = -5$ V.

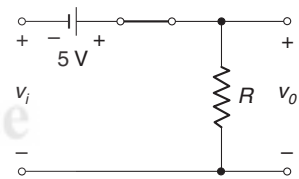


Fig. 2.50 *v_o with diode in the "on" state*

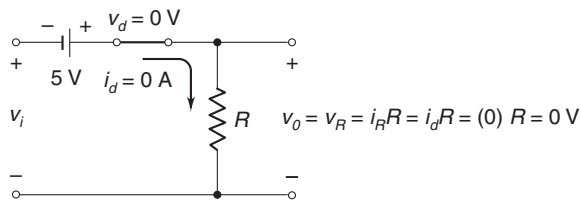


Fig. 2.51 *Determining the transition level for the clipper of Fig. 2.49*

For v_i more negative than –5V the diode will enter its open-circuit state, while for voltage more positive than –5V the diode is in the short-circuit state. The input and output voltages appear in Fig. 2.52.

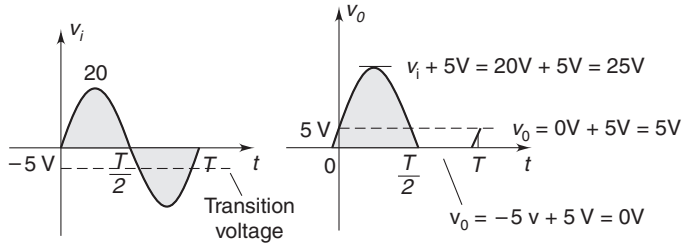


Fig. 2.52 *Sketching v_o for Example 2.5*

2.31.2 Parallel

The network of Fig. 2.53 is the simplest of parallel diode configurations with the output for the same inputs of Fig. 2.42 (The analysis of parallel configurations is similar to that applied to series configuration).

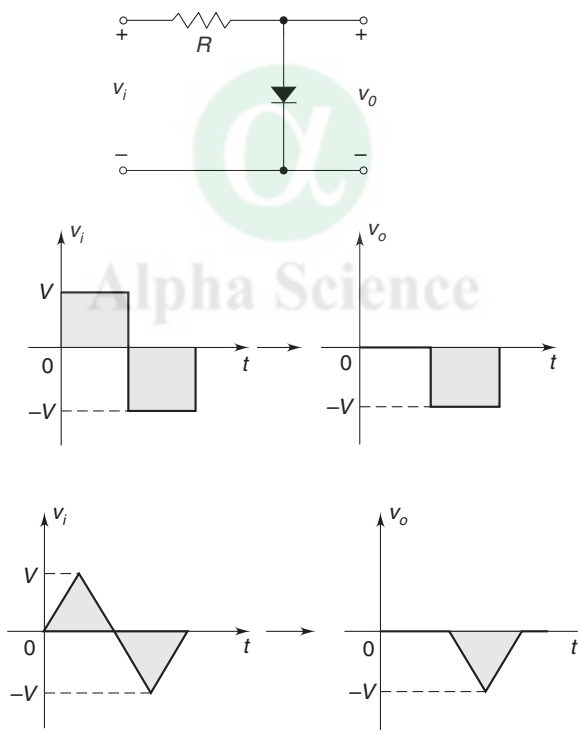


Fig. 2.53 *Response to a parallel clipper*

A variety of series and parallel clippers with the resulting output for the sinusoidal input are provided in Fig. 2.54. In particular, note the response of the last configuration, with its ability to clip off a positive and a negative section as determined by the magnitude of the d.c. supplies.

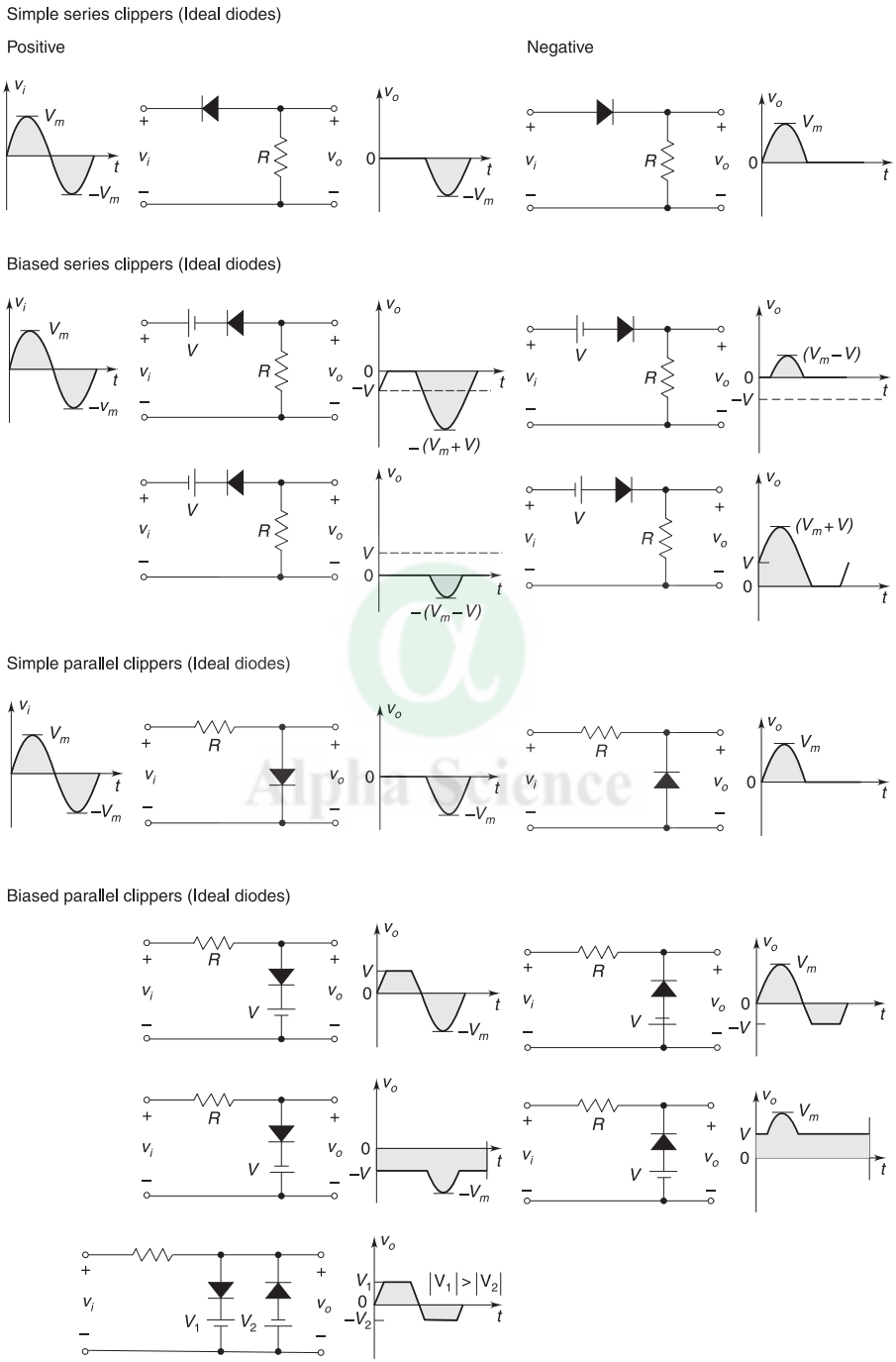
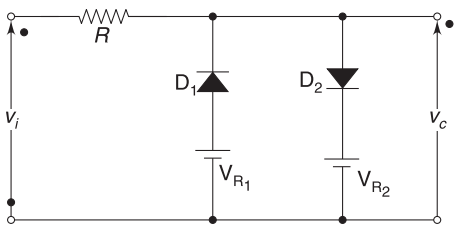


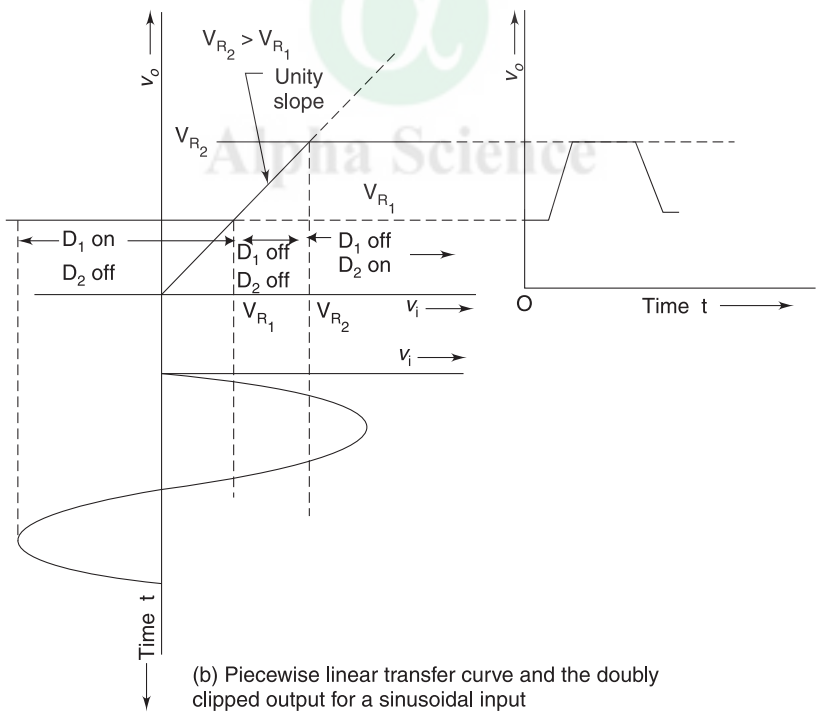
Fig. 2.54 *Clipping circuits*

2.31.3 Clipping at Two Levels

Two clipper diodes may be used in one clipper circuit to perform clipping at two independent levels. The following combinations of the two clipping diodes are possible: (a) both diodes is series arm (b) both diodes in shunt arms and (c) one diode in series arm and the other in shunt arm. Figure 2.55(a) shows a parallel arrangement. Figure 2.55(b) shows the corresponding piecewise linear transmission characteristics and also the output waveform for sinusoidal input.



(a) Double diode clipper



(b) Piecewise linear transfer curve and the doubly clipped output for a sinusoidal input

Fig. 2.55 *A double-diode clipper which clips at two independent levels*

It may be seen from Fig. 2.55 that the transfer curve has two break points one at $v_o = v_i = V_{R_1}$ and the other at $v_o = v_i = V_{R_2}$. Then for $V_{R_2} > V_{R_1} >> V_o$ and $R_F << R$, the transfer curve has the following characteristics:

Input v_i	Output v_o	Diode states
$v_i << V_{R_1}$	$v_o = V_{R_1}$	D_1 ON, D_2 OFF
$V_{R_1} < v_i < V_{R_2}$	$v_o = v_i$	D_1 OFF, D_2 OFF
$v_i \geq V_{R_2}$	$v_o = V_{R_2}$	D_1 OFF, D_2 ON

The double-diode clipper of Fig. 2.55(a) is of often called a slicer since the output is nothing but a slice of the input between two reference levels V_{R_1} and V_{R_2}.

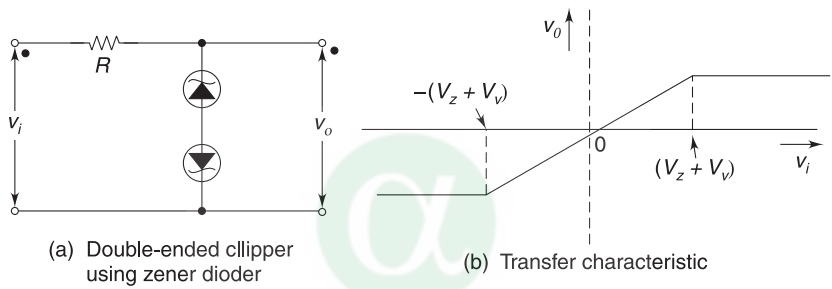


Fig. 2.56 *A double-ended clipper using two zener diodes and its transfer characteristics*

The double-diode clipper circuit of Fig. 2.55(a) is used for converting a sinusoidal waveform into a square waveform. However, in order to generate a symmetrical square wave V_{R_1} and V_{R_2} are made equal and opposite in sign. In that case, the transfer characteristic passes through the origin and the sinusoidal waveform gets clipped symmetrical at the top and at the bottom. The clipped waveform so obtained at the output is not square waveform. However, if the amplitude of the sinusoidal waveform is large in comparison with the difference in the reference levels, then the output waveform is almost square waveform.

Figure 2.56(a) shows a double-diode clipper using two zener diode series opposition. On using diodes with identical characteristics, symmetrical limiter operation is achieved. The transfer characteristic is shown in Fig. 2.56(b), where V_z is the breakdown (zener) voltage and V_v is the cutin voltage of the diode.

2.32 CLAMPERS

The clamping network is one that will "clamp" s signal to a different d.c. level. The network must have a capacitor, a diode, and a resistive element, but it can also employ an independent d.c. supply to introduce an additional shift. The

magnitude of R and C must be chosen such that the time constant $\tau = RC$ is large enough to ensure that the voltage across the capacitor does not discharge significantly during the interval the diode is nonconducting. Throughout the analysis we will assume that for all practical purposes the capacitor will fully charge or discharge in five time constants.

The network of Fig. 2.57 will clamp the input signal to the zero (for ideal diodes). The resistor R can be the load resistor or a parallel combination of the load resistor and a resistor designed to provide the desired level of R.

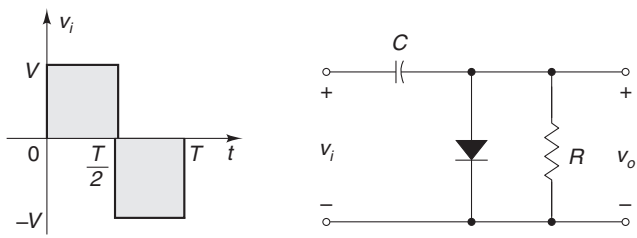


Fig. 2.57 *Clamper*

During the interval $0 \rightarrow T/2$ the network will appear as shown in Fig. 2.58, with the diode in the "on" state effectively "shoring out" the effect of the resistor R. The resulting RC time constant is so small (R determined by the inherent resistance of the network) that the capacitor will charge to V volts very quickly. During this interval the output voltage is directly across the short circuit and $v_o = 0$ V.

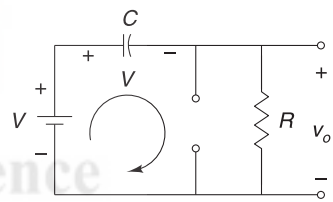


Fig. 2.58 *Diode "on" and the capacitor charging to V volts*

When the input switches to the $-V$ state, the network will appear as shown in Fig. 2.59, with the open-circuit equivalent for the diode determined by the applied signal and stored voltage across the capacitor—both "pressuring" current through the diode from cathode to anode. Now that R is back in the network the time constant determined by the RC product is sufficiently large to establish a discharge period 5τ much greater than the period $T/2 \rightarrow T$ and it can be assumed on an approximate basis that the capacitor holds onto all its charge and therefore voltage (since $V = Q/C$)during this period.

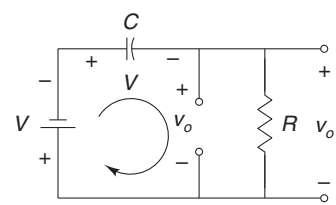


Fig. 2.59 *Determining v_o with the diode "off"*

Since v_o is in parallel with the diode and resistor, it can also be drawn in the alternative position shown in Fig. 2.59. Applying Kirchhoff's voltage law around the input loop will result in

$$-V - V - v_o = 0$$

and $$v_o = -2V$$

The negative sign resulting from the fact that the polarity of 2V is opposite to the polarity defined for v_o. The resulting output waveform appears in Fig. 2.60 with the input signal. The output signal is clamped to 0 V for the interval 0 to $T/2$ but maintains the same total swing (2V) as the input.

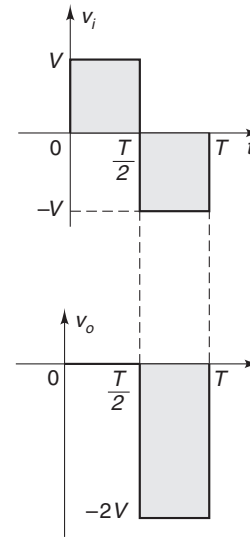


Fig. 2.60 *Sketching v_o for the network of Fig. 2.92*

For a clamping network:

The total swing of the output is equal to the total swing of the input signal.

This fact is an excellent checking tool for the result obtained.

In general, the following steps may be helpful when analyzing clamping networks:

1. *Start the analysis of clamping networks by considering that part of the input signal that will forward bias the diode.*

   The statement above may require skipping an interval of the input signal (as demonstrated in an example to follow), but the analysis will not be extended by an unnecessary measure of investigation.
2. *During the period that the diode is in the "on" state, assume that the capacitor will charge up instantaneously to a voltage level determined by the network.*
3. *Assume that during the period when the diode is in the "off" state the capacitor will hold on to its established voltage level.*
4. *Throughout the analysis maintain a continual awareness of the location and reference polarity for v_o to ensure that the proper levels for v_o are obtained.*
5. *Keep in mind the general rule that the total swing of the total output must match the swing of the input signal.*

Example 2.6 *Determine v_o for the network of Fig. 2.61 for the input indicated.*

Solution Note that the frequency is 1000 Hz, resulting in a period of 1 ms and an interval of 0.5 ms between levels. The analysis will begin with the period $t_1 \rightarrow t_2$ of the input signal since the diode is in its short-circuit state as recommended by comment 1. For this interval the network will appear as

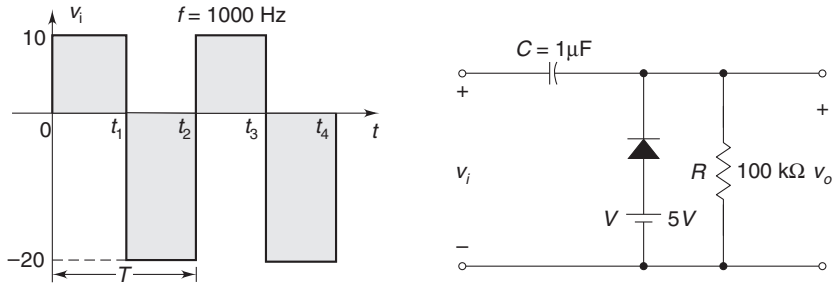


Fig. 2.61 *Applied signal and network for Example 2.6*

shown in Fig. 2.62. The output is across R, but it is also directly across the 5 V battery if you follow the direct connection between the defined terminals for v_o and the battery terminals. The result is v_o = 5 V for this interval. Applying Kirchhoff's voltage law around the input loop will result in

$$-20\ V + V_C - 5V = 0$$

and

$$V_C = 25\ V$$

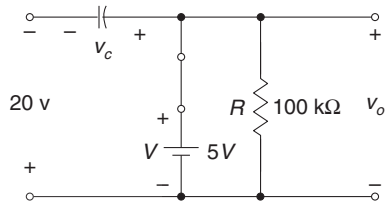


Fig. 2.62 *Determining v_o and V_C with the diode in the "on" state*

The capacitor will therefore charge up to 25 V, as stated in comment 2. In this case the resistor R is not shorted out by the diode but a Thevenin equivalent circuit of that portion of the network which includes the battery and the resistor will result in $R_{Th} = 0\ \Omega$ with $E_{TH} = V = 5$ V. For the period $t_2 \rightarrow t_3$ the network will appear as shown in Fig. 2.63.

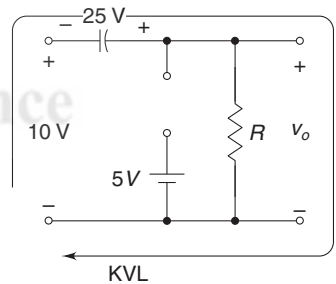


Fig. 2.63 *Determining v_o with the diode in the "off" state*

The open-circuit equivalent for the diode will remove the 5 V battery from having any effect on v_o, and applying Kirchhoff's voltage law around the outside loop of the network will result in

$$+10\ V + 25\ V - v_o = 0$$

and

$$v_o = 35\ V$$

2.33 *P-N* DIODE APPLICATIONS

An ideal *P-N* junction diode is a two terminal polarity sensitive device that has zero resistance (diode conducts) when it is forward biased and infinite resistance (diode does not conduct) when reverse biased. Because of this property, the *P-N* diode finds use in many applications as enumerated on next page.

(i) As rectifiers in d.c. power supplies.
(ii) In demodulation or detector circuits.
(iii) In clamping networks in TV receivers and voltage multipliers.
(iv) In clipping circuits used as wave shaping circuits in computers, radars, radio and TV receivers.
(v) As switches in digital logic circuits.

A homo junction (i.e. *P* and *N* of same material) but with different doping concentrations finds special applications as follows:

(i) As zener diodes in voltage regulators, peak clippers and in switching operation.
(ii) As tunnel diode as a relaxation oscillator at microwave frequencies.
(iii) As light emitting diodes (LEDs) in digital displays.
(iv) As LASER diodes in optical communications.
(v) As varactor diodes in tuning sections of radio and TV receivers.
(vi) As detectors (PIN photodiode) in optical communication circuits.

Questions

2.1 Explain why the discrete energy levels of an isolated atom split into a band of energy when atoms combine together to form a crystal.
2.2 Explain the formation of energy bands in diamond.
2.3 Explain the difference in conductors, insulators and semiconductors using the energy band diagrams.
2.4 Sketch the two dimensional crystal structure of intrinsic silicon at the absolute zero of temperature. Also sketch its energy-band diagram. Sketch the same crystal structure at room temperature. Also sketch its energy band diagram.
2.5 Explain the reason why the conductivity of germanium is more than that of silicon at room temperature. (Hint: see Table 2.1).
2.6 Explain why a pentavalent impurity atom is known as donor-type impurity.
2.7 Of what polarity are the impurity ions in *N*-type and *P*-type semiconductors? Justify your answer in brief (5-6 lines).
2.8 What is a Fermi level. Prove that it lies at half the energy-gap in case of an intrinsic semiconductor.
2.9 What is *P-N* junction diode? How potential barrier is formed in a *P-N* junction diode? Derive an expression for the constant potential.

2.10 What is transition or space charge capacitance for a *P-N* junction? Derive an expression of transition capacitance for a step-graded junction.

2.11 Draw the energy band diagram for a *P-N* junction under open-circuited condition. Clearly indicate various energy levels in *P*-region, depletion region and *N*-region. How will it be modified if *P-N* junction is forward biased? Derive the diode equation for a forward biased diode.

2.12 Discuss behaviour of *P-N* junction diode under forward and reverse biasing, with reference to energy band diagrams.

2.13 Draw sketches to show the method of fabricating alloy diodes and diffused diodes. Briefly explain.

2.14 Explain the breakdown mechanism in a diode.

2.15 Explain how the process of avalanche breakdown occurs in a *P-N* junction diode. How is it different from zener breakdown?

2.16 What is meant by static and dynamic resistance of a diode?

2.17 What is hot carrier effect? Explain saturation of drift velocity at high electric fields for Si.

2.18 Draw the diagrams of a diode (i) series clipper (ii) parallel clipper.

2.19 Draw the four typical diode clipping circuits. Assuming ideal diode and sinusoidal input, draw the output waveforms.

2.20 Draw the circuit of a double-diode clipper with both diodes in shunt arms. Draw the piecewise linear transmission characteristics assuming ideal diodes.

3 Two Terminal Devices and their Applications

3.1 INTRODUCTION (D.C. POWER SUPPLY)

The electrical power is almost exclusively generated, transmitted and distributed in the form of a.c. because of economical consideration but, for operation of most of the electronic devices and circuits, d.c. power supply is required. Dry cells and batteries can be used for this purpose. No doubt, they have the advantages of being portable and ripple-free but their voltages are low, they need frequent replacement and are expensive in comparison to convenient dc power supplies.

Now-a-days, almost all electronic equipment include a circuit that converts a.c. supply into d.c. supply. The part of equipment that converts a.c. into d.c. is called the *d.c. power supply*. In general, at the input of the power supply, there is a power transformer. It is followed by a *rectifier* (a diode circuit), a *smoothing filter*, and then a *voltage regulator* circuit. A block diagram of such a power supply is given in Fig. 3.1.

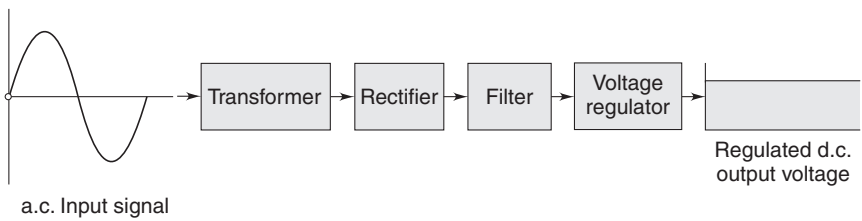


Fig. 3.1 *Block diagram of a d.c. Power supply*

As obvious from this diagram, the basic power supply is constituted by four elements *viz.*, a *transformer*, a *rectifier*, a *filter* and a *regulator* put together. The output of the d.c. power supply is used to provide a constant d.c. voltage across the load. Now, let us briefly outline the function of each of the elements of the d.c. power supply.

Transformer is used to step-up or step-down (usually to step-down) the supply voltage as per need of the devices and circuits to be supplied by the d.c. power supply. It can provide isolation from the supply line – an important safety consideration.

Rectifier is a device which converts the sinusoidal a.c. voltage into either positive or negative pulsating d.c. *P-N* junction diode, which conducts when forward biased and practically does not conduct when reverse biased, can be used for rectification i.e., for conversion of a.c. into d.c. The rectifier typically needs one, two or four diodes. Rectifiers may be either halfwave or fullwave (centre-tap or bridge-) type.

The output voltage from a rectifier circuit has a pulsating character i.e., it contains unwanted a.c. components (components of supply frequency f and its harmonics) along with d.c. component. To reduce ac components, a *filter circuit* is required. A filter is a device which passes d.c. component to the load and blocks a.c. components of the rectifier output. A filter is typically formed from reactive circuit elements such as capacitor and/or inductors and resistors.

The magnitude of output d.c. voltage may vary with the variation of either the input a.c. voltage or the magnitude of load current. So, at the output of a rectifier-filter combination, a voltage regulator is required as shown in Fig. 3.1, to provide an almost constant d.c. voltage at the output of the regulator. The voltage regulator may be constructed from a zener-diode and/or discrete transistors, and/or integrated circuits (ICs). Its main function is to maintain a constant d.c. output voltage. However, it also rejects any a.c. ripple voltage that is not removed by the filter. The regulator may also include protective devices such as short-circuit protection, current-limiting, thermal shut down, or overvoltage protection.

3.2 RECTIFIERS

Rectifiers may be put into the following two categories depending upon the period of conduction:

(a) Halfwave Rectifiers A halfwave rectifiers is one which converts an a.c. voltage into a pulsating voltage using only one half-cycle of the applied a.c. voltage. The rectifying diode conducts during one half-cycle only.

(b) Fullwave Rectifiers A fullwave rectifiers is one which converts an a.c. voltage into a pulsating voltage using both half-cycles of the applied a.c. voltage. It typically uses two diodes, one of which conducts and provides output during (say) the positive half-cycle while the other diode conducts during the negative half-cycle.

3.3 HALFWAVE RECTIFIERS

Figure 3.2(a) gives the basic circuit of halfwave rectifier. The ac voltage to be rectified is applied to the input of the power transformer and the voltage across the secondary is available for rectification. Let this voltage v_i at the input of the rectifier be sinusoidal and given by,

$$v_i = V_m \sin \omega t \tag{3.1}$$

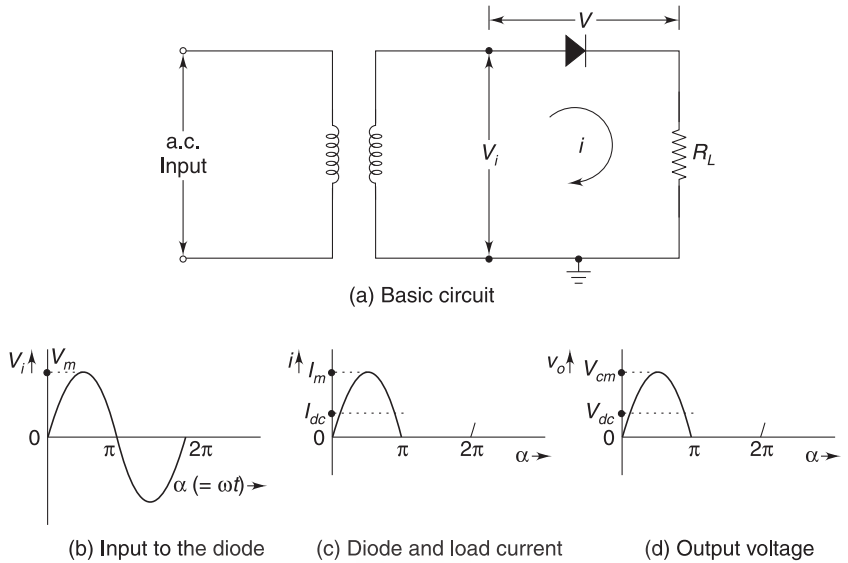


Fig. 3.2 *Halfwave rectifier*

Let the peak value V_m be very large as compared with the cut-in voltage V_v of the diode. Accordingly we assume here that $V_v = 0$. Further let diode be idealized with resistance R_f in the ON state and open-circuit ($R_r = \infty$) in the OFF state. Then the current i in the diode or the load resistance R_L is given by,

$$i = I_m \sin \alpha \qquad \text{for } 0 \le \alpha \le \pi \tag{3.2a}$$

and

$$i = 0 \qquad \text{for } \pi \le \alpha \le 2\pi \tag{3.2b}$$

where I_m is the peak current and is given by

$$I_m \equiv \frac{V_m}{R_f + R_L} \tag{3.3}$$

Figure 3.2(b) shows the voltage v_i while Fig. 3.2(c) shows the current i flowing through the diode and the load resistor R_L. It may be noted that this current i is a unidirectional current. The average current I_{dc} is then given by,

$$I_{dc} = \frac{1}{2\pi} \int_0^{2\pi} i \, dx \tag{3.4}$$

where i is given by Eqs. (3.2a) and (3.2b). No current flows during the interval π to 2π.

Hence,

$$I_{dc} = \frac{1}{2\pi} \int_0^{\pi} I_m \sin \alpha dx = \frac{I_m}{\pi} \tag{3.5}$$

$$= \frac{V_m}{\pi (R_f + R_L)} \tag{3.6}$$

3.3.1 The d.c. Output Voltage

The d.c. output voltage is given by,

$$V_{dc} = I_{dc} \cdot R_L = \frac{I_m R_L}{\pi} \tag{3.7}$$

$$= \frac{V_m}{\pi \left(1 + \frac{R_f}{R_L}\right)} \tag{3.8}$$

3.3.2 The Diode Voltage

The instantaneous voltage across the diode is $i.R$, where R is the diode resistance and equals R_f during the ON state and α in the OFF state. Accordingly the instantaneous voltage v across the diode is as shown in Fig. 3.3.

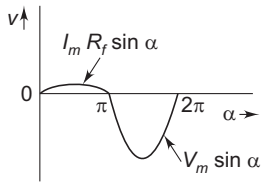


Fig. 3.3 *Instantaneous voltage across the diode*

The d.c. voltage V_{dc} across the diode is obtained on integration and is given by

$$V'_{dc} = \frac{1}{2\pi}\left[\int_0^{\pi} I_m R_f \sin \alpha \, d\alpha + \int_{\pi}^{2\pi} V_m \sin \alpha \, d\alpha\right] \tag{3.9}$$

$$= \frac{1}{\pi}\left[I_m R_f - V_m\right] \tag{3.10}$$

$$= \frac{1}{\pi}\left[I_m R_f - I_m (R_f + R_L)\right]$$

or

$$V'_{dc} = -\frac{I_m R_L}{\pi} \tag{3.11}$$

Thus V'_{dc} comes out to be negative. This signifies that the cathode is positive with respect to the anode.

From Eqs. (3.7) and (3.11) we conclude that the d.c. diode voltage is negative of the d.c. voltage across the load resistor R_L. This is to be expected since the sum of the two d.c. voltages around the complete circuit must add up to zero.

3.3.3 RMS Current and Voltage

The square of the r.m.s. value of a periodic function of time is, as per definition, given by the area of one full cycle of the cycle of the curve representing the square of the function, divided by the base. Thus mathematically r.m.s. value is given by,

$$I_{\text{rms}} = \left[\frac{1}{2\pi}\int_0^{2\pi} i^2 \, d\alpha\right]^{1/2} \tag{3.12}$$

Making use of Eqs. (3.11) and (3.12), we get

$$I_{\text{rms}} = \left[\frac{1}{2\pi}\int_0^{\pi} Im^2 \sin^2\alpha \, d\alpha\right]^{1/2} = \frac{I_m}{2} \tag{3.13}$$

$$= \frac{V_m}{2(R_f + R_L)} \tag{3.14}$$

An rms ammeter (or voltmeter) is constructed such that it indicates the effective or rms current (or voltage). However, most a.c. meters are of the rectifier type, instead of being true rms reading instruments.

RMS voltage across the load is,

$$V_{\text{rms}} = I_{\text{rms}} R_L = \frac{V_m}{2(1 + R_f/R_L)} \tag{3.15}$$

3.3.4 Frequency Components of Rectified Output

Fourier analysis of the current waveshape of a halfwave rectifier gives the following frequency components:

$$i = I_m\left[\frac{1}{\pi} + \frac{1}{2}\sin \omega t - \frac{2}{3\pi}\cos 4\omega t - \frac{2}{15\pi}\cos 4\omega t - \ldots\right] \tag{3.16}$$

It is assumed that the diode conducts during the first half of the ac cycle so that Eq. (3.1) applies. The first term of the series gives the dc or the average component and is the same as given by Eq. (3.5). The second term has its frequency same as the supply frequency and its peak value is $I_m/2$. The effective or the r.m.s. value of I_1 of this fundamental frequency component is given by

$$I_1 = \frac{\text{Peak value}}{\sqrt{2}} = \frac{I_m/2}{\sqrt{2}} = \frac{I_m}{2\sqrt{2}} \tag{3.17}$$

The third term gives the second harmonic term. Its r.m.s. value I_2 is given by,

$$\sqrt{2}\, I_2 = \frac{2}{3\pi} I_m$$

Hence

$$I_2 = \frac{\sqrt{2}\, I_m}{3\pi} \tag{3.18}$$

Similarly the rms value of the rest of the terms may be calculated. These terms are found to be of continuously diminishing values.

3.3.5 Rectifier Efficiency

In a rectifier the useful power output is the d.c. power P_{dc} developed across the load and is given by,

$$P_{dc} = I_{dc}^2 R_L = \frac{I_m^2 R_L}{\pi^2}$$

$$= \frac{E_m^2}{(R_f + R_L)} \cdot \frac{R_L}{\pi^2} \tag{3.19}$$

Out of the total power from the a.c. voltage source, a portion P_a is dissipated at the junction of the diode and the rest P_r is dissipated in the load resistance R_L. Since the rectifier itself is assumed to be ideal, the dissipation is assumed to take place in the resistance R_f. Then, we get

$$P_a = I_{\text{rms}}^2 R_f = \frac{I_m^2}{4} R_f \tag{3.20}$$

$$P_r = I_{\text{rms}}^2 R_L = \frac{I_m^2}{4} R_L \tag{3.21}$$

Hence total input power is,

$$P_{in} = P_a + P_r = \frac{I_m^2}{4} (R_f + R_L) \tag{3.22}$$

Efficiency of the rectifier is defined as the ratio of the d.c. output power P_{dc} to the a.c. input power P_{in}. Hence the rectifier efficiency η is given by,

$$\begin{aligned} \eta = \frac{P_{dc}}{P_{in}} &= \frac{I_m^2 R_L/\pi^2}{(I_m^2/4)(R_f + R_L)} \\ &= \left(\frac{2}{\pi}\right)^2 \cdot \frac{R_L}{R_f + R_L} \\ &= \frac{0.406}{1 + R_f/R_L} \end{aligned} \tag{3.23}$$

Percentage rectifier efficiency

$$= \frac{40.6}{1 + R_f/R_L} \tag{3.24}$$

From Eq. (3.23) we find that the rectifier efficiency increases as the ratio R_f/R_L reduces. Theoretically the maximum value of rectifier efficiency of a halfwave rectifier is 40.6% corresponding to value of R_f/R_L equal to zero.

3.3.6 Ripple Factor

The ripple factor is defined as the ratio of the effective value of the a.c. component of voltage or current to the direct or average value.

Thus the ripple factor gives an idea about the waviness of the rectified voltage. For a voltage or current consisting of a d.c. component, the fundamental and a number of harmonic terms, the effective value is given by the square root of the sum of the squares of the effective values of those components. Thus the effective value of current is given by,

$$\begin{aligned} I &= \sqrt{I_{dc}^2 + I_1^{\,2} + I_2^{\,2} + I_3^{\,2} + \ldots} \\ &= \sqrt{I_{dc}^2 + I_{ac}^2} \end{aligned} \tag{3.25}$$

where I_1, I_2 etc., are the r.m.s. values of the fundamental, second harmonic etc., terms and I_{ac}^2 is the sum of the squares of the r.m.s. values of the a.c. components.

Then the ripple factor γ is given by

$$\gamma = \frac{I_{ac}}{I_{dc}} = \frac{\sqrt{I^2 - I_{dc}^2}}{\sqrt{I_{dc}^2}}$$

$$= \sqrt{\left(\frac{I}{I_{dc}}\right)^2 - 1} \tag{3.26}$$

For a halfwave rectifier,

$$I = \frac{I_m}{2} \quad \text{and} \quad I_{dc} = \frac{I_m}{\pi}$$

Hence ripple factor

$$\gamma = \sqrt{\left(\frac{\pi}{2}\right)^2 - 1} = 1.21 \tag{3.27}$$

3.3.7 Regulation

By regulation is meant the variation of d.c. output voltage with change in d.c. load current. Thus percentage regulation is defined as,

Percentage regulation

$$= \frac{V_{nL} - V_L}{V_L} \times 100 \tag{3.28}$$

where V_{nL} is the no load d.c. output voltage i.e. with zero current and V_L is the d.c. output voltage with normal load current.

An ideal power supply should have zero percentage regulation i.e. the output voltage should be independent of the load current.

For halfwave rectifier, combining Eqs. (3.6) and (3.7), we get

$$V_{dc} = \frac{V_m R_L}{\pi (R_f + R_L)}$$

$$= \frac{V_m}{\pi}\left[1 - \frac{R_f}{R_f + R_L}\right]$$

$$= \frac{V_m}{\pi} - I_{dc} \cdot R_f \tag{3.29}$$

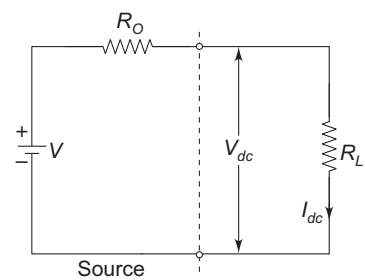


Fig. 3.4 *Thevenin's model for a rectifier*

Equation (3.29) suggests that the halfwave rectifier functions as if it were constant voltage source $V = V_m/\pi$ in series with an internal resistance (output resistance $R_o = R_f$) as shown in Fig. 3.4 From this model, we find that V_{dc} equals V_m/π with no load ($I_{dc} = 0$) and that the d.c.

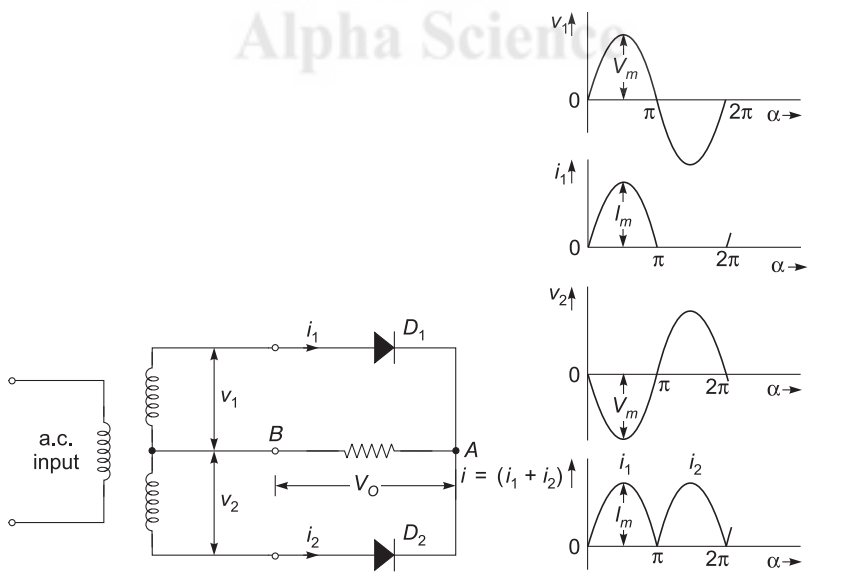
output voltage decreases linearly with the increase of d.c. output current I_{dc}. In a physical circuit, the resistance R_o of the transformer secondary comes in series with the diode so that in Eq. (3.29), R_o should be added to R_f.

Now, according to Thevenin's theorem, any two terminal linear network may be replaced by a voltage generator equal to the open circuit voltage between the terminals with a series resistance equal to the output impedance at the port. Obviously this Fig. 3.4 represents Thevenin's model for the halfwave rectifier with $V = V_m/\pi$ and $R_o = R_f$.

3.4 FULLWAVE RECTIFIER

Figure 3.5(a) gives the basic circuit of a fullwave rectifier. Making use of the centretapped secondary of the transformer, two equal and opposite voltages v_1 and v_2 get available. During one half-cycle, diode D_1 conducts and the input voltage v_1 causes a current i_1 though diode D_1 and load resistor R_L. Diode D_2 remains non-conducting. During the other half-cycle, diode D_1 is non-conducting but diode D_2 conducts and voltage v_2 sends a current i_2 through diode D_2 and load resistor R_L. Figure 3.5(b) shows waveforms of these voltages v_1 and v_2 and currents i_1 and i_2. The total current flowing through R_L is the sum of the individual currents i_1 and i_2. Thus,

$$i = i_1 + i_2 \tag{3.30a}$$



(a) Basic circuit diagram

(b) Waveforms of applied a.c. voltage and output currents

Fig. 3.5 *Fullwave rectifier*

Let the voltage v_1 be given by,

$$v_1 = V_m \sin \omega t = V_m \sin \alpha \tag{3.30b}$$

Let the peak voltage V_m be very large in comparison with the cutin voltage V_o. Then we may assume $V_o = 0$. Then each of the diodes may be idealized with resistance R_f in the ON condition and open circuit ($R_r = \infty$) in the OFF condition. The current i_1 through diode D_1 and the load resistor R_L is as usual given by,

$$v_1 \equiv I_m \sin \alpha \qquad \text{for } 0 \le \alpha \le \pi \tag{3.31}$$

$$i_1 = 0 \qquad \text{for } \pi \le \alpha \le 2\pi \tag{3.32}$$

where
$$I_m = \frac{V_m}{R_f + R_L} \tag{3.33}$$

Similarly, i_2 is given by

$$i_2 = 0, \qquad \text{for } 0 \le \alpha \le \pi \tag{3.34}$$

$$i_2 = I_m \sin \alpha \qquad \text{for } \pi \le \alpha \le 2\pi \tag{3.35}$$

3.4.1 D.C. or Average Current

Since the current is of the same form in the two halves of the a.c. cycle,

$$I_{dc} = \frac{1}{\pi} \int_0^{\pi} i \, d\alpha$$

$$= \frac{1}{\pi} \int_0^{\pi} I_m \sin \alpha \, d\alpha = \frac{2I_m}{\pi} \tag{3.36}$$

3.4.2 D.C. Output Voltage

$$V_{dc} = I_{dc} \, R_L = \frac{2I_m \, R_L}{\pi} \tag{3.37}$$

We note that this d.c. outout voltage is twice that in the halfwave rectifier.

3.4.3 RMS Current

$$I_{\text{rms}} = \sqrt{\frac{1}{\pi} \int_0^{\pi} i^2 \, d\alpha} = \sqrt{\frac{I_m^2}{\pi} \int_0^{\pi} \sin \alpha \, d\alpha} = \frac{I_m}{\pi} \tag{3.38}$$

3.4.4 RMS Voltage across the Load

$$V_{\text{rms}} = I_{\text{rms}} \, . \, R_L = \frac{I_m \, R_L}{\pi} \tag{3.39}$$

3.4.5 Rectifier Efficiency

D.C. output power

$$P_{dc} = I_{dc}^2 \, . \, R_L = \frac{4I_m^{\,2}}{\pi^2} \, R_L \tag{3.40}$$

A.C. input power

$$P_{in} = I_{\text{rms}}^2 (R_f + R_L) = \frac{I_m^2}{2} (R_f + R_L) \tag{3.41}$$

Hence rectifier efficiency is given by

$$\eta = \frac{P_{dc}}{P_{in}} = \frac{\dfrac{4I_m^2}{\pi^2} R_L}{\dfrac{I_m^2}{2} (R_f + R_L)}$$

$$= \frac{8}{\pi^2} \cdot \frac{R_L}{R_f + R_L} = \frac{0.812}{1 + \dfrac{R_f}{R_L}} \tag{3.42}$$

$$\text{Percentage efficiency} = \frac{81.2}{1 + \dfrac{R_f}{R_L}} \tag{3.43}$$

We note that fullwave rectifier has efficiency twice that of halfwave rectifier.

3.4.6 Frequency Components of Fullwave Rectifier

Fourier analysis of the waveform of current i_1 yields,

$$i_1 = I_m \left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{3\pi} \cos 2\omega t - \frac{2}{15\pi} \cos 4\omega t - \ldots \right] \tag{3.44}$$

Current i_2 is 180° out of phase with i_1 and hence is given by,

$$I_2 = I_m \left[\frac{1}{\pi} - \frac{1}{2} \sin \omega t - \frac{2}{3\pi} \cos 2\omega t - \frac{2}{15\pi} \cos 4\omega t - \ldots \right] \tag{3.45}$$

Hence the total current i, being the sum of i_1 and i_2 is given by,

$$i = I_m \left[\frac{2}{\pi} - \frac{4}{3\pi} \cos 2\omega t - \frac{4}{15\pi} \cos 4\omega t - \ldots \right] \tag{3.46}$$

It can be seen from Eq. (3.46) that the lowest frequency term in the rectified output of a fullwave rectifier is twice the supply frequency. This is twice the lowest frequency in the ripple voltage of halfwave rectifier. Hence in a fullwave rectifier, it is comparatively easy to filter the ripple voltage components from the rectified output voltage.

3.4.7 Ripple Factor

The form factor of the rectified output voltage of a fullwave rectifier is given by,

$$F = \frac{I_{\text{rms}}}{I_{dc}} = \frac{I_m/\sqrt{2}}{2I_m/\pi} = 1.11$$

Hence ripple factor

$$v = \sqrt{(1.11)^2 - 1} = 0.48 \tag{3.47}$$

3.4.8 Regulation

From Eqs. (3.33) and (3.37), we get

$$V_{dc} = \frac{2I_m R_L}{\pi} = \frac{2V_m R_L}{\pi (R_f + R_L)}$$

$$= \frac{2V_m}{\pi}\left[1 - \frac{R_f}{R_f + R_L}\right]$$

or

$$V_{dc} = \frac{2V_m}{\pi} - I_{dc} R_f \tag{3.48}$$

Equation (3.48) leads to Thevenin's model of Fig. 3.4 with

$$V = \frac{2V_m}{\pi} \quad \text{and} \quad R_o = R_f.$$

3.5 PEAK INVERSE VOLTAGE

Each rectifier circuit has a certain maximum voltage which it can withstand. This voltage is called the *peak inverse voltage* since it occurs during that period of the cycle when the diode is non-conducting. From Fig. 3.2, it is evident that the halfwave rectifier has peak inverse voltage equal to V_m. It can be shown that fullwave rectifier has peak inverse voltage twice as large, i.e., $2V_m$. Thus consider the time instant when v_1 and v_2 are at their peak values and diode D_1 is conducting and diode D_2 is non-conducting. On applying KVL along the outside loop and neglecting the small voltage drop across D_1, we get $2V_m$ as the peak inverse voltage across diode D_1. This is true irrespective of the nature of the load.

3.6 COMPARISON OF HALFWAVE AND FULLWAVE RECTIFIER

The following points are of significance in this connection:

(i) In a halfwave rectifier, current flows through the secondary of the power transformer in the same direction always whereas in a fullwave rectifier equal currents flow through the two halves of the centre-tapped secondary of the power transformer in opposite directions. The d.c. saturation of the core of the transformer is thus avoided. The saturation of the core increases the magnetizing current and hysteresis losses and produces harmonics in the secondary output.

(ii) Halfwave rectifier has the chief advantage of simplicity and low cost compared with a fullwave rectifier.

(iii) Halfwave rectifier gives high ripple amplitude and low ripple frequency which necessitates the use of comparative expensive smoothing filter.

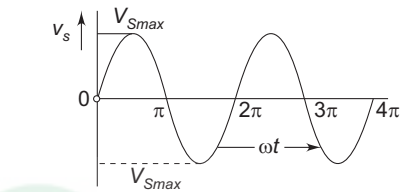
(iv) Halfwave rectifier has low output voltage and low transformer efficiency. Low transformer efficiency results because (a) the ripple voltages and currents are high and (b) the current flows through the secondary in the same direction and saturates the core.

(v) Efficiency of fullwave rectifier is twice that of a halfwave rectifier.

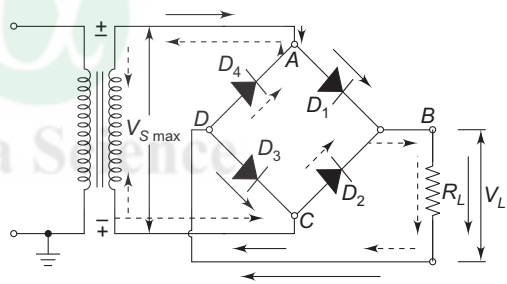
3.7 BRIDGE RECTIFIER

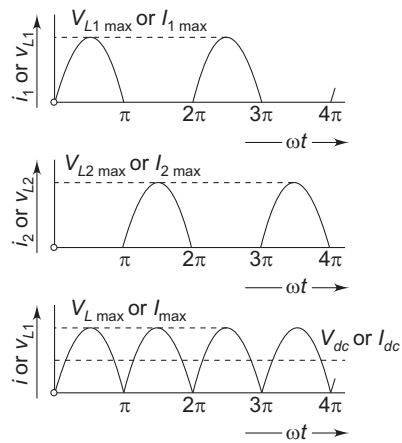
In the bridge circuit four diodes are connected in the form of a Wheatstone bridge, two diametrically opposite junctions of the bridge are connected to the secondary of a transformer and the other two are connected to the load, as shown in Fig. 3.6.

When the upper end of the transformer secondary winding is positive, say during first half cycles of the input supply, diodes D_1 and D_3 are forward biased and current flows through arm AB, enters the load at positive terminal, leaves the load at negative terminal, and returns back flowing through arm DC. During this half of each input cycle, the diodes D_2 and D_4 are reverse biased and so the current is not allowed to flow in arms AD and BC. The flow of current is indicated by solid arrows in the figure. In the second half of the input cycle the lower end of ac supply becomes positive, diodes D_2 and D_4 become forward biased and current flows through arm



Input voltage waveform





Rectified output voltage/current waveforms

Fig. 3.6 *Bridge rectifier*

CB, enters the load at the positive terminal, leaves the load at negative terminal and returns back flowing through arm *DA*. Flow of current has been shown by dotted arrows in the figure. Thus the direction of flow of current through the load resistance R_L remains the same during both half cycles of the input supply voltage.

Peak Inverse Voltage

Figure 3.6 shows a bridge rectifier circuit. Let us consider the instant the secondary voltage attains its positive peak value $V_{S\,\max}$. Now diodes D_1 and D_3 are conducting whereas diodes D_2 and D_4 are non-conducting (being reverse biased). The conducting diodes D_1 and D_3 have almost zero resistance (i.e., zero voltage drop across them). Point *B* has the same potential as point *A* and similarly, point *D* has the same potential as point *C*. The entire voltage of the transformr secondary winding ($V_{S\,\max}$) is developed across the load resistance R_L. The same voltage i.e., $V_{S\,\max}$ acts across each of the non-conducting diodes D_2 and D_4. Thus

$$\text{PIV} = V_{S\,\max} \tag{3.49}$$

Circuit Analysis

As compared to fullwave rectifier circuit, in bridge rectifier circuit two diodes conduct during each half-cycle and forward – resistance becomes double ($2R_F$) and in a bridge rectifier circuit, $V_{s\max}$ is the maximum voltage across the transformer secondary winding across each half of the secondary winding in case of centre-tap fullwave rectifier.

1. Peak Current

Instantaneous value of voltage applied to the rectifier is given as

$$v_s = V_{S\,\max} \sin \omega t \tag{3.50}$$

If the diode is assumed to have a forward resistance of R_F ohms and reverse resistance equal to infinity, then current flowing through the load resistance is given as

$$i_1 = I_{\max} \sin \omega t$$

and $i_2 = 0$ for first half-cycle (3.51a)

and $i_1 = 0$

and $i_2 = I_{\max} \sin \omega t$ for second half-cycle (3.51b)

The total current flowing through the load resistance R_L, being the sum of currents i_1 and i_2, is given as

$$i = i_1 + i_2$$
$$= I_{\max} \sin \omega t \qquad \text{for the whole cycle} \qquad (3.52)$$

where peak value of current flowing through the load resistance R_L is given as

$$I_{\max} = \frac{V_{S\,\max}}{R_F + R_L} \qquad \text{in case of centre-tap rectifier} \qquad (3.53)$$

and $$I_{\max} = \frac{V_{S\,\max}}{2R_F + R_L} \qquad \text{in case of bridge rectifier} \qquad (3.54)$$

2. Output Current

Since the current is the same through the load resistance R_L in the two halves of the a.c. cycle, magnitude of direct current I_{dc}, which is equal to the average value of alternating current, can be obtained by integrating the current i_1 between 0 and π or current i_2 between π and 2π.

So,
$$I_{dc} = \frac{1}{\pi}\int_0^{\pi} i_1 \, d(\omega t)$$
$$= \frac{1}{\pi}\int_0^{\pi} I_{\max} \sin \omega t \, d(\omega t) = \frac{2I_{\max}}{\pi} \qquad (3.55)$$

3. D.C. Output Voltage

Average or d.c. value of voltage across the load is given as

$$V_{dc} = I_{dc} R_L = \frac{2}{\pi} I_{\max} R_L \qquad (3.56)$$

4. RMS Value of Current

RMS or effective value of current flowing through the load resistance R_L is given as

$$I_{\text{rms}}^2 = \frac{1}{\pi}\int_0^{\pi} i_1^2 \, d(\omega t)$$
$$= \frac{1}{\pi}\int_0^{\pi} I_{\max}^2 \sin^2 \omega t \, d(\omega t) = \frac{I_{\max}^2}{2}$$

or
$$I_{\text{rms}} = \frac{I_{\max}}{\sqrt{2}} \qquad (3.57)$$

5. RMS Value of Output Voltage

RMS value of voltage across the load is given as

$$V_{L\,\text{rms}} = I_{\text{rms}} R_L = \frac{I_{\max}}{\sqrt{2}} R_L \qquad (3.58)$$

6. Form Factor and Peak Factor

Form factor,

$$K_f = \frac{\text{RMS value}}{\text{Average value}}$$

$$= \frac{I_{\text{rms}}}{I_{dc}} = \frac{I_{\max}/\sqrt{2}}{2I_{\max}/\pi}$$

$$= \frac{\pi}{2\sqrt{2}} = 1.11 \tag{3.59}$$

Peak factor,

$$K_p = \frac{\text{Peak value}}{\text{RMS value}}$$

$$= \frac{I_{\max}}{I_{\max}/\sqrt{2}} = \sqrt{2} \tag{3.60}$$

7. Output Frequency

With a halfwave rectifier, the output frequency equals the input frequency. But the frequency of the fullwave signal is double the input frequency. A fullwave output has twice as many cycles as the sinewave input has. The fullwave rectifier inverts each negative half-cycle, so that we get the number of positive half-cycles. The effect is to double the frequency. Thus, for fullwave rectifier

$$f_{out} = 2f_{in} \tag{3.61}$$

8. Rectification Efficiency

Power delivered to load,

$$P_{dc} = I_{dc}^2 R_L = \left(\frac{2}{\pi} I_{\max}\right)^2 R_L = \frac{4}{\pi^2} I_{\max}^2 R_L \tag{3.62}$$

A.C. input power,

$$P_{ac} = I_{\text{rms}}^2 (R_L + R_F) = \frac{I_{\max}^2}{2} (R_L + R_F) \tag{3.63}$$

Rectification efficiency,

$$\eta = \frac{P_{dc}}{P_{ac}} = \frac{\frac{4}{\pi^2} I_{\max}^2 R_L}{\frac{1}{2} I_{\max}^2 (R_L + R_F)}$$

$$= \frac{8}{\pi^2} \frac{1}{\left(1 + \frac{R_F}{R_L}\right)} = \frac{0.812}{1 + \frac{R_F}{R_L}} \tag{3.64}$$

In case of bridge rectifier, rectification efficiency is given as

$$\eta = \frac{0.812}{1 + \dfrac{2R_F}{R_L}} \tag{3.65}$$

9. Ripple Factor

Form factor of the rectified output voltage of a fullwave rectifier is given as

$$K_f = \frac{I_{\text{rms}}}{I_{\text{av}}} = \frac{I_{\text{rms}}\sqrt{2}}{2\, I_{\max}/\pi} = \frac{\pi}{2\sqrt{2}} = 1.11$$

Ripple factor is given as

$$\gamma = \sqrt{K_f^2 - 1} = \sqrt{(1.11)^2 - 1} = 0.482 \tag{3.66}$$

10. Regulation

From Eq. (3.56)

$$V_{dc} = \frac{2}{\pi} R_L = \frac{2V_{S\max} R_L}{\pi (R_F + R_L)}$$

$$= \frac{2V_{S\max}}{\pi}\left[1 - \frac{R_F}{R_F + R_L}\right]$$

$$= \frac{2V_{S\max}}{\pi} - I_{\text{dc}} R_F \tag{3.67}$$

In case of a bridge rectifier,

$$V_{dc} = \frac{2V_{S\max}}{\pi} - 2I_{\text{dc}} R_F \tag{3.68}$$

Salient Features of a Bridge Rectifier

(i) Current in both the primary and secondary of the supply transformer flows for the entire a.c. cycle and hence for a given power output, power transformer of a small size may be used in comparison with that in a full-wave rectifier.

(ii) No centre-tap is needed in the transformer secondary.

(iii) Since two diodes are present in series in each conduction path, the peak inverse voltage is shared equally by the two diodes. Hence bridge rectifier is eminently suited for high voltage applications.

Bridge rectifier, however, has the limitation that twice the number of rectifier elements are needed compared with a fullwave rectifier.

3.8 FILTER CIRCUITS

The output from any of the rectifier circuits just discussed is not purely d.c. but also has some a.c. components, called *ripples*, along it. The ripples are maximum

in the single phase halfwave rectifier and being reduced in fullwave rectifier and being reduced further with the increase in number of phases. Such supply is not useful for driving sophisticated electronic devices/circuits. Of course, for a circuit such as battery charger the pulsating nature of supply available from a rectifier is no great detriment as long as the d.c. level provided results in charging of battery. But for supply circuits to radio or tape recorder the pulsating d.c. results in 50 (or 100) Hz signal appearing in the output, thereby making the operation of the overall circuit poor. For such applications, as well as for many more, the output d.c. developed will have to be much steady or smoother than that of the pulsating dc obtained directly from halfwave or fullwave rectifier circuits.

Hence, it becomes essential to reduce the ripples from the pulsating d.c. supply available from rectifier circuits to the minimum. This is achieved by using a filter or smoothing circuit which removes (or filters out) the a.c. components and allows only the d.c. component to reach the load. Obviously, a filter circuit should be placed between the rectifier and the load, as shown in Fig. 3.7.

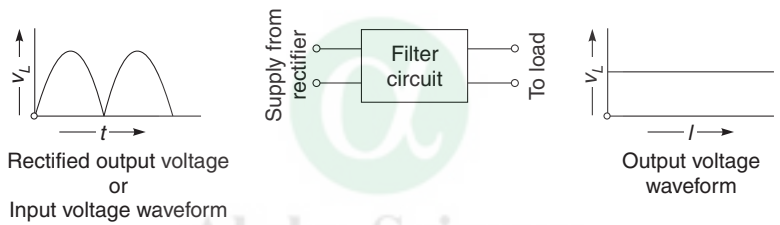


Fig. 3.7

A filter circuit is a device that converts pulsating output of a rectifier into steady d.c. level.

A filter circuit is generally a combination of inductors L and capacitors C. The filtering action of L and C depends upon the facts that an inductor allows d.c. only and capacitor allows a.c. only to pass. So, a suitable L and C network can effectively filter out (or remove) the a.c. components from the rectified output.

Commonly used filter circuits are (a) series inductor filter (b) shunt capacitor filter (c) choke input filter (d) capacitor input or π filter and (e) *R-C* filter. These will be discussed here in brief.

3.9 SERIES INDUCTOR FILTER

In this arrangement a high value inductor or choke L is connected in series with the rectifier element and the load, as illustrated in Fig. 3.8 (a). The filtering action of an inductor filter depends upon its property of opposing any change in the current flowing through it. When the output current of the rectifier increases above a certain value, energy is stored in it in the form of magnetic field and this energy is given up when the output current falls below the average value. Thus

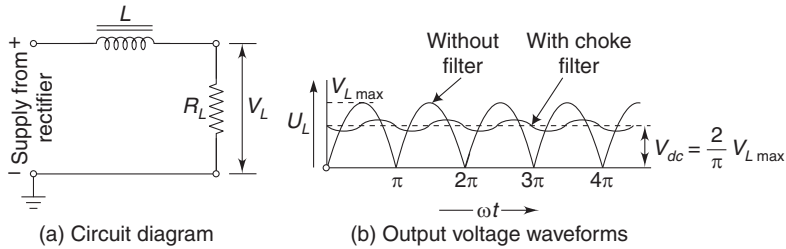


Fig. 3.8 *Fullwave rectifier with series inductor filter*

by placing a choke coil in series with the rectifier output and load, any sudden change in current that might have occurred in the circuit without an inductor is smoothed out by the presence of the inductor *L*.

The function of the inductor filter may be viewed in terms of impedances. The choke offers a high impedance to the ac components but offers almost zero resistance to the desired d.c. components. Thus ripples are removed to a large extent. Nature of the output voltage without filter and with choke filter are shown in Fig. 3.8 (b).

For d.c. (zero frequency), the choke resistance R_C in series with the load resistance R_L forms a voltage divider and d.c. voltage across the load is given as

$$V_L = V_{dc} \times \frac{R_L}{R_C + R_L} = \frac{V_{dc}}{1 + \frac{R_C}{R_L}} \tag{3.69}$$

where V_{dc} is d.c. voltage output from a fullwave rectifier. Usually choke coil resistance R_C is much smaller than R_L and, therefore, almost entire of the d.c. voltage is available across the load resistance R_L.

Fourier series for the rectified output voltage is given as

$$V_L = V_{L\,max} \left[\frac{2}{\pi} - \frac{4}{3\pi} \cos 2\,\omega t - \frac{4}{15\pi} \cos 4\,\omega t - \ldots \right] \tag{3.70}$$

Since the reactance of inductor increases with the increase in frequency, better filtering of the higher harmonic components take place, so effect of third and higher harmonic voltages can be neglected.

As obvious form Eq. (3.69), if choke coil resistance R_C is negligible in comparison to load resistance R_L, then the entire dc component of rectifier output is available across R_L and is equal to $\frac{2}{\pi} V_{L\,max}$. The a.c. voltage partly drops across X_L and partly over R_L. Since choke and R_L are connected in series, the maximum value of voltage drop across load resistance R_L due to second harmonic voltage $\left(\frac{4}{3\pi} V_{L\,max}\right)$

$$V_{ac\ \max} = \frac{4}{3\pi} V_{L\ \max} \cdot \frac{R_L}{\sqrt{R_L^2 + X_L^2}}$$

RMS value of this a.c. voltage drop across load resistance R_L is given as

$$V_{ac\ \text{rms}} = \frac{4}{3\pi\sqrt{2}} \cdot V_{L\ \max} \frac{R_L}{\sqrt{R_L^2 + X_L^2}}$$

Ripple factor,
$$\gamma = \frac{V_{ac\ \text{rms}}}{V_{dc}} = \frac{\frac{4}{3\pi\sqrt{2}} \cdot V_{L\ \max} \frac{R_L}{\sqrt{R_L^2 + X_L^2}}}{\frac{2}{\pi} V_{L\ \max}}$$

$$= \frac{\sqrt{2}\ R_L}{3\sqrt{R_L^2 + X_L^2}} = \frac{\sqrt{2}}{3\sqrt{1 + \frac{X_L^2}{R_L^2}}}$$

Since $X_L = 2\omega L$

$\therefore$ Frequency is doubled in fullwave rectifier

$\therefore$
$$\gamma = \frac{\sqrt{2}}{3\sqrt{1 + \frac{4\omega^2 L^2}{R_L^2}}} \tag{3.71}$$

If $2\omega L >> R_L$.
$$\gamma = \frac{\sqrt{2}\ R_L}{3 \times 2\ \omega L} = \frac{R_L}{3\sqrt{2}\ \omega L} \tag{3.72}$$

It is seen that ripple decreases with the decrease in R_L (or increase in load current I) so inductor filter is more efficient for high load current or small load resistance R_L.

In case the load resistance is infinite, i.e., the output is an open circuit, then ripple factor from Eq. (3.71)

$$\gamma = \frac{\sqrt{2}}{3} = 0.471 \tag{3.73}$$

This is slightly less than the value of 0.482. The difference being attributable to the omission of higher harmonics.

3.10 SHUNT CAPACITOR FILTER

This is the most simple form of the filter circuit and in this arrangement a high value capacitor C is placed directly across the output terminals, as shown in

Fig. 3.9(a). During the conduction period it gets charged and stores up energy in the electrostatic field and discharges through the load resistance R_L delivering energy to it during non-conduction period. Through this process, the time duration during which current flows through the load resistance gets prolonged and ac components or ripples get considerably reduced. It is to be noted here that the capacitor C gets charged to the peak value of input voltage quickly because charging time constant is almost zero. It is so because there is no resistance (except the negligible forward resistance of diode) in the charging path. But the discharging time is quite large (roughly 100 times more than the charging time depending upon the value of R_L) because it discharges through load resistance R_L.

The function of the capacitor filter may be viewed in terms of impedances. The large value capacitor C offers a low impedance shunt path to the ac components or ripples but offers high impedance to the dc component. Thus ripples get bypassed through capacitor C and only dc component flows through the load resistance R_L.

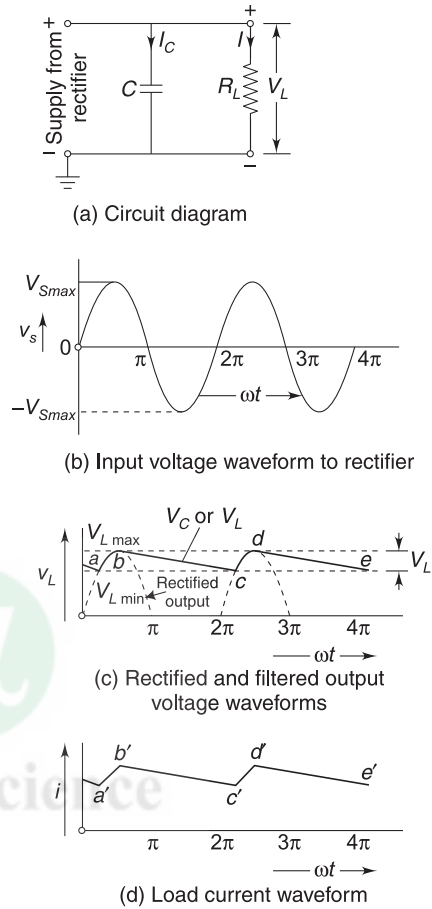


Fig. 3.9 *Halfwave rectifier with shunt capacitor filter*

Capacitor filter is very popular because of its low cost, small size, light weight and good characteristics. It is quite useful for load up to 50 mA as in transistor radio, battery eliminators.

3.10.1 Halfwave Rectifier with Shunt Capacitor Filter

The waveforms of ac input voltage, rectified and filtered output voltages and load current are shown in Figs. 3.9(b), (c) and (d) respectively.

During the positive half cycle of the ac input, the diode of the rectifier is forward biased and so it conducts. This quickly charges the capacitor C to peak value of the supply voltage $V_{S\,\max}$ because of almost zero charging time constant. This is shown by point b in Fig. 3.9(c). After being fully charged, the capacitor

holds the charge till input ac supply to the rectifier goes negative. During the negative half cycle, the diode gets reverse biased and so stops conduction. So the capacitor C discharges through load resistance R_L and losses charge. Voltage across R_L (v_L) or across C (v_C), both being equal, decreases exponentially with time constant CR_L along the curve bc, as illustrated in Fig. 3.9(c). Because of large discharge time constant CR_L, the capacitor does not have sufficient time to discharge appreciably. Due to this fact the capacitor maintains a sufficiently large voltage across R_L, even during the negative half cycle of the input supply. During the next positive half cycle, when the rectified voltage exceeds the capacitor voltage v_C represented by point c in Fig. 3.9(c), the capacitor C again gets charged quickly to $V_{S\,\max}$ (or $V_{L\,\max}$) as represented by point d in the figure. This process of charging and discharging of capacitor C is repeated for each cycle of input supply voltage. It is seen, from the figure, that nearly constant dc voltage appears across load resistance R_L at all times and also the dc component of the output voltage is increased considerably.

In case of a purely resistive load, the output current waveform is of the same nature as the waveform of output voltage. This is illustrated in Fig. 3.9(d). During the periods $a'\ b'$ and $c'\ d'$ the current is supplied by the diode and during periods $b'\ c'$ and $d'\ e'$ etc., by capacitor C.

Diode current flows during short intervals of time ab and cd etc. in Fig. 3.9(c). During these time intervals, diode output voltage exceeds the capacitor (or load) voltage. Hence diode current is short-duration pulses (i.e. surging current). A small resistor, called the surge limiting resistor is, therefore, always connected in series with the diode to limit this surge current.

The noteworthy points about shunt capacitor filter are:

1. For a fixed-value filter capacitance larger the load resistance R_L larger will be discharge time constant CR_L and therefore, lower the ripples and more the output voltage. On the other hand lower the load resistance (or more the load current), lower will be the discharge time constant and therefore, more the ripples and lower the output voltage.
2. Similarly smaller the filter capacitor, the less charge it can hold and more it will discharge. Thus the peak-to-peak value of the ripple will increase, and the average d.c. level will decrease. Larger the filter capacitor, the more charge it can hold and the less it will discharge. Hence the peak-to-peak value of the ripple will be less, and the average dc level will increase. But, the maximum value of the capacitance that can be employed is limited by another factor. The larger the capacitance value, the greater is the current required to charge the capacitor to a given voltage. The maximum current that can be handled by a diode is limited by the figure quoted by the manufacturer. Thus the maximum value of the capacitance, that can be used in the shunt filter capacitor is limited.

Approximate Analysis Total change in output voltage is equal to V_r, the ripple component of output voltage. The average or d.c. value of output voltage, V_{dc} is almost between the peak value $V_{L\,\max}$ and the minimum value given by points *b* and *c* respectively in Fig. 3.9(c). Thus

$$V_{dc} = V_{L\,\max} - \frac{V_r}{2} \tag{3.74}$$

The total charge lost during non-conduction (or discharge) duration T_2 through load is given as

$$Q_{\text{discharge}} = I_{dc}\, T_2 \tag{3.75}$$

This charge is replenished during time interval T_1, in which voltage across the capacitor increases by V_r volts. So charge gained by capacitor,

$$Q_{\text{charge}} = CV_r \tag{3.76}$$

In steady-state,

$$Q_{\text{charge}} = Q_{\text{discharge}}$$

or $$CV_r = I_{dc}\, T_2$$

or $$V_r = \frac{I_{dc}\, T_2}{C}$$

Assuming $T_1 << T_2$,

$$T_2 = T = \frac{1}{f}$$

So, $$V_r = \frac{I_{dc}}{f\,C} \tag{3.77}$$

Hence from Eq. (3.74)

$$V_{dc} = V_{L\,\max} - \frac{I_{dc}}{2f\,C} \tag{3.78}$$

From Eq. (3.77), it is revealed that ripple voltage varies directly with the load current I_{dc} and inversely with the capacitance *C*.

Ripple Factor The r.m.s. value of the ripple component of almost triangular wave is independent of the slope or the length of the almost straight lines *bc* and *cd* but depends only on the peak value of V_r. The r.m.s. value is given as

$$V_{ac\,\text{rms}} = \frac{V_r}{2\sqrt{3}}$$

Hence ripple factor,

$$\gamma = \frac{V_{ac\,\text{rms}}}{V_{dc}} = \frac{V_r}{2\sqrt{3}\; I_{dc}\, R_L}$$

$$= \frac{I_{dc}}{2\sqrt{3}\, I_{dc}\, R_L f\, C}$$

$$= \frac{1}{2\sqrt{3}\, f\, CR_L} \tag{3.79}$$

3.10.2 Fullwave Rectifier with Shunt Capacitor Filter

The filtering action of shunt capacitor filter on a fullwave rectifier is shown in Fig. 3.10. In this case capacitance C discharges twice during one cycle. Because both the diodes conduct, non-conduction period has reduced. The result is that ripple voltage V_r has been reduced to half and V_{dc} has been increased relative to halfwave rectifier.

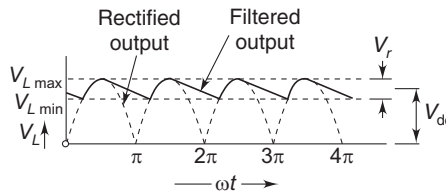


Fig. 3.10 *Rectified and filtered output voltage waveform from fullwave rectifier with shunt filter capacitor*

In this case because

$$T_2 = \frac{T}{2} = 1/2\, f$$

So
$$V_r = \frac{I_{dc}}{2f\, C} \tag{3.80}$$

$$V_{dc} = V_{L\,\max} - \frac{V_r}{2} = V_{L\,\max} - \frac{I_{dc}}{4f\, C} \tag{3.81}$$

$$V_{ac\ \text{rms}} = \frac{V_r}{2\sqrt{3}}$$

So, ripple factor,

$$\gamma = \frac{V_{ac\ \text{rms}}}{V_{dc}} = \frac{V_r}{2\sqrt{3}\, I_{dc}\, R_L} = \frac{1}{4\sqrt{3}\, f\, CR_L} \tag{3.82}$$

Voltage regulation in this case is better than that in halfwave rectifier.

3.11 CHOKE-INPUT OR L-SECTION FILTER

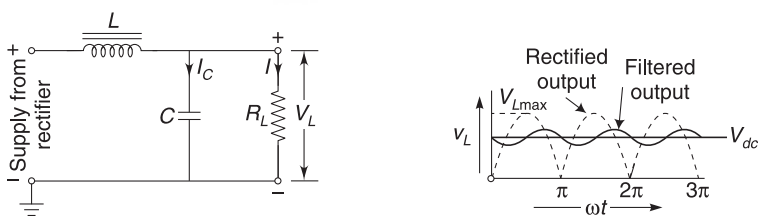
A simple series inductor reduces both the peak and effective values of the output current and output voltage. On the other hand simple shunt capacitor filter reduces the ripple voltage but increases the diode current. The diode may get damaged due to large current and at the same time it causes greater heating of supply transformer resulting in reduced efficiency.

In an inductor filter, ripple factor increases with the increase in load resistance R_L while in a capacitor filter it varies inversely with load resistance R_L.

From economical point of view also, neither series inductor nor shunt capacitor type filters are suitable.

Practical filter-circuits are derived by combining the voltage stabilizing action of shunt capacitor with the current smoothing action of series choke coil. By using combination of inductor and capacitor ripple factor can be lowered, diode current can be restricted and simultaneously ripple factor can be made almost independent of load resistance (or load current). Two types of most commonly used combination are choke-input or *L*-section filter and capacitor-input or π-filter. Choke-input filter is explained below.

Choke-input filter consists of a choke *L* connected in series with the rectifier and a capacitor *C* across the load, as shown in Fig. 3.11(a). This is also sometimes called the *L-section filter* because in this arrangement inductor and capacitor are connected, as an inverted *L*. In Fig. 3.11(a) only one filter section is shown, but several identical sections are often employed to improve the smoothing action. The choke *L* on the input side of the filter readily allows dc to pass but opposes the flow of ac components because its dc resistance is negligibly small but ac impedance is large. Any fluctuation that remains in the current even after passing through the choke are largely bypassed around the load by the shunt capacitor because X_C is much smaller than R_L. Ripples can be reduced effectively by making X_L greater than X_C at ripple frequency. However, a small ripple still remains in the filtered output and this is considered negligible if it is less than 1%. The rectified and filtered output voltage waveforms from a fullwave rectifier with choke-input filter are shown in Fig. 3.11(b).



(a) Circuit diagram

(b) Rectified and filtered output voltage waveform fullwave rectifier with choke-Input filter

Fig. 3.11

Regulation The output voltage of the rectifier is given as

$$v_L = \frac{2}{\pi} V_{L\,\max} - \frac{4}{3\pi} V_{L\,\max} \cos 2\,\omega t \tag{3.83}$$

(refer to Eq. (3.70) Section 3.9.)

Considering that the inductor has negligibly small (almost zero) resistance

$$V_{dc} = \frac{2}{\pi} V_{L\,\max} \tag{3.84}$$

If R_C is the choke resistance, then

$$V_{dc} = \frac{2}{\pi} V_{L\,\max} - I_{dc}\, R_C \tag{3.85}$$

Ripple Factor The main object of the filter is to suppress the harmonic components in the system and for this it is necessary that reactance of the choke coil X_L is made much greater than combined parallel impedance of the capacitor C and load resistor R_L. The parallel impedance of capacitor C and load resistor R_L can be made small by making the reactance of capacitor, X_C much smaller than load resistor R_L. Very little error is caused if it is assumed that the entire alternating current is flowing through the capacitor and none through the load resistor R_L. Under these conditions the net impedance across the input terminals of the filter circuit is approximately $X_L = 2\omega L$, the reactance of the inductor at the second harmonic frequency. A.C. current through the circuit is given as

$$I_{ac\ \text{rms}} = \frac{4V_{L\,\max}}{3\pi\sqrt{2}\,X_L} = \frac{\sqrt{2}}{3} \cdot \frac{V_{dc}}{X_L} \tag{3.86}$$

$$\therefore \qquad V_{dc} = \frac{2}{\pi} V_{L\,\max}$$

The a.c. voltage across the load (the ripple voltage) is equal to the voltage across the capacitor so

$$V_{ac\ \text{rms}} = I_{ac\ \text{rms}}\, X_C = \frac{\sqrt{2}}{3} V_{dc} \frac{X_C}{X_L} \tag{3.87}$$

where $X_C = \dfrac{1}{2\omega C}$, the reactance of the capacitor at the second harmonic frequency.

Ripple factor,
$$\gamma = \frac{V_{ac\ \text{rms}}}{V_{dc}}$$

$$= \frac{\sqrt{2}}{3} \frac{X_C}{X_L} = \frac{\sqrt{2}}{3} \cdot \frac{1}{2\omega C} \cdot \frac{1}{2\omega L}$$

$$= \frac{1}{6\sqrt{2}\,\omega^2 LC} \tag{3.88}$$

Critical Inductance During the above discussion it has been assumed that a current flows through the circuit at all times. In the absence of inductor, current flows through the diode circuit for a small portion of the cycle, and the capacitor is charged to the peak transformer secondary voltage in each cycle (neglecting diode forward and transformer resistance). When a small inductance is inverted in the circuit, the diode current will exist for a longer duration but cutout may still occur. With the continuous increase in inductance, a value is reached for which

diode current exists for the whole cycle. This value of inductance is referred to as the *critical inductance* L_C.

Current flowing through the load is made up of two components I_C given as $\frac{V_{dc}}{R_L}$ and a.c. component of peak value $\frac{4V_{L\,\max}}{3\pi X_L}$.

For continuous flow of current through diode it is necessary that I_{dc} should always exceed the negative peak value of a.c. component so

$$\frac{V_{dc}}{R_L} \geq \frac{4V_{L\,\max}}{3\pi X_L}$$

or $$X_L \geq \frac{2R_L}{3} \qquad \because V_{dc} = \frac{3V_{L\,\max}}{\pi}$$

or $$L_C \geq \frac{R_L}{3\omega} \tag{3.89}$$

The *L-C* filter was filter popular at one time. Now, it is becoming obsolete in typical power supplies because of the size and cost of inductors. For low-voltage power supplies, the *L-C* filter has been replaced by IC voltage regulators, active filters that reduce ripple and hold the output d.c. voltage constant.

3.12 CAPACITOR-INPUT OR π-FILTER

Such a filter consists of a shunt capacitor C_1 at the input followed by an *L*-section filter formed by series inductor *L* and shunt capacitor C_2. This is also called the π-*filter* because the shape of the circuit diagram for this filter appears like Greek letter π (*pi*) (Fig. 3.12(a)). Since the rectifier feeds directly into the capacitor so it is also called *capacitor input filter*.

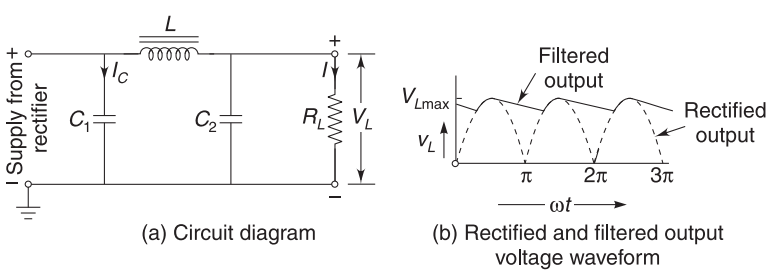


Fig. 3.12 *Fullwave rectifier with capacitor input filter*

As the rectifier output is fed directly into a capacitor C_1, such a filter can be used with a halfwave rectifier (series inductor and *L*-section filters cannot be used with halfwave rectifiers). Usually electrolytic capacitors are used even though their capacitances are large but they occupy minimum space. Usually both

capacitors C_1 and C_2 are enclosed in one metal container. The metal container serves as the common ground for the two capacitors.

A capacitor-input or π-filter is characterized by a high voltage output at low current drains. Such a filter is used, if, for a given transformers, higher voltage than that can be obtained from an *L*-section filter is required and if low ripple than that can be obtained from a shunt capacitor filter or *L*-section filter is desired. In this filter, the input capacitor C_1 is selected to offer very low reactance to the ripple frequency. Hence major part of filtering is accomplished by the input capacitor C_1. Most of the remaining ripple is removed by the *L*-section filter consisting of a choke *L* and capacitor C_2.

The action of this filter can best be understood by considering the action of *L*-section filter, formed by L and C_2, upon the triangular output voltage wave from the input capacitor C_1. The charging and discharging action of input capacitor C_1 has already been given in Section 3.10. The output voltage is roughly the same as across input capacitor C_1 less the d.c. voltage drop in inductor. The ripples contained in this output are reduced further by *L*-section filter. The output voltage of π-filter falls off rapidly with the increase in load-current and, therefore, the voltage regulation with this filter is very poor.

Ripple Factor From Eq. (3.89), the ripple voltage in case of a fullwave rectifier with shunt capacitor is given as

$$V_r = \frac{I_{dc}}{2f\,C_1} \qquad \because \text{here } C = C_1 \tag{3.90}$$

The r.m.s. value of second harmonic voltage is given as

$$V_{ac\ \text{rms}} = \frac{V_r}{\pi\sqrt{2}} = \frac{1}{\pi\sqrt{2}} \cdot \frac{I_{dc}}{2f\,C_1} = \sqrt{2}\ I_{dc}\ X_{C_1} \tag{3.91}$$

where $X_{C_1} = \dfrac{1}{2\omega C_1} = \dfrac{1}{4\pi f\,C_1}$, reactance of input capacitor C_1 at second harmonic frequency.

Now $V_{ac\ \text{rms}}$ is applied to *L*-section so the ripple voltage can be obtained by multiplying X_{C_2}/X_L i.e.,

$$V'_{ac\ \text{rms}} = V_{ac\ \text{rms}} \times \frac{X_{C_2}}{X_L} = \sqrt{2}\ I_{dc}\ X_{C_1} \times \frac{X_{C_2}}{X_L} \tag{3.92}$$

Now the ripple factor,

$$\gamma = \frac{V'_{ac\ \text{rms}}}{V_{dc}} = \frac{\sqrt{2}\ I_{dc}\ X_{C_1}\ X_{C_2}}{V_{dc}\ R_L\ X_L} = \frac{\sqrt{2}\ X_{C_1}\ X_{C_2}}{R_L\ X_L} \tag{3.93}$$

or

$$\gamma = \frac{\sqrt{2}}{R_L} \cdot \frac{1}{2\omega C_1} \cdot \frac{1}{2\omega C_2} \cdot \frac{1}{2\omega L}$$

$$= \frac{\sqrt{2}}{8\omega^3 C_1 C_2 L R_L} \tag{3.94}$$

3.13 SALIENT FEATURES OF L-SECTION AND Π-FILTERS

1. In π-filter the dc output voltage is much larger than that can be had from an *L*-section filter with the same input voltage.
2. In π-filter ripples are less in comparison to those in shunt capacitor or *L*-section filter. So smaller valued choke is required in a π-filter in comparison to that required in *L*-section filter.
3. In π-filter, the capacitor is to be charged to the peak value hence the rms current in supply transformer is larger as compared in case of *L*-section filter.
4. Voltage regulation in case of π-filter is very poor, as already mentioned. So π-filters are suitable for fixed loads whereas *L*-section filter can work satisfactorily with varying loads provided a minimum current is maintained.
5. In case of a π-filter PIV is larger than that in case of an *L*-section filter.

3.14 ZENER DIODES

The current that flows under reverse breakdown is not necessarily destructive if the amount of current is limited to a value such that the diode does not become overheated. By suitably controlling the amount of impurity and making the *P-N* junction free from surface imperfections, the voltage at which breakdown takes place can be made to occur very distinctly and sharply, as shown in Fig. 3.13(a). This type of diode is called a Zener diode and is represented by the symbol shown in Fig. 3.13(b). Although the Zener diode is used in the reverse biased condition, the symbol is the same as for rectifying diodes. That is it is normally operated

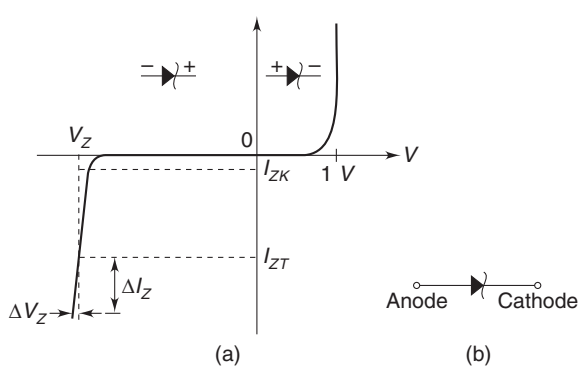


Fig. 3.13 *(a) Zener diode characteristics showing the sharp reverse breakdown; (b) symbol for the Zener diode*

with the node negative and the cathode positive to provide the characteristics shown in the third quadrant of Fig. 3.13(a).

Here it is seen that, once the diode has brokendown, very little change in voltage across the diode results from a change in current through the diode. Thus, a Zener diode of known breakdown voltage can be used as a reference voltage when operated in the reverse mode and can be used to regulate voltages.

3.15 ZENER DIODE SPECIFICATIONS

Zener Voltages

The manufacturers specify the value of break-down voltage known as the Zener voltage, V_Z, at some value of test current, I_{ZT}. This is on the linear portion of the reverse characteristic and corresponds to approximately one-quarter of the maximum power dissipation capability of the diode. Values of V_Z are available at various values from 2.4 to 200 V with accuracies between 5 and 20%, depending upon cost.

Power Dissipation

Power dissipation in the diode is the product of V_Z and reverse current, I_Z, with maximum power ratings ranging from 150 mW to 50 W.

Breakover Current

Since there is some curvature of the reverse characteristic at low values of I_Z, there may be specified some value of current, I_{ZK}, in the neighborhood of the breakover knee, where the voltage across the diode starts to differ greatly from V_Z.

Dynamic Impedance

There is another specification provided by the manufacturer, which helps to describe how "vertical" is the reverse characteristic shown in Fig. 3.13(a). This is called the Zener dynamic impedance, Z_Z, and is defined as:

$$Z_Z = \frac{\Delta V_Z}{\Delta I_Z} \tag{3.95}$$

This impedance is usually evaluated in the region of the specified test current, I_{ZT}, and is the reciprocal of the slope of the reverse characteristic. Ideally, Z_Z is zero for a perfectly vertical breakdown curve, but in practice may vary from several ohms to several hundred ohms, depending upon the particular Zener diode voltage and the operating current as shown in Fig. 3.14. Actually, the equation is more useful in the form:

$$\Delta V_Z = Z_Z \, \Delta I_Z, \tag{3.96}$$

since it gives the change in voltage that will occur across the diode due to a change in current through the diode.

For example, a 12 V Zener diode operating at 5 mA has a Z_Z of approximately 25 ohms (from Fig. 3.14) and will have a change of 25 mV across it if the current changes by 1 mA, while at 40 mA operation the same diode will have only 5 mV change for the same change in current. Evidently a Zener diode operated at a higher value of reverse current will have a more constant voltage across it in the face of any changes in diode current. This important property will now be illustrated in the design of a voltage regulator circuit.

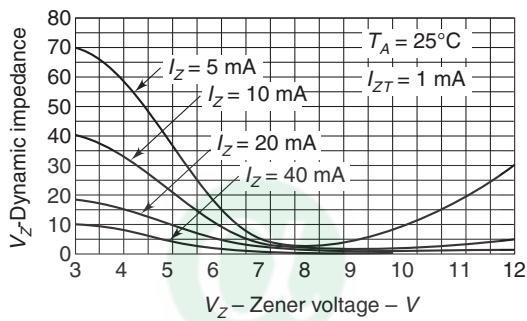


Fig. 3.14 *Typical dynamic impedance vs. zener voltage as a function of zener current*

3.16 ZENER DIODE EQUIVALENT CIRCUIT

The fact that the characteristic curve drops down and away from the horizontal axis rather than up and away for the characteristic positive V_D region reveals that the current in the Zener region has a direction opposite to that of a forward biased diode.

This region of unique characteristics is employed in the design of Zener diodes which have the graphic symbol appearing in Fig. 3.15. Both the semiconductor diode and zener diode are presented side by side in Fig. 3.15 to ensure that the direction of conduction of each is clearly understood together with the required polarity of the applied voltage. For the semiconductor diode the "on" state will support a current in the direction of the arrow in the symbol. For the Zener diode the direction of conduction is opposite to that of the arrow in the symbol as pointed out in the introduction to this section. Note also that the polarity of V_D and V_Z are the same as would be obtained if each were a resistive element.

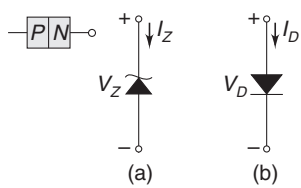


Fig. 3.15 *Conduction direction: (a) zener diode; (b) semiconductor diode*

The location of the Zener region can be controlled by varying the doping levels. An increase in doping, producing an increase in the number of added impurities, will decrease the Zener potential. Zener diodes are available having Zener potentials of 1.8 to 200 V with power ratings from 1/4 to 50 W. Because of its higher temperature and current capability, silicon is usually preferred in the manufacture of Zener diodes.

The complete equivalent circuit of the Zener diode in the Zener region includes a small dynamic resistance and d.c. battery equal to the Zener potential, as shown in Fig. 3.16. For all applications to follow, however, we shall assume as a first approximation that the external resistors are much larger in magnitude than the Zener-equivalent resistor and that the equivalent circuit is simply the one indicated in Fig. 3.16

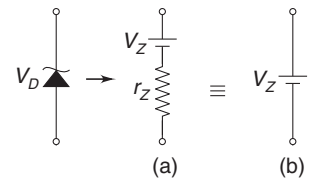


Fig. 3.16 *Zener equivalent circuit: (a) complete; (b) approximate*

3.17 THE VOLTAGE REGULATOR CIRCUIT

The way in which a Zener diode is used to regulate the voltage across a load in the face of supply voltage or load current variations is shown in Fig. 3.17.

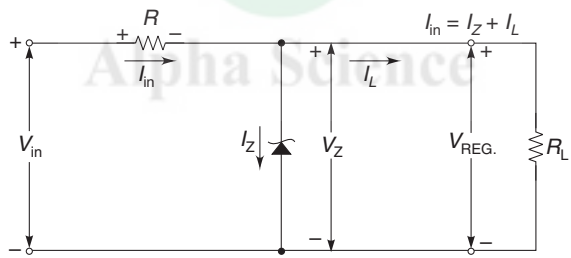


Fig. 3.17 *Circuit of a zener diode voltage regulator showing the series dropping resistor, R*

Note that the reverse characteristics of the Zener diode are being used. The symbol still points in the direction of conventional current flow in the forward-biased connection, even though the diode is not used in this direction. The diode is selected with a V_Z equal to the voltage desired across the load. The Zener diode acts as a "bypass valve," through which more current can pass when an increase in input voltage or a decrease in load current occurs, maintaining the voltage at the output nearly constant at V_Z. It should be noted that resistor R is the element that drops the excess voltage from the input to give the desired output across the load; hence this circuit cannot compensate for voltages on the

input that fall below the desired output. The selection of R is made according to the following equation:

$$R = \frac{V_{in} - V_{out}}{I_{in}} \tag{3.97}$$

where $I_{in} = I_Z + I_L$, using an average value for I_L.

3.18 ZENER DIODE ANALYSIS

The analysis of networks employing Zener diodes is quite similar to that applied to the analysis of semiconductor diodes. First the state of the diode must be determined followed by a substitution of the appropriate model and a determination of the other unknown quantities of the network. Unless otherwise specified, the Zener model to be employed for the "on" state will be as shown in Fig. 3.18(a). For the "off" state as defined by a voltage less than V_Z but greater than 0 V with the polarity indicated in Fig. 3.18(b), the Zener equivalent is the open circuit that appears in the same figure.

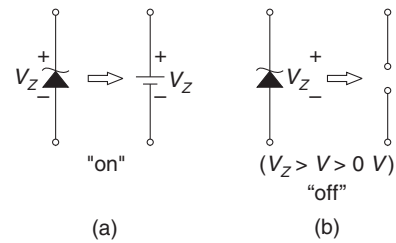


Fig. 3.18 *Zener diode equivalent for the (a) "on" and "off" states*

V_i and R Fixed

The simplest of Zener diode networks appears in Fig. 3.19. The applied d.c. voltage is fixed, as is the load resistor. The analysis can fundamentally be broken-down into two steps.

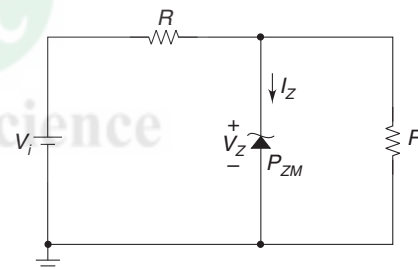


Fig. 3.19 *Basic zener regulator*

1. Determination the state of the Zener diode by removing it from the network and calculating the voltage across the resulting open-circuit.

Applying step 1 to the network of Fig. 3.19 will result in the network of Fig. 3.20, where an application of the voltage divider rule will result in

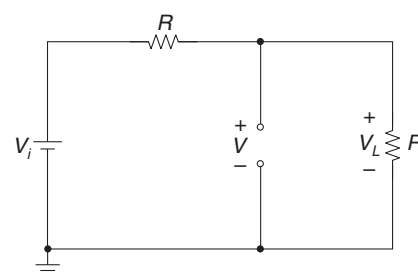


Fig. 3.20 *Determining the state of the zener diode*

$$V = V_L = \frac{R_L V_i}{R + R_L} \tag{3.98}$$

If $V \geq V_Z$, the Zener diode is "on" and the equivalent model of Fig. 3.18(a) can be substituted. If $V < V_Z$, the diode is "off" and the open-circuit equivalence of Fig. 3.18(b) is substituted.

2. Substitute the appropriate equivalent circuit and solve for the desired unknowns.

For the network of Fig. 3.19 the "on" state will result in the equivalent network of Fig. 3.21. Since voltages across parallel elements must be the same, we find that

$$V_L = V_Z \tag{3.99}$$

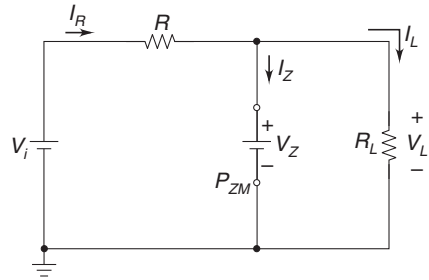


Fig. 3.21 *Substituting the zener equivalent for the "on" situation*

The Zener diode current must be determined by an application of Kirchhoff's current law. That is,

$$I_R = I_Z + I_L$$

and

$$I_Z = I_R - I_L \tag{3.100}$$

where

$$I_L = \frac{V_L}{R_L}$$

and

$$I_R = \frac{V_R}{R} = \frac{V_i - V_L}{R}$$

The power dissipated by the Zener diode is determined by

$$P_Z = V_Z I_Z \tag{3.101}$$

which must be less than the P_{ZM} specified for the device.

Before continuing, it is particularly important to realize that the first step was employed only to determine the *state of the Zener diode*. If the Zener diode is the "on" state, the voltage across the diode is not V volts. When the system is turned on the Zener diode will turn "on" as soon as the voltage across the Zener diode is V_Z volts. It will then "lock in" at this level and never reach the higher level of V volts.

Zener diodes are most frequently used in *regulator* networks or as a *reference* voltage. Figure 3.19 is a simple regulator designed to maintain a fixed voltage across the load R_L. For values of applied voltage greater than required to turn the Zener diode "on", the voltage across the load will be maintained at V_Z volts. If the Zener diode is employed as a reference voltage, it will provide a level for comparison against other voltages.

Fixed V_i, Variable R_L

Due to the offset voltage V_Z, there is a specific range of resistor values (and therefore load current) which will ensure that the Zener is in the "on" state. Too

small a load resistance R_L will result in a voltage V_L across the load resistor less than V_Z and the Zener device will be in the "off" state.

To determine the minimum load resistance of Fig. 3.19 that will turn the Zener diode on, simply calculate the value of R_L that will result in a load voltage $V_L = V_Z$. That is,

$$V_L = V_Z = \frac{R_L V_i}{R_L + R}$$

Solving for R_L, we have

$$R_{L_{\min}} = \frac{RV_Z}{V_i - V_Z} \tag{3.102}$$

Any load resistance value greater than the R_L obtained from Eq. (3.102) will ensure that the Zener diode is in the "on" state and the diode can be replaced by its V_Z source equivalent.

The condition defined by Eq. (3.102) establishes the minimum R_L, but in turn specifies the maximum I_L as

$$I_{L_{\max}} = \frac{V_L}{R_L} = \frac{V_Z}{R_{L_{\min}}} \tag{3.103}$$

Once the diode is in the "on" state, the voltage across R remains fixed at

$$V_R = V_i - V_Z \tag{3.104}$$

and I_R remains fixed at

$$I_R = \frac{V_R}{R} \tag{3.105}$$

The Zener current

$$I_Z = I_R - I_L \tag{3.106}$$

resulting in a minimum I_Z when I_L is a maximum and a maximum I_Z when I_L is a minimum value since I_R is constant.

Since I_Z is limited to I_{ZM} as provided on the data sheet, it does affect the range of R_L and therefore I_L. Substituting I_{ZM} for I_Z establishes the minimum I_L as

$$I_{L_{\min}} = I_R - I_{ZM} \tag{3.107}$$

and the maximum load resistance as

$$R_{L_{\max}} = \frac{V_Z}{I_{L_{\min}}} \tag{3.108}$$

Fixed R_L, Variable V_i

For fixed values of R_L in Fig. 3.19, the voltage V_i must be sufficiently large to turn the Zener diode on. The minimum turn-on voltage $V_i = V_{i_{\min}}$ is determined by

$$V_L = V_Z = \frac{R_L V_i}{R_L + R}$$

and
$$V_{i_{\min}} = \frac{(R_L + R)V_Z}{R_L} \tag{3.109}$$

The maximum value of V_i is limited by the maximum Zener current I_{ZM}. Since $I_{ZM} = I_R - I_L$,

$$I_{R_{\max}} = I_{ZM} + I_L \tag{3.110}$$

Since I_L is fixed at V_Z/R_L and I_{ZM} is the maximum value of I_Z, the maximum V_i is defined by

$$V_{i_{\max}} = V_{R_{\max}} + V_Z$$
$$V_{i_{\max}} = I_{R_{\max}} R + V_Z \tag{3.111}$$

3.19 DESIGN OF A VOLTAGE REGULATOR CIRCUIT

As an example, say $V_{\text{in}} = 20$ V, V_{out} required = 15 V over a load current range from 10 to 20 mA. Assume a 15 V Zener diode has been selected with an I_{ZT} of 6 mA. Let us determine the value of R and the rating of the diode. Assuming an average load current of 15 mA, the total current $I_{\text{in}} = I_Z + I_L = 6 + 15 = 21$ mA, as shown in Fig. 3.22.

$$R = \frac{V_{\text{in}} - V_{\text{out}}}{I_{\text{in}}} = \frac{20 - 15}{21 \text{ mA}} = \frac{5}{21 \times 10^{-3}}$$

$$= \frac{5000}{21} = 238 \text{ ohms.}$$

Resistor power rating $= \frac{V^2}{R} = \frac{5^2}{238} = \frac{25}{238} = 0.1$ W.

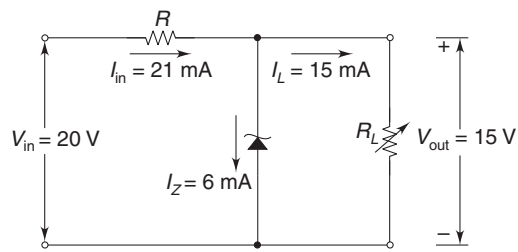


Fig. 3.22 *Condition in voltage regulator circuit at average load current*

Now let us examine the Zener diode current under minimum and maximum load current conditions, with the input voltage constant at 20 V. These conditions are depicted in Fig. 3.23.

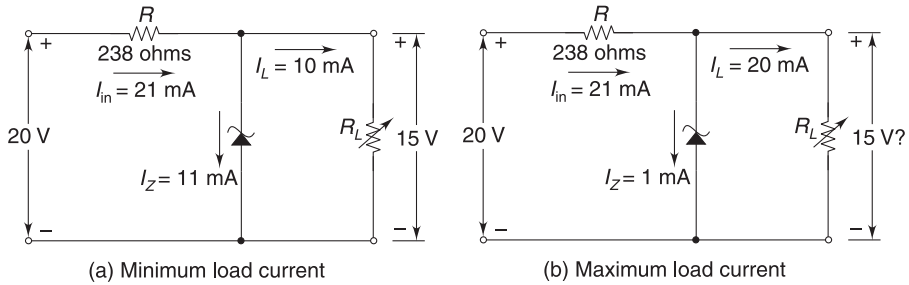


Fig. 3.23 *Voltage regulator circuit conditions for (a) minimum load current, (b) maximum load current*

Note that with this design, when the current to the load falls to 10 mA (from 15 mA, average level used) an extra 5 mA will flow through the Zener diode, giving a total of 11 mA. Most likely, the Zener voltage will still be close to 15 V. However, when the load increases to 20 mA, the extra 5 mA must be supplied from the diode, which means the Zener current must drop to 1 mA. It is doubtful the Zener voltage will still be 15 V at this low level of current and the voltage probably will be less than 15 V, since the "knee" of the curve is being approached.

Before remedying this situation, let us get some idea of the voltage variation across the diode (and load) over the current variation of 10 mA. From Fig. 3.14, the impedance of the 15 V diode at an average Zener current of 6 mA is in the order of 40 ohms. Thus, assuming the "knee" of the curve is avoided, the change in voltage that occurs is:

$$\Delta V_Z = Z_Z \, \Delta I_Z = 40 \times 10 \times 10^{-3} \text{ volts} = 0.4 \text{ V},$$

which corresponds to a voltage regulation of approximately 3%. Now let us see what the effect would be of selecting R so that more than the Zener test current I_{ZT} (at which V_Z is specified to be 15 V) flows through the Zener diode when the average load current is 15 mA. Use a value of 10 mA for example, instead of 6 mA. Then

$$I_{\text{in}} = I_Z + I_L = 10 + 15 = 25 \text{ mA}$$

$$R = \frac{V_{\text{in}} - V_{\text{out}}}{I_{\text{in}}} = \frac{20 - 15}{25 \text{ mA}} = \frac{5}{25 \times 10^{-3}}$$

$$= \frac{5000}{25} = 200 \text{ ohms}.$$

$$\text{Resistor power rating} = \frac{V^2}{R} = \frac{5^2}{200} = \frac{25}{200}$$

$$= 0.125 \text{ watt}.$$

Now, when I_L increases to 20 mA, I_Z will decrease only to 5 mA, at which point the Zener voltage will still be close to 15 V. If I_L decreases to 10 mA, I_Z will increase to 15 mA. This is the worst case of power dissipation in the diode and must be checked:

$$\text{Maximum dissipation} = V_Z I_Z = 15 \text{ V} \times 15 \text{ mA}$$
$$= 225 \text{ mW}.$$

Many small glass Zener diodes have ratings of only 150 mW, so this is an important factor that must be checked.

Since the Zener diode is now operating at a higher average value of current, its dynamic impedance is in the order of 10 ohms, so that the change in voltage that will occur across the load is now only 0.1 V. This corresponds to less than 1% voltage regulation. Of course one could improve this voltage regulation even more by using a Zener diode whose value of V_Z is 15 V at an I_{ZT} of 170 mA for example, where the dynamic impedance is only 3 ohms. But this would require a 10 watt rating for the diode, with its attendant increase in cost, and would also increase the "idling" current from the supply.

Notice that no matter what the value of the resistor, in each of the resistor, in each of the above cases it is used to drop 5 V to give the desired output, independent of load current, since I_{in} is constant at 21 mA in the first case and 25 mA in the second case. This is the condition where the input voltage remains constant, and only the load current varies.

3.20 ZENER DIODE APPLICATIONS

Zener diodes have numerous applications in transistor circuitry. Some of their common applications are explained below.

1. Zener Diode Shunt Regulator (or Voltage Regulation)

Voltage regulation is a measure of a circuit's ability to maintain a constant output voltage even when either input voltage or load current varies.

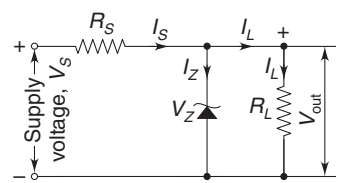


Fig. 3.24

Figure 3.24 shows how a Zener diode can be used as a voltage regulator to provide a constant voltage from a source whose voltage may vary appreciably. A resistor R_S is necessary to limit the reverse current through the diode to safer value. The voltage source V_S and the resistor R_S are so selected that the diode operates in the breakdown region. The diode voltage in this region, which is also the voltage across the load R_L, is called Zener voltage V_Z and the diode current is called the Zener current I_Z.The series resistor R_S absorbs the output voltage

fluctuations so as to maintain voltage across the load constant. Zener diode is reverse connected across the input voltage whose variations are to be regulated.

As long as voltage across the load R_L is less than the breakdown voltage V_Z, the zener diode does not conduct, the resistors R_S and R_L constitute a potential divider across V_S. At an increased supply voltage V_S, the voltage drop across load resistor (or zener diode) becomes greater than the Zener breakdown voltage. It then operates in its breakdown region. The series resistor R_S limits the Zener current I_Z from exceeding its rated maximum $I_{Z\,\max}$. Current through resistor R_S is given as

$$I_S = \frac{V_S - V_Z}{R_S} \tag{3.112}$$

The current from the power supply splits at the junction of Zener diode and the load resistor R_L.

So $$I_S = I_Z + I_L \tag{3.113}$$

When the zener diode operates in its breakdown region, the voltage across it, V_Z remains fairly constant even through the current I_Z flowing through it may vary considerably.

When the supply voltage V_S increases, the current through both the zener diode and load resistor R_L increases. At the same time, however, the zener diode resistor decreases and the current through the diode increases more than proportionately. As a result, a greater voltage drop will occur across the series resistor R_S and the output voltage V_{out} (voltage across the diode or load resistor R_L) will become very close to the original value. The reverse is also true. Thus a Zener diode can maintain the output voltage V_{out} within a fraction of a volt when the supply or input voltage V_S may vary over a range of several volts. The Zener diode will maintain a constant voltage across the load as long as the supply voltage is more than the Zener voltage.

Let us examine the other cause of the output voltage variation. When the load resistance R_L decreases for constant input voltage V_S, load current I_L increases. This additional current is not supplied from the source of supply but the demand of additional load current is met by decrease in Zener current I_Z. This keeps the voltage drop across series resistance R_S constant and so the output voltage V_{out}. The worst case occurs for minimum source voltage and maximum load current because the Zener current reduces to a minimum. In such a case,

$$I_{S\,\min} = \frac{V_{S\,\min} - V_Z}{R_{S\,\max}}$$

or $$R_{S\,\max} = \frac{V_{S\,\min} - V_Z}{I_{S\,\min}} \tag{3.114}$$

From Eq. (3.113) $I_Z = I_S - I_L$

In worst case, this may be written as

$$I_{Z\,\min} = I_{S\,\min} - I_{L\,\max}$$

The critical point occurs when $I_{L\,\max} = I_{S\,\min}$. At this point, the Zener current, I_Z reduces to zero, and regulation is lost.

By substituting $I_{L\,\max}$ for $I_{S\,\min}$ in Eq. (3.114), we have

$$R_{S\,\max} = \frac{V_{S\,\min} - V_Z}{I_{L\,\max}} \tag{3.115}$$

where $R_{S\,\max}$ is the critical value of series resistance, $V_{S\,\min}$ is the minimum source voltage, V_Z is Zener voltage and $I_{L\,\max}$ is maximum load cuurret.

The *critical resistance*, $R_{S\,\max}$ is the maximum allowable series resistance. The series resistance R_S must always be less than the critical value; otherwise; breakdown operation is lost, and the regulator stops its operation.

The advantages of Zener diode voltage regulators over other voltage regulators are that they are smaller, lighter, more rugged and have a longer life. They are also simpler and inherently cheaper.

Such regulators have the following drawbacks:

1. Their efficiency is low for heavy load currents because there is a considerable power loss in series resistor R_S and the diode.
2. The output voltage varies slightly due to Zener impedance r_Z.
3. The output voltage cannot be choosen independently but depends upon the breakdown voltage of the Zener diode.

This regulator is employed when there are small variations in load current and supply voltage.

2. Meter Protection

Figure 6.9 shows how does a Zener diode protect a meter (AVO meter or multimeter) movement or circuit against damage from an accidental application of excessive voltage. In this arrangement Zener diode is connected in parallel with the meter, as shown in Fig. 3.25. In case of accidental overload most of the current will pass through the Zener diode and so the meter movement will be saved from damage. Two Zener diodes connected, as shown in Fig. 3.26 can provide overload protection regardless of the applied polarity.

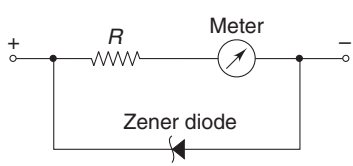


Fig. 3.25

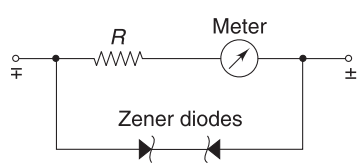


Fig. 3.26

3. Zener Diode as a Peak Clipper

Zener diodes can be used for clipping off the input waveform in waveshaping circuits, as illustrated in Fig. 3.27.

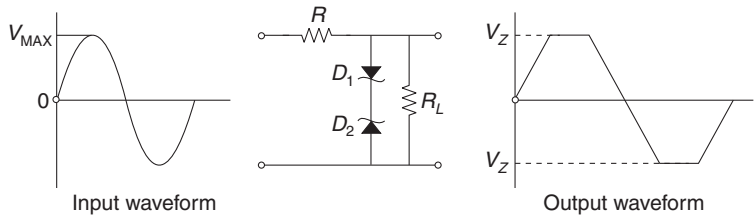


Fig. 3.27

Two similar Zener diodes D_1 and D_2 are shunted across the input sinusoidal voltage of peak value of V_{max} volts. Let the Zener voltage of each of Zener diodes be V_Z volts (V_Z being lesser than V_{max}.

We have seen that a Zener diode acts like a 'short' (or very low resistance) when forward biased and 'open' (or very high resistance) when reverse biased till it goes into breakdown at V_Z. The action of the circuit is explained below.

For positive input half-cycle diode D_1 is forward biased and acts like a 'short' while diode D_2 acts like an open (being reverse biased) up to V_Z. Thereafter diode D_2 goes into breakdown and holds the output voltage constant at V_Z till input voltage falls below V_Z in the latter part of the positive half-cycle. This is because during this duration of time both of the diodes acts as 'short'. When the supply voltage falls below V_Z, diode D_2 comes out of the breakdown and again acts as an open across the load circuit. During the negative input half-cycle, the actions of diodes D_1 and D_2 are reversed. The result is that the output voltage waveform is clipped off on both peaks, as illustrated in Fig. 3.27.

4. Switching Operations

Zener diode can produce a sudden change from low current to high current, so it is useful for switching operations. The use of Zener diode in switching operations has made possible as extremely fast performance in computer applications.

5. Zener Diode as a Reference Element

There are many applications in which it is desirable that a constant voltage is maintained between two points in a circuit, and use this voltage as reference for comparing other voltages against it. The difference between the two voltages (the voltage under comparison and reference voltage) is usually amplified and then used to perform some control function. Such an arrangement is employed in power supply voltage regulator circuits, measurement circuits and

servomechanism circuits. The constant voltage characteristic of a zener diode in its breakdown region makes it desirable for such an application.

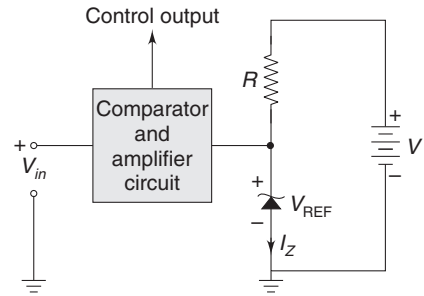


Fig. 3.28

A circuit employing zener diode as a reference element is given in Fig. 3.28. The value of R is so chosen that the diode operates well within its breakdown region. V_{ref} equals the Zener breakdown voltage V_z. The difference of input voltage V_{in} and reference voltage V_{ref} (i.e., $V_{in} - V_{ref}$) gives the control output. Obviously V_{in} should be more than V_{ref}.

Example 3.1 *For the Zener regulator shown in Fig. 3.29, find the maximum and minimum current flowing through the Zener diode.*

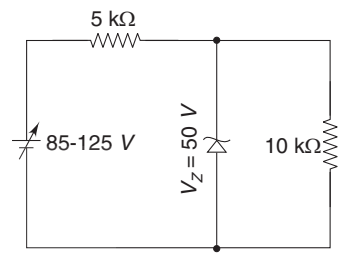


Fig. 3.29

Solution Load current,

$$I_L = \frac{V_{out}}{R_L} = \frac{V_Z}{R_L}$$

$$= \frac{50}{10\ \text{k}\Omega} = 5\ \text{mA}$$

Zener current will be maximum when input voltage is maximum i.e., 125 V.

Corresponding current through series resistance of 5 kΩ

$$I_{S\,max} = \frac{V_{S\,max} - V_{out}}{R_S} = \frac{125 - 50}{5\ \text{k}\Omega} = 15\ \text{mA}$$

Corresponding zener current,

$$I_{Z\,max} = I_{S\,max} - I_L = 15 - 5 = 10\ \text{mA}\ \textbf{Ans.}$$

The zener current will be minimum when input voltage is minimum i.e., 85 V. Corresponding current through series resistance (5 kΩ).

$$I_{S\,min} = \frac{85 - 50}{5\ \text{k}\Omega} = 7\ \text{mA}$$

Corresponding zener current,

$$I_{Z\,min} = I_{S\,min} - I_L = 7 - 5 = 2\ \text{mA}\ \textbf{Ans.}$$

3.21 TUNNEL DIODE

A conventional *P-N* diode is doped by impurity atoms in the concentration 1 part in 10^8.With this order of doping, the width of the depletion layer is of the

order of a micron. The potential barrier corresponding to this width of depletion layer restrains the flow of holes from *P* to *N*-region and electrons from *N* to *P*-region. Now if impurity concentration is greatly increased say to 1 part in 10^3, (corresponding to impurity concentration exceeding 10^{19} per cm^3), then the depletion layer width reduces to about 100 Å and the device characteristics get completely changed. Esaki in the year 1958 announced this new diode and gave appropriate theoretical explanation for its volt-ampere characteristic. This diode utilizes the phenomenon called tunneling and hence the diode is referred to as the *tunnel diode* or *Esaki diode*.

Tunneling Phenomenon

We know that the width of the junction barrier *W* varies inversely as the square root of the impurity concentration. Hence with large impurity concentration of the order of 1 part in 10^3, barrier width *W* reduces to less than 100Å. This width of 100 Å is very small, being only about one-fifth of the wavelength of visible light. As per classical theory, a particle must have an energy at least equal to the height of potential energy barrier in order to cross the junction. However, if the barrier is extremely thin, such as encountered in Esaki diode, the Schrodinger wave equation shows that there is non-zero probability that an electron may penetrate through the barrier. This behaviour is called the *tunneling* and hence the corresponding diode is called *tunnel diode*.

Energy Band Structure of a Highly Doped *P-N* Diode

We have seen that for tunneling to take place, the barrier thickness should be less than 100 Å. This condition is essential for tunneling but not sufficient condition. The additional requirement is that the occupied energy states exist on the side from which the electron tunnels and that allowed empty states exist on the other side at the same energy level. We now consider the energy band structure. In the energy band diagram for a lightly doped *P-N* junction the Fermi level E_F lies inside the forbidden energy gap. We shall see shortly that in an extremely heavily doped diode, the Fermi level E_F lies outside the forbidden band.

We have*

$$E_F = E_C - kT \ln \frac{N_C}{N_D}$$

---

* Concentration of electrons in the conduction band is given by

$$n = N_C \exp\left[-\frac{(E_C - E_F)}{kT}\right]; \qquad N_C = 2\left(\frac{2\pi\, m_e\, kT}{h^2}\right)^{3/2}$$

where E_C is the energy at the lower edge of the conduction band. For *n*-type semiconductor $n \approx N_D$

$$\therefore \qquad \frac{N_C}{N_D} = \exp\left[\frac{E_C - E_F}{kT}\right] \qquad \text{or} \qquad kT \ln\left(\frac{N_C}{N_D}\right) = (E_C - E_F)$$

or $$E_C = E_F + kT \ln\left(\frac{N_C}{N_D}\right)$$

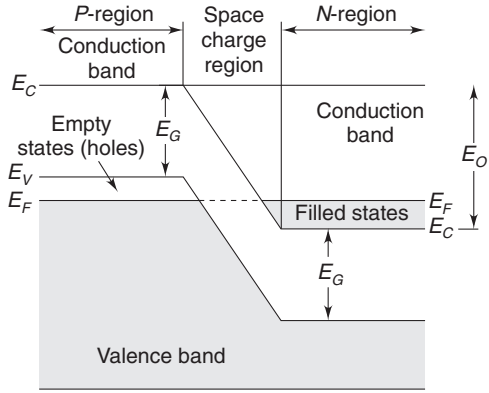


Fig. 3.30 *Energy band diagram in a heavily doped p-n diode under open-circuit condition*

For a lightly doped *n*-type semiconductor, $N_C < N_D$, with the result that $\ln \frac{N_C}{N_D}$ is positive and hence $E_F < E_C$ signifying that the Fermi level lies inside the forbidden band as in Fig. 3.30. Taking $N \approx 10^{19}/\text{cm}^3$, corresponding to doping exceeding 1 part in 10^3, the quantity $\ln N_C/N_D$ becomes negative so that $E_F > E_C$. Thus the Fermi level in the *n*-material lies in the conduction band. Following similar reasoning we find that for a heavily doped *P*-region, $N_A > N_V$ so that the Fermi level E_F lies in the valence band.* Now for Esaki diode $E_0 > E_G$, i.e., the contact potential energy E_0 exceeds the forbidden energy gap energy E_G. Hence for open-circuit condition, the band structure of a heavily doped *P-N* junction has to be as depicted in Fig. 3.30. The Fermi level E_F in the *P*-region is at the same energy level as the Fermi level E_F in the *N*-region. It may be seen from this diagram that there are no filled states on one side of the junction which are at the same energy level as the empty allowed states on the other side. Accordingly there result no flow of charge in either direction across the junction. The current flow through the diode is zero, as should be expected for open-circuited condition.

Volt-ampere Characteristics

To understand the tunneling phenomenon, let us ground the *P*-material and let a voltage be applied across the diode shifting the potential of *N*-side relative to that of the *P*-side. Thus with reverse-bias applied, the height of the barrier gets increased above its open circuit value E_0. This requires that the *N*-side levels must shift downward with respect to the *P*-side levels by an amount corresponding to E_0 plus applied reverse bias magnitude as shown in Fig. 3.31. The result of this is that now we have some energy states (shown heavily shaded in Fig. 3.31) in the valence band of the *P*-region which are at the same level as the allowed empty

* $(E_F - E_V) = kT \ln \frac{N_V}{N_A}$

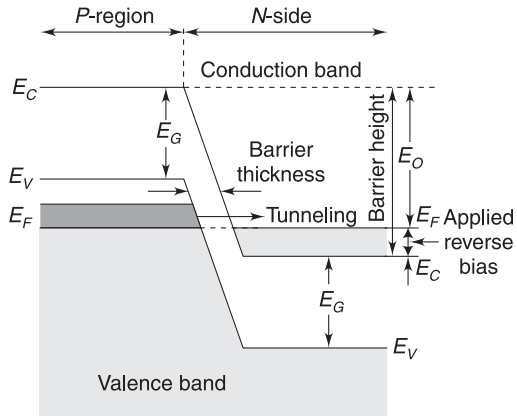


Fig. 3.31 *Energy band diagram in a heavily doped P-N diode with applied reverse bias*

states in the conduction band of the *N*-region. The electrons, therefore, tunnel from the *P*-region to the *N*-region, resulting in the reverse diode current. With the increase of the reverse bias, the heavily shaded region increases in size thereby causing increase in the reverse current as shown by region *OA* in Fig. 3.33.

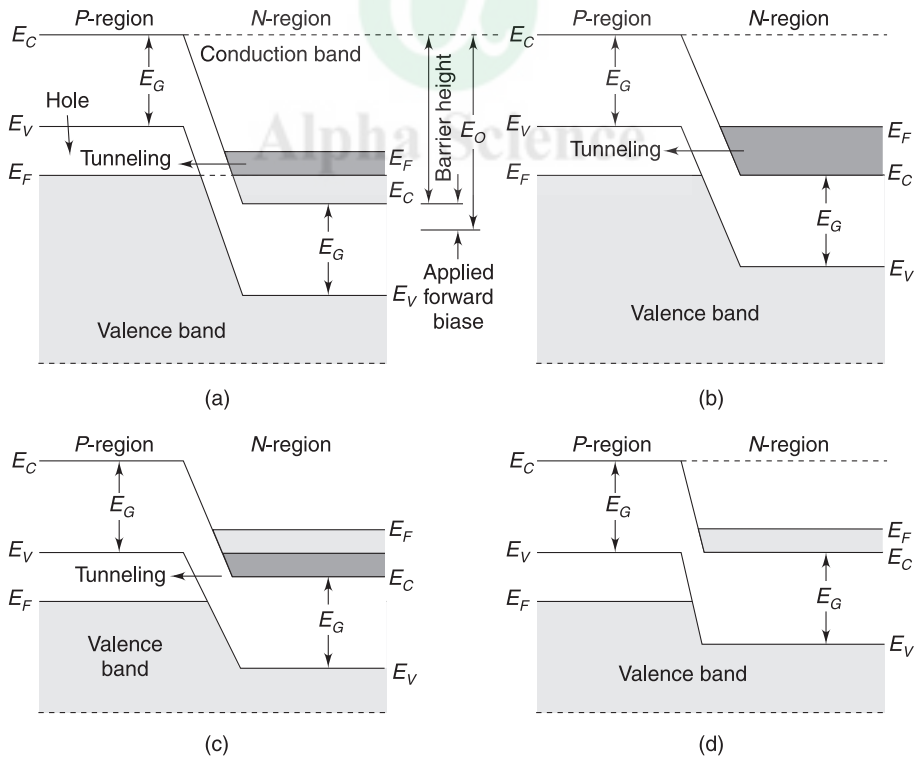
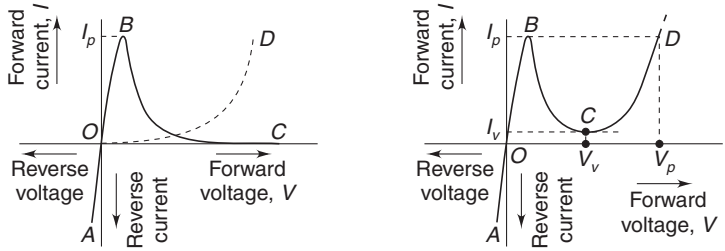


Fig. 3.32 *Energy band diagram of heavily-doped P-N diode with a forward bias. Bias magnitude increases progressively from (a) to (d)*



(a) Tunneling current (solid) and injection current (dashed)

(b) Total tunnel diode current

Fig. 3.33 *Tunnel diode voltampere characteristic*

Next let a forward bias be applied to the diode. This causes the potential barrier to decrease below the value E_0. To do so the *N*-side levels must shift upward relative to the levels on the *P*-side as shown in Fig. 3.32(a). Under this forwards bias condition, there exist occupied states in the conduction band of the *N*-material (shown heavily shaded) which are at the same energy as the allowed empty states (holes) in the valence band of the *P*-material. Electrons, therefore, tunnel from the *N*-region to the *P*-region resulting in the forward current shown by the region *OB* in Fig. 3.33.

Let the magnitude of the forward bias be increased resulting in the condition shown in Fig. 3.32(b). Under this condition, the number of electrons that can leave the occupied states in the *N*-region and tunnel through the barrier to the empty states in the *P*-region is maximum. The heavily shaded region is maximum as shown in Fig. 3.32(b). This condition corresponds to the maximum forward current shown as I_p in Fig. 3.33. On still increasing the forward bias magnitude, the heavily shaded region decreases as shown in Fig. 3.32(c) and the tunneling current decreases as depicted by region *BC* in Fig. 3.33. For even larger value of forward bias, the heavily shaded region vanishes as shown in Fig. 3.32(d). Under this condition, there are no empty allowed states on one side of the junction at the same energy level as the occupied states on the other side. The tunneling current is, therefore, zero.

The solid curve in Fig. 3.33(a) gives the tunneling current, as a function of applied voltage. In addition to this tunneling current, the usual *P-N* junction injection current (given by Eq.*) also flows. This injection current is shown by dashed curve in Fig. 3.33(a). The total current in the tunnel diode is the sum of these two currents and is shown in Fig. 3.33(b).

---

* $I_0 = Aq\left[\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A}\right] n_i^2$

where $n_i^2 = A_0 T^3 e^{-E_{G0}/kT}$

Discussion of the Tunnel Diode Volt-ampere Characteristic

From the *V-I* characteristic of Fig. 3.33(b) we observe that the tunnel diode has an excellent conduction in the reverse direction, i.e., with the *P*-side made negative with respect to the *N*-side. Further, for small forward voltages, up to about 50 mV for germanium diode, the forward resistance remains small, of the order of 5 Ω. At point *B*, the current is maximum and is indicated by I_p. This corresponds to the forward voltage V_p. At this point *B*, the slope *dI*/*dV* is zero. In the region *BC* immediately to the right of *B*, the diode exhibits a *negative-resistance characteristics* between the peak current I_p and the *valley current* (minimum current) I_v. The valley voltage V_v at which current $I = I_v$, again the slope of current-voltage curve is zero, i.e., the conductance *dI*/*dV* is zero. Beyond the point *C*, the current again reaches the value I_p at the so-called *peak forward voltage* V_F. For sill greater forward voltages, the current exceeds the peak current I_p.

It is evident from Fig. 3.33(b), that for current values lying between I_v and I_p, the curve is triple valued, i.e., each current value may be obtained for three different values of voltage *V*. This multi-valued feature of the *I-V* characteristic of tunnel diode makes the device useful in pulse and digital circuits.

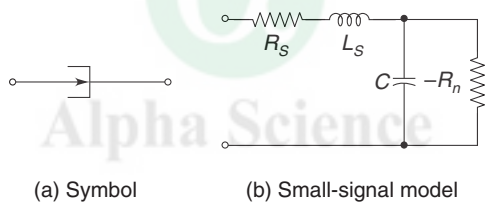


Fig. 3.34 *Symbol and small-signal model for a tunnel diode*

Figure 3.34(a) shows the circuit symbol for a tunnel diode while Fig. 3.34(b) shows its small-signal model for operation in the negative-resistance region. The negative resistance $-R_n$ has a minimum value at some points between the values I_p and I_v, called the point of inflection. Resistance R_s represents the ohmic resistance of the device. The series inductance L_s of the device is caused by the lead length and by the geometry of the diode package. The magnitude of the junction capacitance *C* is a function of the bias. *C* is usually measured at the valley point. Typical values of the circuit parameters depend on the value of peak current. Thus for I_p = 10 mA; the typical values are: R_s = 1 Ω, L_s = 4 nH, *C* = 20 μF and $-R_n = -30$ Ω.

Applications

Tunneling takes place at the speed of light, the transient response being limited only by the total shunt capacitance (junction capacitance *C* plus stray wiring

capacitance) and the peak driving current. Hence tunnel diode is, useful as a very high-speed switch, with switching time as low as 50 ps. However, typically a switching time of the order of a nano-second is used. Another application of the tunnel diode is as a microwave oscillator.

Typical Parameters

Commercial tunnel diodes are made from Ge and gallium arsenide. However, it is rather difficult to manufacture a silicon tunnel diode with the requisite high ratio of peak-to-valley current I_p/I_v. Table 3.1 gives the typical values of I_p/I_v, and voltages V_p, V_v, and V_F. It may be seen from this table that I_p/I_v is highest for gallium arsenide tunnel diode. Similarly, the voltage swing ($V_F - V_P$) is about 0.95 volt for GaAs as against 0.45 volt for germanium.

Table 3.1 *Typical Tunnel Diode Parameters*

Parameter	Ge	GaAs	Si
I_P/I_V	8	15	3.05
V_P (volts)	0.05	0.15	0.06
V_V (volts)	0.35	0.50	0.42
V_F (volts)	0.50	1.10	0.70

The peak current I_p depends on the impurity concentration and the junction area. Tunnel diodes with I_p in the range of 1 to 100 mA are commonly used in computer applications. The peak point $B(I_P, V_P)$ is not a very sensitive function of temperature. On the other hand, the valley point (V_V, I_V) which is influenced by the junction current, is quite temperature-sensitive.

Mertits and Drawbacks

The principal merits of the tunnel diode are its low cost, low noise, simplicity of construction, high speed of switching, low power and environmental immunity. The main drawback of tunnel diode is its low output voltage swing. The other drawback is the fact that it is a two terminal device so that there is no isolated between input and the output. This causes serious circuit design difficulties.

3.22 VARACTOR DIODE

The *varactor* [also called the *varicap*, *voltcap*, *epicap*, VVC (voltage-variable capacitance) or tuning] diode is a semiconductor, voltage-dependent, variable capacitor. Basically, it is just a reverse-biased junction diode whose mode of operation is based on its transition capacitance, The *P*- and *N*-regions (away from the space-charge region) are essentially low resistance areas because of high concentration of majority carriers. The space-charge or depletion region, which is depleted of majority carriers, serves as a *dielectric*. The *P*- and *N*-regions act as

the plates of the capacitor and the depletion region acts as an insulating dielectric. The reverse-biased *P-N* junction thus possesses junction capacitance, called the *transition capacitance* C_T. The value of junction capacitance is given as

$$C_T = \frac{\varepsilon A}{W} \tag{3.116}$$

where ε is the permittivity of the semiconductor material, A is the area of the *P-N* junction and W is the width of the space-charge region.

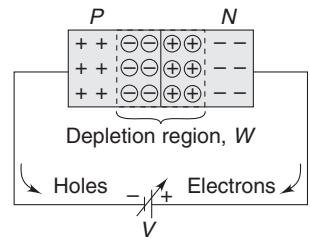


Fig. 3.35 *Depletion region in a reverse biased P-N junction*

With the increase in reverse bias potential, the width of the space-charge region increases, which in turn reduces the transition capacitance C_T. Figure 3.36 illustrates graphically how the capacitance of a varactor diode varies with the reverse bias voltage. The normal range of voltage V_R for varactor diodes is limited to about 20 V. In terms of applied reverse bias V_R,the transition capacitance is given approximately by the following equation

$$C_T = K/(V_K + V_R)^n \tag{3.117}$$

where K is a constant determined by the semiconductor material and fabrication technique, V_K is the diode knee voltage or barrier potential, V_R is the reverse bias applied in volts and n = 1/2 for alloyed junctions and 1/3 for diffused junctions.

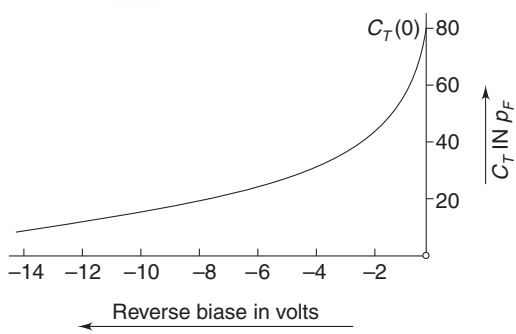


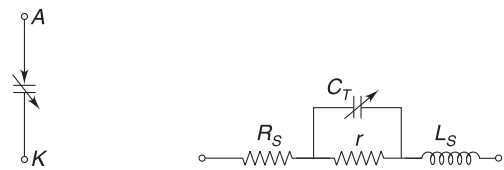
Fig. 3.36

In terms of the capacitance at zero-bias condition C_T (0). The capacitance as a function of reverse bias voltage is given as

$$C_T = \frac{C_T(0)}{\left(1 + \frac{V_R}{V_k}\right)^n} \tag{3.118}$$

where V_k is the diode barrier potential (e.g. 0.7 V for Si).

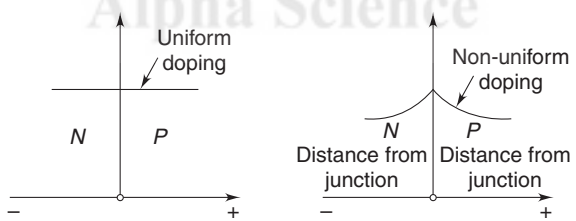
The systematic symbol and a simple equivalent circuit for a varactor diode are given in Fig. 3.37. The resistance R_S represents the body (ohmic) series resistance of the diode. Typical values of C_T and R_S are 20 pF and 8.5 Ω, respectively at a reverse bias of 4 V. The reverse diode resistance r shunting C_T is very large (exceeding 1 MΩ). To ensure that r is as large (for minimum leakage current) as possible, silicon is normally used in varactor diodes. The fact that the device is used at very high frequencies, series inductance L_S is required to be inclined in the equivalent circuit, even though it is measured in nanohenries.



(a) Symbol (b) Equivalent circuit in reversebias mode

Fig. 3.37 *Varactor diode*

The tuning range of a varactor diode depends on the doping level. Figure 3.38(a) shows the doping profile for an abrupt junction diode (ordinary type of diode). Note that the doping is uniform on both sides of the junction i.e., the number of free electrons and holes is equally distributed. The tuning range of an abrupt-junction diode is between 3:1 and 4:1.



(a) Abrupt doping profile (b) Hyper-abrupt doping profile

Fig. 3.38

To get larger tuning ranges, some varactors have a hyper abrupt junction, whose doping profile appears like that shown in Fig. 3.38(b). In this case concentration of charge carriers increases as the distance from the junction reduces. The heavier concentration leads to a narrower depletion region and a larger capacitance. Such a varactor diode has a tuning range of about 10:1, enough to tune an AM radio through its frequency range (535 to 1.605 kHz).

Varactor diodes possessing voltage controlled capacitance have replaced mechanically tuned capacitors in many applications such as television receivers and automobile radios. A varactor diode in parallel with an inductor gives a

resonant tank circuit, whose resonant frequency can easily be changed by varying the reverse bias voltage applied to the diode. These diodes find wide use in the high-frequency (as defined by the small capacitance levels) areas which may include FM modulators, automatic frequency control devices, adjustable bandpass filters and parametric amplifiers.

Varactor Diode in Tuning Circuit

Figure 3.39 depicts two varactor diodes D_1 and D_2 which provide total variable capacitances in a parallel resonant circuit. Here, V_C is a variable d.c. voltage that controls the reverse bias and thus the capacitance of varactor diodes and L is the inductance of the resonant circuit in henry. The resonant frequency of the circuit shown in Fig. 3.39 is

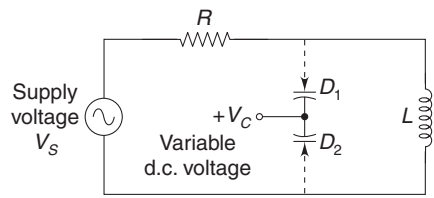


Fig. 3.39 *Varactor diode in tuning circuit*

$$f_0 = \frac{1}{2\pi \sqrt{LC_T}} \tag{3.119}$$

where $C_T = \dfrac{C_1 C_2}{C_1 + C_2}$ if C_1 and C_2 be the maximum and minimum values of diode capacitances.

3.23 PHOTOCONDUCTIVITY

When excess electrons and holes are produced in a semiconductor, there is a corresponding increase in conductivity of a sample an indicated by the following equation

Conductivity of semiconductor,

$$\sigma = \frac{J}{E} = e\,(p\mu_h + n\mu_e) \tag{3.120}$$

When the excess carries in a semiconductor are due to optical luminescence, the resulting conductivity is called *photoconductivity*. This is an important effect, with useful applications in the analysis of semiconductor materials and in the operation of different types of devices.

The photoconductive effect is explained as follows:

The conductivity of a material is proportional to the concentration of charge carriers present, as indicated by Eq. (3.120). Radiant energy supplied to the semiconductor causes covalent bonds to be broken, and EHPs in excess of those generated thermally are produced. These increased current carriers reduce the

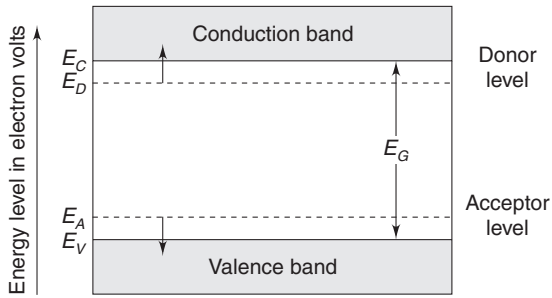


Fig. 3.40 *Photo excitation in semiconductors*

resistance of the material and therefore, such a device is called a *photoconductor* or *photoresistor*.

Energy diagram of the semiconductor having both acceptor and donor impurities is given in Fig. 3.40. When this specimen is illustrated by photons of sufficient energies, the possible transitions are as follows:

An EHP can be generated by a high energy photon, in what is called *intrinsic excitation*; a photon may excite a donor electron into the conduction band, or a valence electron may move into an acceptor state. The last two transitions are called the *impurity excitations*. Since the density of states in the conduction and valence bands greatly exceeds the density of impurity states, *photoconductivity* is due principally to intrinsic excitation.

Spectral Response

The minimum energy of a photon required for intrinsic excitation is the forbidden-gap energy E_G, in electron volts, of the semiconductor material. The long-wavelength threshold of the material is defined as the wavelength corresponding to the energy gap E_G and is given by Eq. 3.40.

$$\lambda_C = \frac{1.24}{E_G} \text{ microns} \tag{3.121}$$

For silicon, $E_G = 1.1$ eV and $\lambda_C = 1.13$ microns

whereas for germanium, $E_G = 0.72$ eV and $\lambda_C = 1.73$ microns at room temperature.

The spectral-sensitivity curves for silicon and germanium are given in Fig. 3.41. The noteworthy point is that the long-wavelength limit is slightly greater than the calculated values of λ_C. This is due to impurity excitations. With the decrease of the wavelength, the response increases and attains a maximum value.

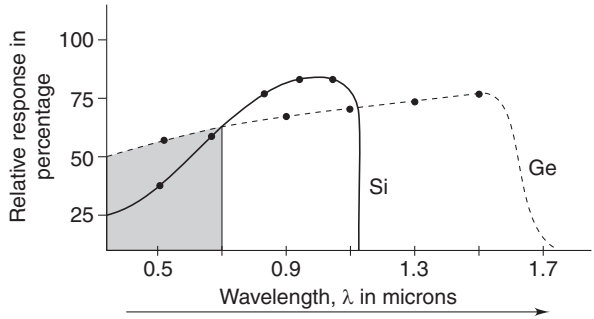


Fig. 3.41 *Relative spectral response of Silicon and Germanium*

Photoconductive Current

The carriers generated by photoexcitation will move under the influence of an applied field. If they survive recombination, they will reach the ohmic contacts at the ends of the semiconductor bar, and thus the device current will be constituted. The current may be given by

$$I = \frac{eG_L \tau}{t_t} \tag{3.122}$$

where G_L is the generation rate of excess carriers, in cm^{-3}-s^{-1}, produced by optical luminescence, τ is the average lifetime of the newly generated carriers and t_t is the average transit time for carriers to reach the ohmic contacts.

3.23.1 Photoconductor

Figure 3.42 depicts a semiconductor bar with ohmic contacts at each end and a voltage applied between the terminals. The initial thermal-equilibrium conductivity is given by

$$\sigma_o = e\,(n_o\,\mu_e + p_o\,\mu_h) \tag{3.123}$$

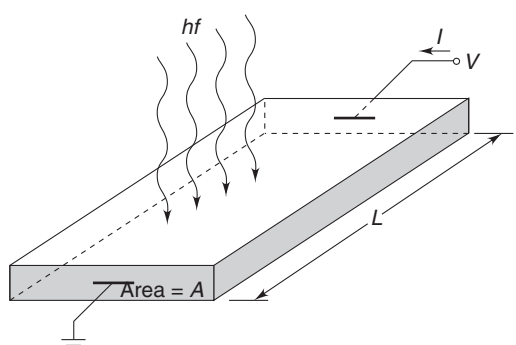


Fig. 3.42 *Photoconductor*

When the excess carriers are created in the semiconductor, the conductivity becomes

$$\sigma = e\ [(n_o + \delta n)\mu_e + (p_o + \delta p)\mu_h] \tag{3.124}$$

where δn and δp are the excess electron and hole concentration, respectively. Considering an *N*-type semiconductor, it can be assumed that $\delta n = \delta p$. Using δp as the concentration is given by

$$\delta p = G_L\ \tau_h$$

where G_L is the generation rate of excess carriers and τ_h is the lifetime of excess minority carriers.

The conductivity from Eq. (3.124) can be written as

$$\sigma = e\ (\mu_e\ n_o + \mu_h\ p_o) + e\delta p\ (\mu_e + \mu_h) \tag{3.125}$$

The change in conductivity due to the optical excitation, called a *photoconductivity*, is then

$$\Delta\sigma = e\ \delta\ p\ (\mu_e + \mu_h) \tag{3.126}$$

Applied voltage induces an electric field in the semiconductor and a current is caused by this electric field. The current density is given by

$$J = (J_o + J_L) = (\sigma_o + \Delta\sigma)E \tag{3.127}$$

where J_0 is the current density in the semiconductor prior to optical excitation and J_L is the photocurrent density. The photocurrent density, $J_L = \Delta\sigma E$. In case of uniform generation of excess electrons and holes throughout the semiconductor, photocurrent is given as

$$I_L = J_L\ A = \Delta\sigma\ .\ AE = eG_L\ \tau_h\ (\mu_e + \mu_h)\ AE \tag{3.128}$$

where *A* is the cross-sectional area of the semiconductor device. The photocurrent varies directly as the excess carrier generation rate, which in turn is proportional to the incident photo flux.

If excess carriers (electrons and holes) are not generated uniformly throughout the semiconductor material, then the total photocurrent is determined by integrating the photoconductivity over the cross-sectional area.

Since $\mu_e E$ is the electron drift velocity, the electron transit time (time required for an electron to flow through the photoconductor) is given as

$$t_n = \frac{L}{\mu_e\ E} \tag{3.129}$$

where *L* is the length of the semiconductor device.

The photocurrent is given by Eq. (3.128),

$$I_L = eG_L\ \frac{\tau_h}{t_n}\left(1 + \frac{\mu_h}{\mu_e}\right) AL \tag{3.130}$$

Photoconductor gain can be defined as the ratio of the rate at which charge is collected by the contacts to the rate at which charge is created within the photoconductor i.e.,

$$\text{Photoconductor gain} = \frac{I_L}{eG_L AL} \tag{3.131}$$

and substituting the value of I_L from Eq. (3.130) in Eq. (3.131) we have

$$\text{Photoconductor gain} = \frac{\tau_h}{t_n}\left(1 + \frac{\mu_h}{\mu_e}\right) \tag{3.132}$$

3.24 PHOTODIODE

Photodiode is a two-terminal semiconductor *P-N* junction device and is designed to operate with reverse bias. The operating principle of a semiconductor photodiode is that when a reverse biased *P-N* junction is illuminated, the current varies almost linearly with the light flux. The device consists of a *P-N* junction embedded in a clear plastic, as shown in Fig. 3.43(a). Radiation is allowed to fall upon one surface across the junction. The remaining sides of the plastic are either painted black or enclosed in a metallic case. The entire unit is extremely small and has dimensions of the order of 25 mm. The basic biasing arrangement and symbols for the device are given in Fig. 3.43(b).

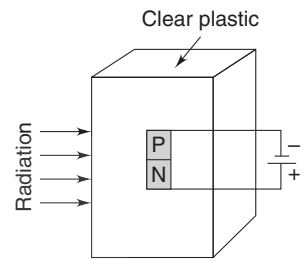


Fig. 3.43(a) *Semiconductor diode construction*

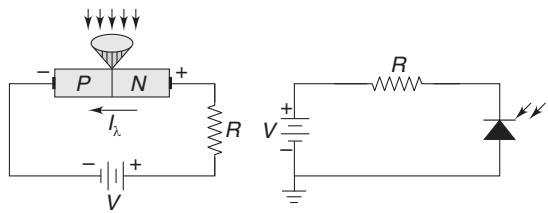


Fig. 3.43(b) *Basic biasing arrangement*

The output voltage is taken from across a series-connected load resistor *R*. This resistance may be connected between the diode and ground or between the diode and the positive terminal of the supply, as illustrated in Fig. 3.43(b).

When the *P-N* junction is reverse biased, a reverse saturation current flows due to thermally generated holes and electron being swept across the junction as the minority carriers. With the increase in temperature of the junction more

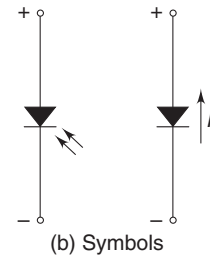


Fig. 3.44

and more hole-electron pairs are created and so the reverse saturation current I_0 increases. The same effect can be had by illuminating the junction. When light energy bombards a *P-N* junction, it dislodges valence electrons. The more light striking the junction the larger the reverse current in a diode. It is due to generation of more and more charge carriers with the increase in the level of illumination. This is clearly shown in Fig. 3.45 for different intensity levels. The *dark current* is the current that exists when no light is incident. It is to be noted here that current becomes zero only with a positive applied bias equals to V_0. The almost equal spacing between the curves for the same increment in luminous flux reveals that the reverse saturation current I_0 increases linearly with the luminous flux as shown in Fig. 3.46. Increase in reverse voltage does not increase the reverse current significantly, because all available charge carriers are already being swept across the junction. For reducing the reverse saturation current I_0 to zero, it is necessary to forward bias the junction by an amount equal to barrier potential. Thus the photodiode can be used as *photoconductive device*.

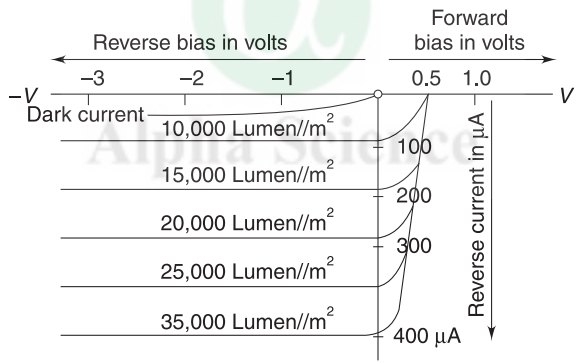


Fig. 3.45 *Photodiode characteristics*

On removal of reverse bias applied across the photodiode, minority carriers continue to be swept across the junction while the diode is illuminated. This has the effect of increasing the concentration of holes in *P*-side and that of electrons in the *N*-side. But the barrier potential is negative on the *P*-side and positive on the *N*-side, and was created by holes flowing from *P* to *N*-side and electrons from *N* to *P*-side during fabrication of junction. Thus the flow of minority carriers tends to reduce the barrier potential.

When an external circuit is connected across the diode terminals, the minority carriers return to the original side via the external circuit. The electrons which crossed the junction from *P* to *N*-side now flow out through the *N*-terminal and

into the *P*-terminal. This means that the device is behaving as a voltage cell with the *N*-side being the negative terminal and the *P*-side the positive terminal. Thus, the photodiode is a *photovoltaic* device as well as *photoconductive* device. When the device is operated with e reverse bias, it operates as a photoconductive device and when operated without the reverse bias, it operates as a photovoltaic device.

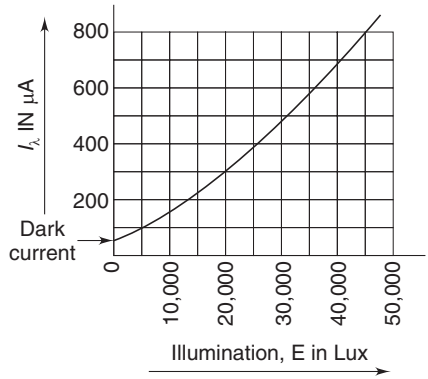


Fig. 3.46

The reverse saturation current I_0 in a *P-N* diode is proportional to the concentration p_{no} and n_{po} of minority carriers in the *N* and *P* regions, respectively, when a reverse biased *P-N* junction is illuminated, the number of hole-electron pairs is proportional to the number of incident photons. So, the current under large reverse bias is $I = I_0 + I_{sc}$, where I_{sc}, the short-circuit current, is proportional to the light intensity. Hence the *V-I* characteristics is given by

$$I = I_{sc} + I_0 \left(1 - e^{V/\eta V_T}\right)$$

where I, I_{sc} and I_0 represent the magnitude of the reverse current, and V is positive for a forward voltage and negative for a reverse bias. The parameter η is unity for germanium and 2 for silicon, and V_T is the volt equivalent of temperature.

With typical photodiodes, the reverse current is in the tens of microamperes.

Since the rise and fall times (change of state parameters) are very small for this device (in the nanosecond range), it can be employed for high speed counting or switching applications.

3.25 LIGHT-EMITTING DIODE

The process of giving off light by applying an electrical source of energy is called *electroluminescence*.

The increasing use of digital displays in calculators, watches and other forms of instruments has contributed to the current extensive interest in structures that will emit light when properly biased. The two types of displays in common use today to perform this function are the light-emitting diode (LED) and the liquid crystal display (LCD).

As the name implies, the light-emitting diode (LED) is a diode that will give off visible light when it is energized electrically. In any forward biased *P-N* junction, there is, within the structure and primarily close to the junction,

a recombination of holes and electrons. This recombination requires that the energy possessed by the unbound free electron be transferred to another state. In all semiconductor *P-N* junctions some of this energy will be given off as heat and some in the form of photons. In Si and Ge, the greater percentage is given up in the form of heat. In other semiconductor materials e.g., GaAsP, or GaP, the number of photons of light energy emitted is sufficient to create a very visible light source (electroluminescence).

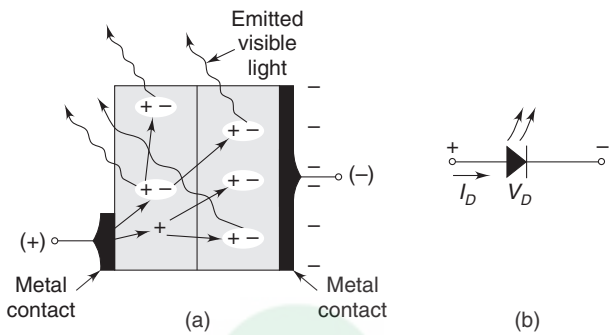


Fig. 3.47 *(a) Process of electroluminescence in the LED (b) graphic symbol*

As shown in Fig. 3.47, with its graphic symbol, the conducting surface connected to the *P*-material is much smaller, to permit the emergence of the maximum number of photons of light energy. Note in the figure that the recombination of the injected carriers due to the forward biased junction results in emitted light at the site of recombination. There may, of course, be some absorption of the packages of photon energy in the structure itself, but a very large percentage are able to leave, as shown in the figure.

It is to be noted that photons may be emitted only if an electron and a hole recombine by a direct band-to-band recombination process in a direct band gap material. The emission wavelength is given by

$$\lambda = \frac{hc}{E_g} = \frac{1.24}{E_g}\ \mu\text{m}.$$

where E_g is the energy gap measured in electron volts, i.e., the wavelength of the output signal of an LED is determined by the energy band gap of the semiconductor. GaAs, a direct band gap material has a band gap energy of 1.42 eV which yields a wavelength of $\lambda = 0.873\ \mu$m. Comparing this wavelength to the visible spectrum, we find that the output of a GaAs LED is not in the visible range. For a visible output, the emitted wavelength should be in the range 0.4 to 0.72 μm. This range of λ corresponds to band gap energies between ~ 1.7 eV to 3.1 eV.

$GaAs_{1-x}P_x$ is a direct band gap material for $0 \le x \le 0.45$. At $x = 0.40$, $E_g = 1.9$ eV which would produce an optical output in the red range. By using

planar technology, Ga $As_{0.6}$ $P_{0.4}$ monolithic arrays have been fabricated for numeric and alphanumeric displays. (For $x > 0.45$, the material changes into an indirect band gap semiconductor (i.e. maximum of C.B. does not lie below the minimum of V.B. and the efficiency is greatly reduced). Table 3.2 provides a list of common compound semiconductors and the light they produce. In addition, the typical range of forward bias voltage for each is listed.

Table 3.2 *Light-emitting diodes*

Coulour	Construction	Typical forward voltage in volts
Amber	AlInGaP	2.1
Blue	GaN	5.0
Green	GaP	2.2
Orange	GaAsP	2.2
Red	GaAsP	1.8
White	GaN	4.1
Yellow	AlInGaP	2.1

Figure 3.48(a) shows a circuit symbol for a LED. The arrows indicate radiation emitted by the diode. Figures 3.48(b) and (c) gives two curves used to determine LED operating characteristics. Figure 3.48(b) is forward bias *V-I* curve for a

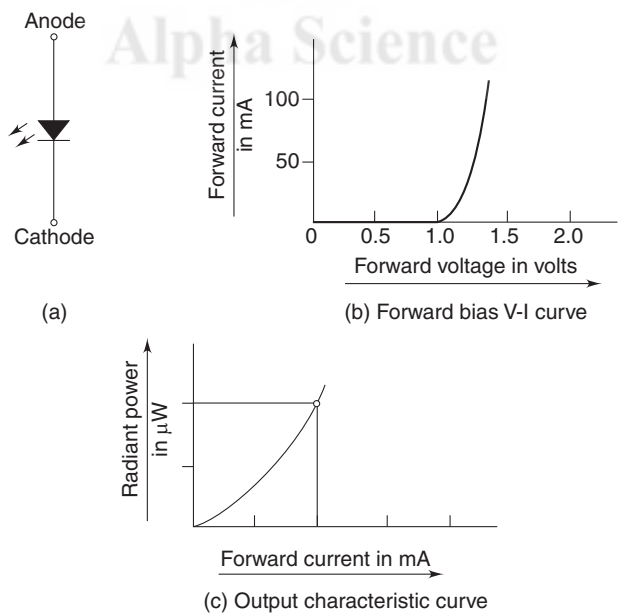


Fig. 3.48 *LED, (a) circuit symbol, (b) forward bias V-I curve, (c) radiant power vs. forward current*

typical LED employed in burglar alarms. Forward bias of approximately one volt is required to give significant forward current. Figure 3.48(c) provides radiant power-forward current curve. The radiant output power is rather small and indicates a very low efficiency of electrical to radiant energy conversion.

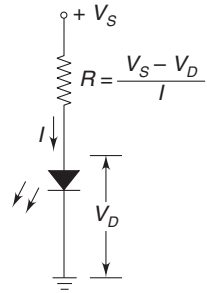


Fig. 3.49 *LED circuit*

Figure 3.49 shows a source connected to a series resistor R_S and a LED. The outward arrows symbolize the radiated light. The forward resistance of LEDs is very low which means that once the forward bias exceeds, the current through it will increase rapidly for only a very small increase in voltage. Thus it becomes necessary to use an external series current limiting resistor. The value of series resistor R_S can be determined from the following equation

$$R_S = \frac{V_S - V_D}{I}$$

where V_S is the supply voltage, V_D is LED forward bias voltage and I is the desired current.

For most of the commercially available LEDs, the typical voltage drop (diode voltage) is from 1.5-2.5 V for currents between 10 and 50 mA. The exact voltage drop depends on the LED current, colour, tolerance etc.

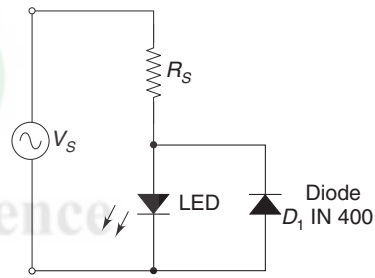


Fig. 3.50 *LED as an indicator in ac circuit*

A LED can be used as an indicator in a.c. circuit by wiring it in inverse parallel with a normal diode, as shown in Fig. 3.50, to prevent the LED being reverse biased; for a given brightness, the value of R_S should be halved relative to that of the d.c. circuit.

3.26 PHOTOVOLTAIC OR SOLAR-CELL

These cells are semiconductor junction devices used for converting radiation energy into electrical energy. These cells generate a voltage proportional to electromagnetic radiation intensity and are called the *photovoltaic cells* because of their voltage generating capability. An example of such a cell is a silicon solar cell which converts radiant energy of the Sun into electrical energy.

Selenium and silicon are the most widely used materials for solar cells, though gallium arsenide, indium arsenide and cadmium sulphide are also used.

Consider a *P-N* junction depicted in Fig. 3.51 with a resistive load. Even when there is no bias applied to the diode, an electric field exists in the space

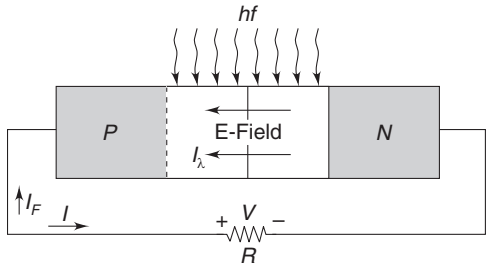


Fig. 3.51 *P-N junction solar cell with resistive load*

charge region as illustrated in the figure. Incident photon illumination can generate EHPs in the space charge region that will be swept out generating photo-current I_λ in the reverse bias direction as illustrated.

The photocurrent I_λ causes a voltage drop across the resistive load which forward biases the *P-N* junction and therefore, a forward bias current I_F as shown in the figure. The resultant *P-N* junction current, in the reverse bias direction is given by

$$I = I_\lambda - I_F = I_\lambda - I_S\,(e^{eV/kT} - 1) \tag{3.133}$$

where the ideal diode equation has been employed. As the diode becomes forward biased, the magnitude of the electric field in the space charge region reduces, but does not reduce to zero or change direction. The photocurrent is always in the reverse bias direction and the resultant current is also always in the reverse bias direction.

The short-circuit condition occurs when load resistance $R = 0$. The current in this case is called the *short-circuit current*, or

$$I = I_{SC} = I_\lambda \tag{3.134}$$

The open-circuit condition occurs when load resistance $R \to \infty$. The resultant current is zero and the voltage developed is the *open-circuit voltage*. The photocurrent is just balanced by the forward-biased junction current and, therefore, we have

$$I = 0 = I_\lambda - I_S\,[e^{eVoc/kT} - 1] \tag{3.135}$$

The open-circuit voltage V_{OC} can, therefore, be determined by equation

$$V_{OC} = V_T \log_e \left(1 + \frac{I_\lambda}{I_S}\right) \tag{3.136}$$

I-V characteristics is shown in Fig. 3.52.

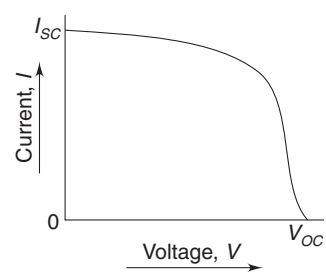


Fig. 3.52 *Current-voltage characteristics of a P-N junction solar cell*

The open-circuit voltage V_{OC} and short-circuit I_{SC} are determined for a given level of light by the cell properties.

The power delivered to the load is given by the product of voltage and current, therefore,

$$P = VI = V[I_\lambda - I_S\,(e^{eV/kT} - 1)] \tag{3.137}$$

The power delivered to the load will be maximum if its derivative is zero i.e.,

$$\frac{dP}{dV} = 0.$$

or
$$I_\lambda - I_S\,(e^{eV_m/kT} - 1) - I_S\,V_m\,\frac{e}{kT}\,e^{eV_m/kT} = 0 \tag{3.138}$$

where V_m is the voltage that makes the power delivered to the load maximum. The above Eq. (3.139) may be rewritten as

$$\left(1 + \frac{V_m}{V_T}\right) e^{eV_m/kT} = 1 + \frac{I_\lambda}{I_S} \tag{3.139}$$

The value of V_m can be found by trial and error. It is to be noted that the maximum power delivered to the load illustrated by the shaded rectangle in Fig. 3.53 is less than the $V_{OC}\,I_{SC}$ product. The ratio $V_m\,I_m/(V_{OC}\,I_{SC})$ is called the *fill factor*, and is a figure of merit for solar cell design.

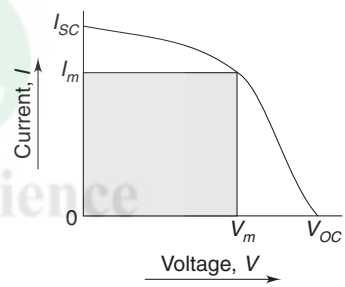


Fig. 3.53 *Maximum power rectangle of solar cell I-V characteristic*

The noteworthy point is that J_S is a function of the semiconductor doping concentrations. With the increase in doping concentrations, J_S decreases, which results in increase of open-circuit voltage V_{OC}. However, as the open-circuit voltage V_{OC} is a function of log of I_λ and I_S, the V_{OC} is not a strong function of these parameters.

The construction and cross section of a typical power solar cell for use as an energy converter are given in Fig. 3.54. The surface layer of *P*-type material is extremely thin so that light can penetrate to the junction. The nickel-plated ring around the *P*-type material is the positive output terminal, and the plating at the bottom of the *N*-type material is the negative output terminal. Power solar cells are also available in flat strip form for efficient coverage of available surface areas. The circuit symbol often used for a photovoltaic cell is given in Fig. 3.55(a).

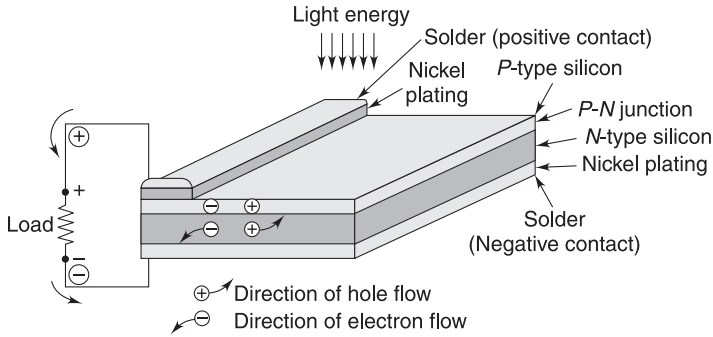
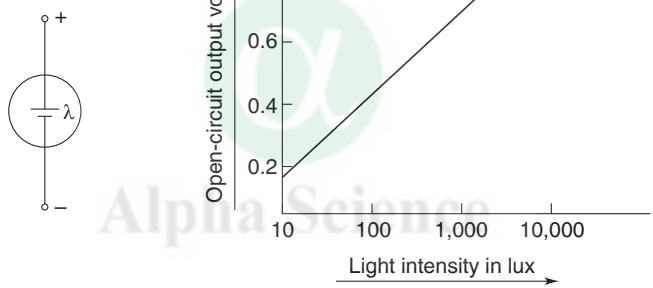
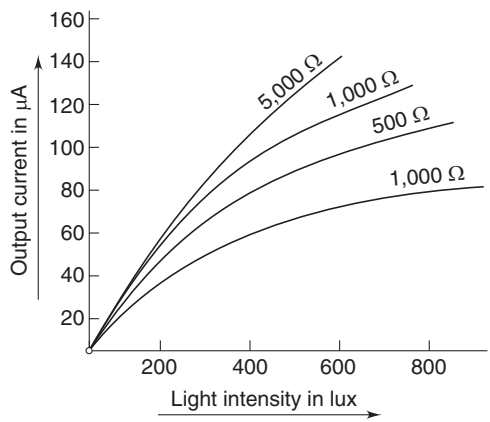


Fig. 3.54 *Solar cell*



(a) Circuit symbol (b) Output voltage characteristic



(c) Output current characteristic

Fig. 3.55 *Photovoltaic cell*

The open-circuit output voltage characteristic of a typical photovoltaic cell is given in Fig. 3.55(b), the graph is *logarithmic* on *light intensity axis*. This characteristic shows that the cell is more sensitive for low light levels, because a small change in light intensity (say from 10 to 100 lux) can produce the same increase in output voltage as a large change in light intensity (say from 100 to 1,000 lux) at a higher light intensity level.

The output current characteristics of a typical photovoltaic cell for various load resistances are given in Fig. 3.55(c). These characteristics show that the output current of a photovoltaic cell increases with the increase in light intensity. The output current of such a cell is very low and is measured in microamperes. The available output current depends upon the light intensity, cell efficiency (typically only a few per cent), and on the size of the active area of the cell face. The conversion efficiency depends upon the spectral content and the illumination. Photocells can be stacked in parallel, in order to increase their output current capability.

The photovoltaic cell can be operated satisfactorily over a wide range of temperature (say from -100 to 125°C). The temperature variations have little effect on short-circuit current but affect the open-circuit output voltage considerably. These variations may be of the order of a few mV per °C in output voltage. An individual solar cell generates an open-circuit voltage of about 500 mV (depending on light intensity) when active. However, photocells can be connected in series to increase the available terminal voltage.

The advantages of such devices are their ability to generate a voltage without any bias and their extremely fast response i.e., these devices can be used as energy convertors directly.

Questions

3.1 (a) Draw the circuit of a half wave rectifier using *P-N* diode and resistive load and describe its working.

(b) Assuming zero in voltage, and $R_s = \infty$, draw waveforms for the output current and voltage for sinusoidal input voltage.

(c) Derive expressions for d.c. and r.m.s. currents, d.c. and r.m.s. output voltages, d.c. output power, ac input power, rectifier efficiency and ripple factor.

3.2 (a) Draw the circuit of a full wave rectifier using two *P-N* diodes and resistive load and describe its working.

(b) Assuming zero cut in voltage and $R_r = \infty$ draw waveformer for the output current and voltage for sinusoidal input voltage.

(c) Derive expressions for d.c. and r.m.s. load currents, d.c. and r.m.s. output voltages, d.c. output power, a.c. input power, rectifier efficiency and ripple factor.

3.3 Define the term *regulation*. Derive the regulation equation for a fullwave rectifier.

3.4 (a) Define *peak inverse voltage*.

(b) What are the values of peak inverse voltage for halfwave and full wave rectifier using ideal diode and sinunoidal input.

3.5 Compare the performance of half wave and fullwave rectifiers.

3.6 Draw the circuit and describe the operation of a bridge rectifier. What are the merits and limitations of this circuit?

3.7 With a neat sketch, explain the working of a (i) centre tap full wave rectifier (ii) fullwave bridge rectifier.

3.8 Derive an expression for the efficiency of a halfwave rectifier.

3.9 Derive an expression for the efficiency of a fullwave rectifier.

3.10 Derive an expression for efficiency and regulation for a bridge rectifier.

3.11 What is a ripple factor? What are its values for a halfwave and fullwave rectifier?

3.12 Describe the action of the following filter circuits: (i) capacitor filter (ii) choke input filter and (iii) capacitor input filter.

3.13 (a) Describe the working of a fullwave rectifier with a (i) *L*-section filter (ii) π-section filter.

(b) Describe salient features of *L*-section and π-filters.

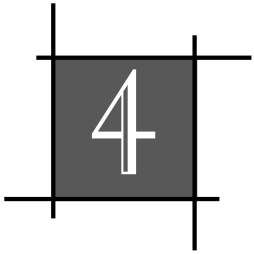
3.14 What is a zener diode? Explain how zener diode maintains constant voltage across the load?

3.15 Why is zener diode used as a voltage-regulator?

3.16 Why is capacitor input filter preferred to choke input filter?

3.17 What is a photodiode? How does it function? Give its characteristics and describe one of its applications.

3.18 What is a tunnel diode? Give the *V-I* characteristics of a tunnel diode and briefly describe the properties of the curve in different regions. Explain what is meant by negative resistance? What is the use of a tunnel diode?



The Bipolar Junction Transistors

4.1 INTRODUCTION

In 1947, Bardeen and Brattain demonstrated that the current through a forward biased point junction on a semiconductor could control the current to a reverse biased third electrode, mounted very close to the first contact. The resultant three electrode semiconductor control device was named the transistor (a contraction of the words "transfer resistor"). The junction form of transistor was later proposed by Shockley, to overcome some of the problems associated with point contacts. These devices operate with both electrons and holes and are accordingly known as *bipolar junction transistors* (BJTs).

(Another semiconductor control device is the *field effect transistor* (FET), also proposed by Shockley. Control of current in a FET is by an electric field across the conduction region, employing only majority carriers, the FET is therefore said to be *unipolar*).

4.2 THE PNP AND NPN TRANSISTORS

The bipolar junction transistor (BJT) is a three layer sandwich of *P* and *N* materials, as illustrated in Fig. 4.1. The transistor is formed as two opposed *P-N* junctions, very close together in a common wafer of semiconductor material. The three layers may be arranged in *PNP* or *NPN* order; the polarizing bias voltages and charge carriers are reversed in the two types, as shown in Fig. 4.1(a) and (c). The positive directions of the currents are directed inward, by definition.

The three layers in the transistor are designated as *emitter*, *base*, and *collector*. Junctions are present at the interfaces between the *N* and *P*-regions. The first junction is operated with forward bias and lies between the emitter and the base. The emitter acts as the source of the mobile charge carriers and the emitter-base potential controls the current diffusing into the base. The second junction is operated with reverse bias and lies between the base and the collector. The latter element receives the charges after transit of the base.

We will describe the operation of the *PNP* form as an example. Biased in the forward direction, or with emitter positive to the base, the normal diode voltage-

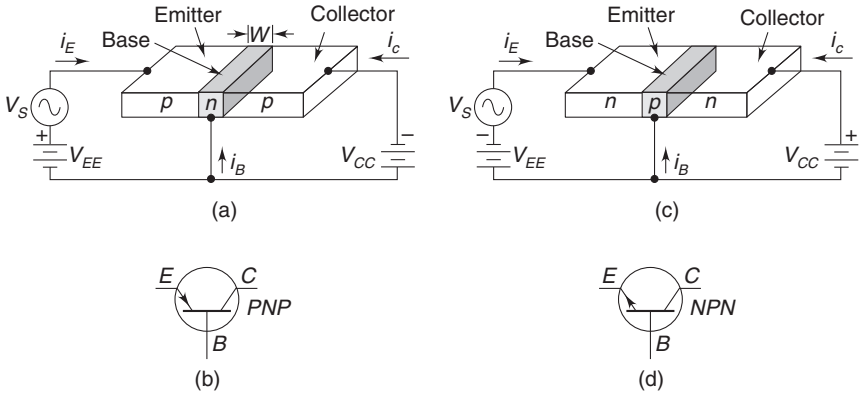


Fig. 4.1 *(a) Transistor, PNP form; (b) PNP symbol; (c) NPN form; (d) NPN symbol*

current law applies to the emitter-base junction. The action of the forward bias lowers the junction potential and with a high ratio of *P* to *N* density between emitter and base, holes are the predominant carrier diffusing into the base, as indicated in Fig. 4.2.

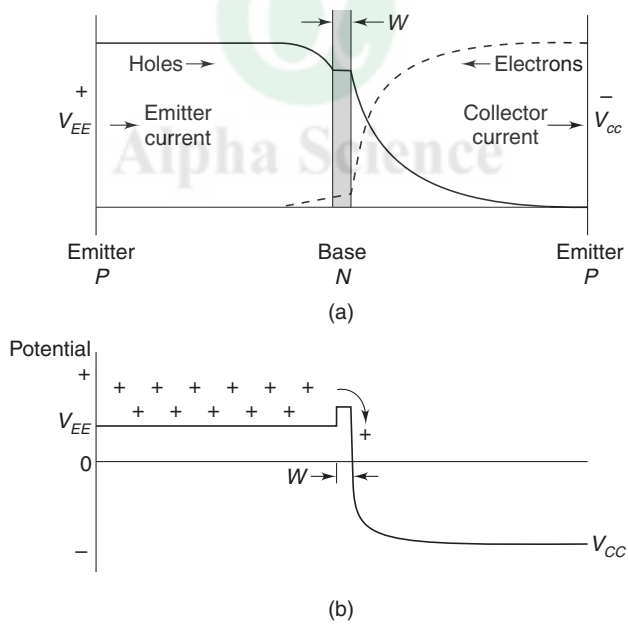


Fig. 4.2 *(a) Hole and electron currents; (b) potentials across junction*

The base width *W* is made very small with respect to the average distance a hole can diffuse before recombination can occur in the *N* base region. Ninety-five per cent or more of the diffusing holes will transit the base and reach the

base-collector junction. The base is so named, because it was the wafer on which the point electrodes of the original transistor made contact.

The base-collector junction is reverse biased, with a negative potential on the *P* collector, with respect to the base. The holes diffusing across the base to this junction are swept through the junction in the direction of easy hole flow, as shown in Fig. 4.2(b). The holes meet electrons in the *N* collector region and recombine. Replacement electrons move into the collector region from the negative supply and constitute the current i_C.

As the hole current into the base is varied by the applied junction potential in accordance with *V* of the diode equation, the collector current varies correspondingly. Therefore, the transistor is a *control element* in which the emitter-base voltage controls the collector-output current.

Some holes do recombine in passing through the base, and this creates a small electron flow in the base lead, to resupply the negative charge lost to recombination. This charge flow constitutes the steady-state base current i_B. With a change in base potential, there must be a change in the charge in the base, and the base current may have a transient component to supply this stored charge whenever the potential changes. This action appears as a storage capacitance, C_s, as in case of the *PN* diode.

At distance remote from the junctions in the transistor the current is composed of holes in the *P* emitter and by electrons in the *P* collector. The terminal currents remove electrons from the emitter, creating holes therein, and supply electrons to the collector from the negative supply. Since conduction is by both holes and electrons, the junction transistor is *bipolar*.

The resistance of the input forward biased junction is low, usually a few hundred ohms. The output circuit includes the reverse biased junction with a resistance of thousands of ohms. Since the same current exists in both regions the output power level is appreciably greater than the input power level, and power amplification is achieved by the transistor.

The *NPN* transistor could be discussed in like manner, after reversal of the bias potentials to maintain the forward and reverse biased junction, and by considering the emitter current as composed of electrons. The arrows on the emitter elements of the circuit symbols in Fig. 4.1 distinguish between *PNP* and *NPN* forms of the transistor by indicating the direction of positive charge flow at the emitter.

4.3 TRANSISTOR OPERATION

Let us consider the *PNP* transistor. We know that for normal operation of a transistor, the emitter-base junction is forward biased and collector-base junction is reverse biased. The transistor-operation can be understood as follows:

In Fig. 4.3, the *PNP* transistor has been drawn without the base-to-collector bias. V_{EE} serves as the forward bias. Due to the applied bias, the depletion region gets reduced resulting in a heavy flow of majority carriers from the *P*- to the *N*-region (i.e. emitter to the base).

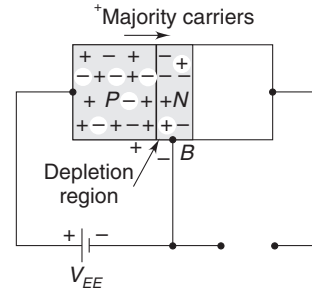


Fig. 4.3 *Forward biased E-B junction of a PNP transistor*

Let us now remove the base-to-emitter bias of the *PNP* transistor as shown in Fig. 4.4. Due to the applied reverse-bias (to the *C*-*B*), the flow of majority carriers is zero, resulting in only a minority-carrier flow (as indicated in Fig. 4.4).

In Fig. 4.5, both biasing potentials have been applied to a *PNP* transistor. The resulting majority- and minority-carrier flow have been indicated. Note that (in Fig. 4.5) the widths of the depletion region, it is less for *E*-*B* junction and more for *C*-*B* junction, due to the applied biases.

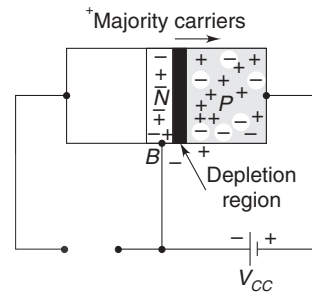


Fig. 4.4 *Reverse-biased C-B junction of a PNP transisitor*

As indicated in Fig. 4.5, a large number of majority carriers will diffuse across the forward biased *P*-*N* junction into the *N*-type material. The question then is whether these carriers will contribute directly to the base current I_B or pass directly into the *P*-type collector. *Since, the sandwiched N-type material is very thin, and has a low conductivity, a very small number of these carriers will take the path of high resistor to the base material.* (The magnitude of the base current is typically on the order of microamperes as compared to milliamperes for the emitter and collector currents). The large number of the majority carriers will diffuse into the *P*-type collector region due to reverse bias on the collector. Further all the minority carriers in the (*C*-*B*) depletion region cross the reverse biased junction. These facts when combined, account for the flow of currents as indicated in Fig. 4.5.

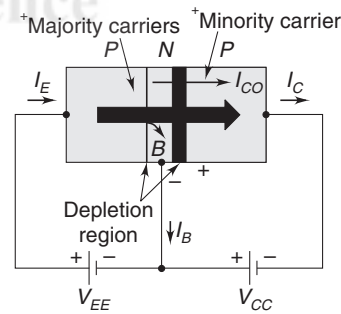


Fig. 4.5 *Majority and minority carrier flow of a PNP transistor*

Applying KCL to the transistor of Fig. 4.5 as if it were a single node, we obtain

$$I_E = I_C + I_B$$

i.e., the emitter current is the sum of the collector and base currents. The collector current, however, is comprised of two components, the majority and minority carriers as indicated in Fig. 4.5. The minority-current component is called the *leakage current* and is given the symbol I_{CO} (i.e., I_C current with emitter terminal open). Therefore,

$$I_C = I_{C_{\text{majority}}} + I_{CO_{\text{minority}}}$$

4.4 EMITTER EFFICIENCY γ, BASE TRANSPORT FACTOR β* AND THE COLLECTOR EFFICIENCY δ

(a) Emitter Efficiency γ

The emitter current I_E consists of two components (1) hole component of the emitter current I_{pE} and (2) electron component of the emitter current I_{nE}. In explaining the performance of a *PNP* transistor, it may be pointed out that the hole component of the current I_{pE} is the link for the emitter circuit to control the collector current. Thus, to have a high efficiency of the emitter circuit in injecting holes in the base region of a *PNP* transistor, the hole component of the emitter current I_{pE} should be as high as possible, so that almost the entire current in the emitter circuit can take part in the transistor action. The emitter efficiency γ is defined as the ratio of the change in the injected component of the hole current I_{pE} to the change in the total emitter current I_E. Thus

$$\gamma = \frac{\Delta I_{pE}}{\Delta I_E} = \frac{\Delta I_{pE}}{\Delta I_{pE} + \Delta I_{nE}}$$

It may be shown that (for a *PNP* transistor),

$$\gamma = 1 - \left(\frac{\sigma_n W}{\sigma_p L_n}\right) = 1 - \left(\frac{\sigma_b W}{\sigma_e L_n}\right)$$

where σ_b and σ_e are the conductivities of the *N*-type base and *P*-type emitter respectively of the *PNP* transistor.

In a *PNP* transistor, σ_p of the *P*-type emitter region is made for larger than that of the *N*-type base region σ_n and the width *W* of the base region (*N*-type) is made far smaller than the diffusion length L_p of holes in the *N*-region, so that almost the entire current flowing in the emitter junction consists of positive hole current, while the electron component of the emitter current is negligible (~ 0.1% of I_E). Thus the emitter efficiency γ is increased. The emitter efficiency of a transistor increases with the conductivity of the emitter and the diffusion length L_n of electrons in the *P*-type emitter region, and falls with the width *W* of the base region and the conductivity of the base region.

(b) Base Transport Factor β*

When the hole component of the current is injected into the base region, a small fraction of the same may be lost due to recombination with the free electrons in the base region giving rise to a small base current. The transport factor β* is defined as the ratio of a small change in the collector current I_C to a small change in the hole component of current I_{pE} i.e.,

$$\beta^* = \frac{\Delta I_C}{\Delta I_{pE}} = \frac{\Delta I_{pc}}{\Delta I_{pE}}$$

It may be shown to be given by

$$\beta^* = 1 - \frac{W^2}{2L_p^{\,2}}$$

The transport factor β* will be high if the highest possible amount of the hole current reaches the collector junction; this will be available as the collector current. The value of β* increases with the reduction of the width W of the base region since a smaller number of the injected holes will be lost in the base region due to recombination.

The short-circuited current gain α (see Section 4.7) is related to the emitter efficiency γ and the transport factor β*:

$$\alpha = \frac{\Delta I_C}{\Delta I_E} = \left(\frac{\Delta I_{pc}}{\Delta I_{pE}}\right)\frac{\Delta I_{pE}}{\Delta I_E} = \beta^*\gamma$$

The factors γ and β* can be made close to unity and the typical value of their product alpha (α) ranges from 0.95 to 0.98. With a higher collector voltage, a third factor has to be taken into account in finding the value of α. This is the avalanche multiplication of the collector current in the high field of the collector depletion layer.

(c) Collector Efficiency δ

The collector efficiency (collector multiplication factor) δ is defined as the ratio of the current leaving the collector region to the hole current entering the collector region from the base. This factor may exceed unity. There is a generation of secondary carriers within the collector depletion region as a consequence of the ionization of atoms by collision with the highly accelerated primary holes. Thus at low collector voltage $\alpha = \beta^*\gamma$ and at a high collector voltage $\alpha = \beta^*\gamma\delta$ where δ is the collector multiplication factor. Through there is no satisfactory theory for this action, there is an empirical relationship which is quite correct in describing this effect. The collector efficiency (or collector multiplication ratio δ) is given by

$$\delta = \frac{1}{1 - \left(\frac{V_C}{V_A}\right)^n}$$

where V_A is the breakdown voltage of the collector junction, V_C is the collector voltage, n is empirically determined and is approximately equal to 3 for an alloyed *PNP* transistor. The value of V_A for a typical audio transistor is of the order of 100 volts. At V_C = 20 volts, the collector efficiency δ would be equal to 1.016. But for a low value of the collector voltage V_C, the value of δ departs negligibly from unity.

Note that although in the case of a *PNP* or *NPN* transistor, the collector and the emitter materials are of the same type; their electrical conductivities are made different by adding different amounts of impurity atoms. Thus, the emitter and the collector materials are not exactly identical so far as their current conducting properties are concerned and therefore they are not interchangeable.

4.5 COMMON-BASE CONFIGURATION

Figure 4.6 shows the notations and symbols used in conjunction with the transistors for the common-base configuration with *PNP* and *NPN* transistors.

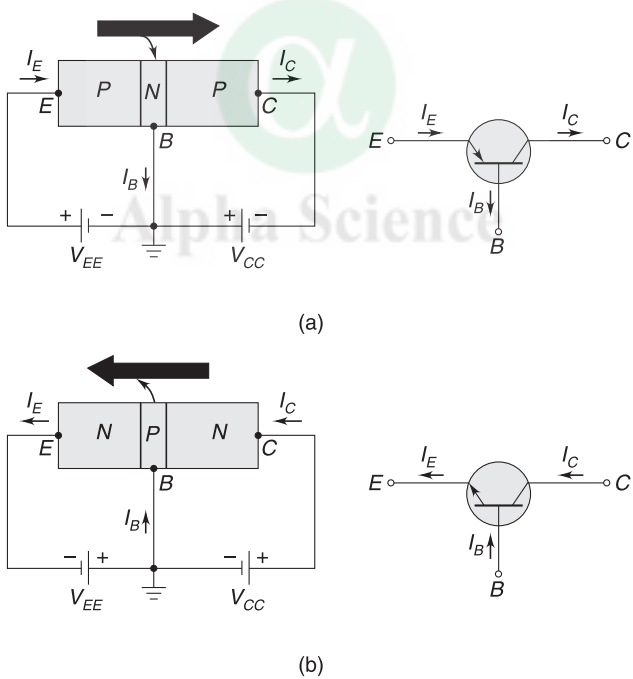


Fig. 4.6 *Notation and symbols used with the common-base configuration (a) PNP transistor, (b) NPN transistor*

The arrows in the diode symbol define the direction of conduction for conventional current and the arrow in the graphic symbol defines the direction of emitter current (conventional flow) through the device.

All the current directions appearing in the Fig. 4.6 are the actual directions as defined by the choice of conventional flow. In each case $I_E = I_C + I_B$. Note that the applied biasing (voltage sources) are such as to establish current in the direction indicated for each branch.

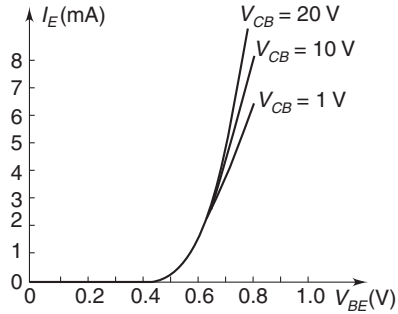


Fig. 4.7 *Input or driving point characteristics for a common-base silicon transistor amplifier*

To fully describe the behaviour of a three-terminal device such as the common-base amplifiers of Fig. 4.6 requires two sets of characteristics—one for the *driving point* or *input* parameters and the other for the *output* side. The input set for the common-base amplifier as shown in Fig. 4.7 will relate an input current (I_E) to an input voltage (V_{BE}) for various levels of output voltage (V_{CB}).

The output set will relate an output current (I_C) to an output voltage (V_{CB}) for various levels of input current (I_E) as shown in Fig. 4.8. The output or *collector* set of characteristics has three basic regions of interest, as indicated in Fig. 4.8: the *active*, *cut-off* and *saturation* regions. The active region is used for linear (undistorted) amplification. In the active region, the collector-base junction i.e., reverse biased and the emitter-base junction is forward biased. (It is defined by the biasing arrangement of Fig. 4.6). The output characteristics are shown in Fig. 4.8.

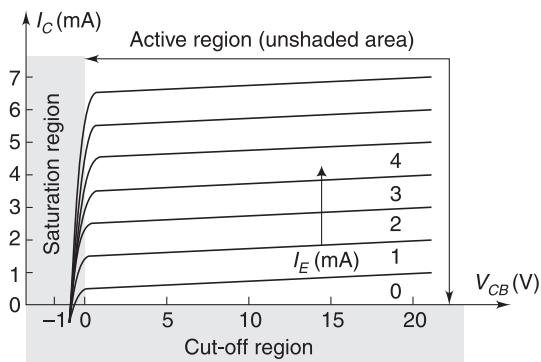


Fig. 4.8 *Output or collector characteristics for a common-base transistor amplifier*

At the lower end of the active region, the emitter current I_E is zero, the collector current is simply that due to reverse saturation current I_{CO} as indicated in Fig. 4.5. The current I_{CO} is so small (~ microamperes) in magnitude compared

to the vertical state of I_C (~ milli-amperes) that it appears on virtually the same horizontal line as $I_C = 0$. The circuit conditions that exist when $I_E = 0$ for the common-base configuration are shown in Fig. 4.9.

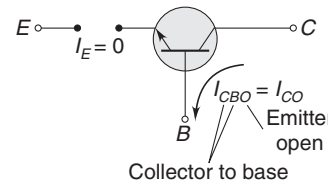


Fig. 4.9 *Reverse saturation current*

The notation most frequently used for I_{CO} on data and specification sheets is (as indicted in Fig. 4.9), I_{CBO} (i.e. collector current with emitter open). Because of improved construction techniques, the level of I_{CBO} for general purpose transistors in the low- and mid-power ranges is usually so low that it can be ignored. However, for higher powers, I_{CBO} still appears in the microampere range. Further, it is temperature sensitive and increases rapidly with temperature.

Note in Fig. 4.8, that in active region, there is almost negligible effect of V_{CB} on the collector current. The curves also indicate that, to a first approximation in the active region:

$$I_C \approx I_E$$

The cut-off region is defined as that region where the collector current is *OA* (Fig. 4.8). Further: *in the cut-off region, the collector-base and emitter-base junctions of a transistor are both reverse biased.*

The saturation region is defined as that region of the characteristics to the left of $V_{CB} = 0$. Note the exponential increase in collector-current as the voltage V_{CB} increases toward zero volts.

In the saturation region, the collector-base and base-emitter junctions are forward biased.

4.6 BASE-WIDTH MODULATION AND PUNCH THROUGH EFFECT

In the active region, the emitter-base (EB) junction is forward biased and collector-base (*CB*) junction is reverse biased, so the barrier width at *E-B* junction is negligible in comparison with the space-charge width at *CB* junction, as illustrated in Fig. 4.10.

The transition or space-charge region at a junction is the region of uncovered charges on both sides of the junction. As the voltage applied across the junction increases, the transition region penetrates deeper into the

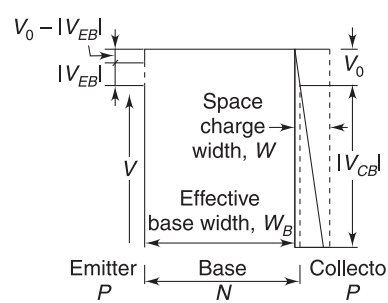


Fig. 4.10 *The potential variation through PNP transistor under biased conditions*

collector and base. Because neutrality of the charge is to be maintained, the number of uncovered charges on each side remains equal. Since the doping in the base is substantially smaller than that of collector, the penetration of the transition region into the base is much larger than that into the collector region. Hence the collector depletion region is neglected and all the immobile charge is indicated in the base region.

As obvious from Fig. 4.10, the effective base width is metallurgical base width less space charge width W. This *modulation of the effective base width* by the collector voltage V_{CB} is known as the *Early effect*.

An increase in magnitude of collector voltage increases space-charge width $W \left[\because V_B = \frac{eN_D W^2}{2\varepsilon}\right]$ at the output junction diode and this, in turn, causes the effective base width W_B to decrease.

The decrease in effective base width W_B with increasing reverse bias has three consequences (Fig. 4.11).

(i) There is less chance for recombination within the base region. Hence the transport factor β^* and also α, increase with an increase in the magnitude of *C-B* junction voltage.

(ii) The concentration gradient of minority carriers is increased within the base, and consequently, the current of minority carriers injected across the emitter junction increases, with increasing reverse collector voltage.

(iii) For extremely large voltages, W_B may be reduced to zero causing voltage breakdown. When effective base width W_B is reduced to zero, the emitter barrier becomes V', which is smaller than normal value $V_0 - |V_{EB}|$ because the collector voltage has reached through the base region. This lowering of the emitter-junction voltage causes an excessively large emitter current, thus placing an upper limit on the magnitude of the collector voltage. This phenomenon is known as *punch-through*.

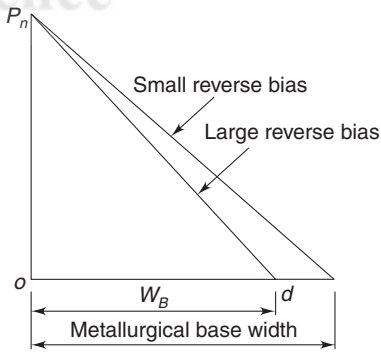


Fig. 4.11 *Minority carrier density within the base*

The punch-through (or reach-through) phenomenon may be defined as the mechanism by which a transistor usefulness may be terminated by increasing the collector voltage.

4.7 ALPHA (α)

In the d.c. mode, the levels of I_C and I_E due to the majority carriers are related by a quantity called *alpha* and defined by the following relation

$$\alpha_{dc} = \frac{I_C}{I_E} \tag{4.1}$$

where I_C and I_E are the levels of current at the point of operation. For practical devices the level of alpha typically extends from 0.90 to 0.998. Above, alpha is defined solely for the majority carriers, Eq. (4.1) becomes

$$I_C = \alpha\, I_E + I_{CO} \tag{4.2}$$

For a.c. situations, where the point of operation moves on the characteristic curve, an ac alpha is defined by

$$\alpha_{ac} = \left.\frac{\Delta I_C}{\Delta I_E}\right|_{V_{CB} = \text{const.}} \tag{4.3}$$

The ac alpha is formally called the *common-base, short circuit, amplification factor*. This equation specifies that a relatively small change in collector current is divided by the corresponding change in I_E with the collector-to-base voltage held constant. (For most situations, the magnitudes of α_{ac} and α_{dc} are quite close, permitting the use of the magnitude of one for the other).

4.8 THE TRANSISTOR COMMON BASE CIRCUIT AS AMPLIFIER

Figure 4.12 shows the basic common-base amplifier using *PNP* transistor. Without any signal applied, the emitter current is the d.c. bias current I_E. Most of this current is swept into the collector portion of the transistor, giving a collector current I_C and leaving only a small base current I_B. Thus, $I_E = I_B + I_C$ (as before), and when v_s is applied, this becomes $i_E = i_B + i_C$. Note that *E-B* junction is forward biased and *C-B* junction is reverse biased. Figure 4.13 shows the common-base amplifier equivalent circuit using the two diode analogy. Since the arrow on the emitter points inward, therefore, the two diodes will also point inward as shown in the Fig. 4.13.

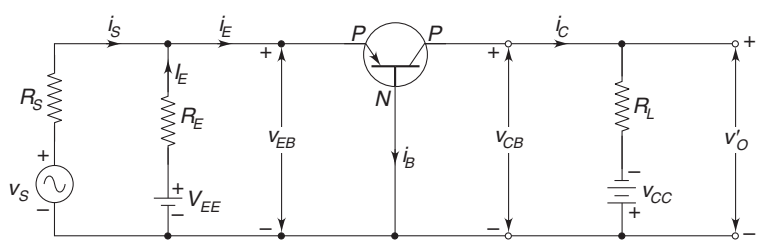


Fig. 4.12 *Basic common-base amplifier with PNP transistor*

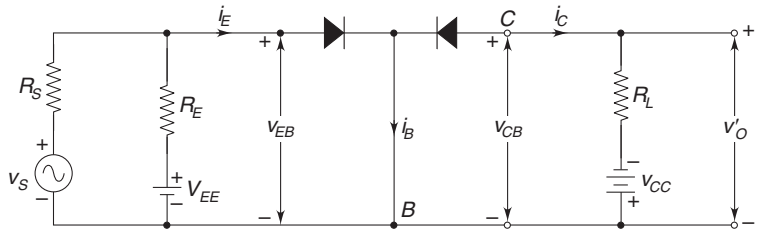


Fig. 4.13 *Using the two-diode analogy to represent a CB amplifier*

The output curves are nearly horizontal lines provided the collector-to-base junction is reverse biased. This indicates a high incremental output resistance of the order of 10^6 ohms. By contrast the incremental input resistance of the transistor is of the order of a few ohms.

In most of the circuit analysis, the current gain of the transistor is of greatest interest, because the transistor is a *current operated device*, and the voltages necessary to sustain suitable currents across the junctions are of secondary importance, e.g., the emitter current versus emitter-to-base voltage curve is only rarely plotted since it is very much temperature dependent $J = (eD_p\ p_{no})/W \times \exp eV_{EB}/kT$. The small emitter-to-base voltage required for forward operation of the junction is readily obtained practically, and interest is chiefly centered on the value of the emitter current.

Instead of going into a complete analysis of the operation of the transistor amplifier is attempted, it would be of interest to show the capabilities of the device under ac operation. This can be done by referring to Fig. 4.14 where a small alternating voltage generator V_e has been placed in series with the input circuit. Assuming that the alternating voltage V_e is much less than the direct emitter-to-base voltage, a small alternating current I_e will flow in the input circuit given by

$$V_e = I_e\ r_i \tag{4.4}$$

where r_i is the incremental input resistance under the given conditions. If the alternating current in the output circuit is I_C, the alternating output voltage across the load is

$$V_0 = I_C\ R_L \tag{4.5}$$

The current gain of the circuit is

$$A_i = \frac{I_C}{I_e} \tag{4.6}$$

and the voltage gain is

$$A_v = \frac{I_C\ R_L}{I_e\ r_i} \tag{4.7}$$

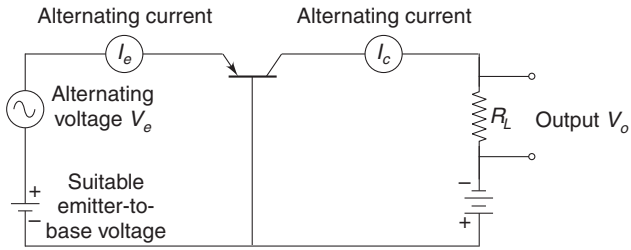


Fig. 4.14 *Inclusion of a small alternating voltage in the input of a CB transistor circuit*

Equations (4.4) through (4.7) apply to small-signal a.c. operation. It can be seen however, that the ratio of the alternating currents I_C and I_e is almost the same as the ratio of the direct currents measured under the same conditions since the mechanism of charge conveyance in the base region is the same at low frequencies as at zero frequency. When the load resistance R_L is zero, and the collector-to-base voltage is a constant, the current gain of the *CB* circuit is the α (alpha) of the transistor and has a value between 0.95 and 0.995 (This is the most important transistor parameter).

In the case shown in Fig. 4.14, we may approximately the condition $R_L = 0$ by $R_L < 0.1\ r_o$, where r_o is the incremental output resistance, since then, the voltage drop across R_L is a small proportion of the total alternating voltage in the collector circuit. Using the further approximately $\alpha = 1$, we have

$$A_v = \frac{R_L}{r_i} \tag{4.8}$$

Typical values would be

$$r_o = 1\ \text{M}\Omega, \quad R_L = 10{,}000\ \Omega, \quad r_i = 50\ \Omega$$

giving

$$A_v = \frac{10{,}000}{50} = 200$$

From the foregoing simplified analysis, the *CB* transistor circuit is characterized by

1. Current gain just less than unity
2. High voltage gain
3. Low input resistance
4. High output resistance

Note that a common-base amplifier exhibits a current gain A_i less than one. However, since the current in the output can flow through a load resistor that is larger than the input resistance, it is possible to have a voltage gain A_v, and therefore, a power gain $Ap = A_v A_i$. In fact, it is the ability of the device to *transfer* current from a low resistance input to a large *resistance* output that originated the name "*transistor*" as a contraction of the phrase "*transfer resistor*".

4.9 COMMON-EMITTER CONFIGURATION

The most frequently encountered transistor configuration appears in Fig. 4.15 for the *PNP* and *NPN* transistors. It is called the *common-emitter configuration* since the emitter is common or reference to both the input and output terminals (in this case common to both the base and collector terminals). Two sets of characteristics are again necessary to describe fully the behaviour of the common-emitter configuration: one for the *input* or *base-emitter* circuit and one for the *output* or *collector-emitter* circuit. Both are shown in Fig. 4.15.

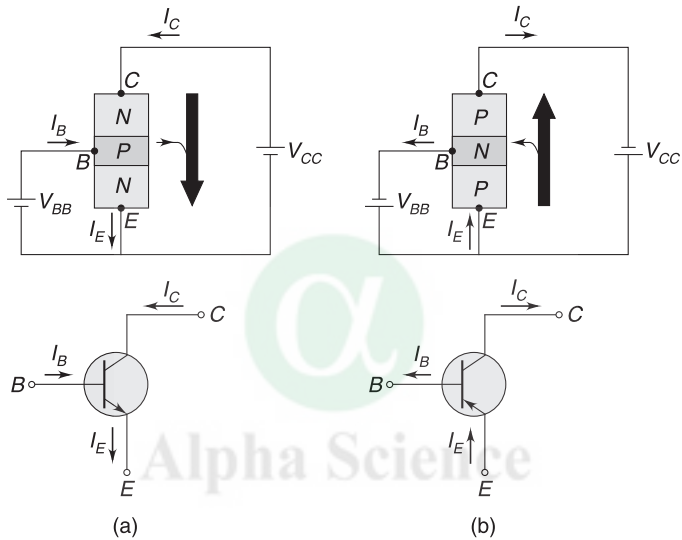


Fig. 4.15 *Notation and symbols used with the common-emitter configuration: (a) NPN transistor; (b) PNP transistor*

The emitter, collector, and base currents are shown in their actual conventional current direction. Even though the transistor configuration has changed, the current relations developed earlier for the common-base configuration are still applicable. That is, $I_E = I_C + I_B$ and $I_C = \alpha I_E$.

For the common-emitter configuration the output characteristics are a plot of the output current (I_C) versus output voltage (V_{CE}) for a range of values of input current (I_B). The input characteristics are a plot of the input current (I_B) versus the input voltage (V_{BE}) for a range of values of output voltage (V_{CE}).

Note that on the characteristics of Fig. 4.16 the magnitude of I_B is in microamperes, compared to milliamperes of I_C. Consider also that the curves of I_B are not as horizontal as those obtained for I_E in the common-base configuration, indicating that the collector-to-emitter voltage will influence the magnitude of the collector current.

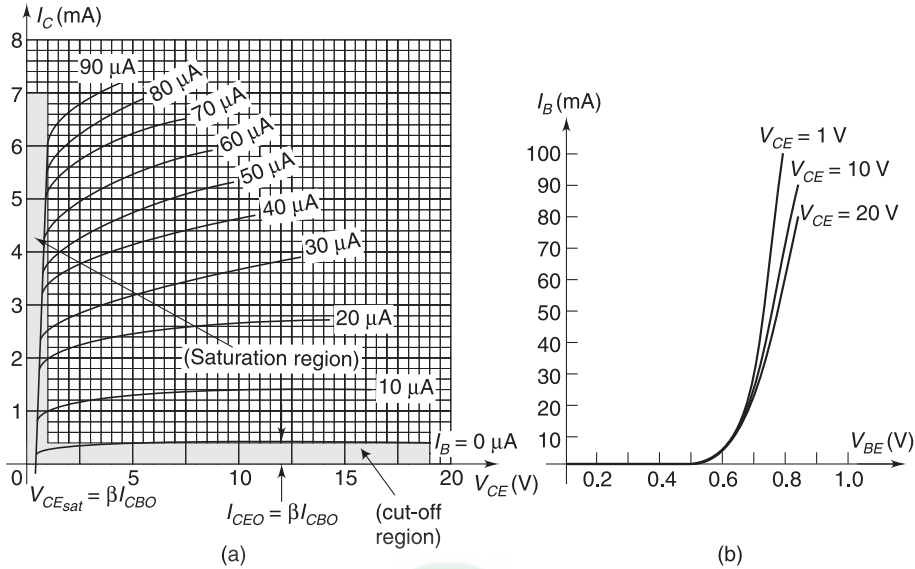


Fig. 4.16 *Characteristics of a silicon transistor in the common-emitter configuration: (a) collector characteristics; (b) base characteristics*

The active region for the common-emitter configuration is that portion of the upper-right quadrant that has the greatest linearity, that is, that region in which the curves for I_B are nearly straight and equally spaced. In Fig. 4.16(a), a this region exists to the right of the vertical dashed line at $V_{CE_{sat}}$ and above the curve for I_B equal to zero. The region to the left of $V_{CE_{sat}}$ is called the saturation region.

In the active region of a common-emitter amplifier the collector-base junction is reverse biased, while the base-emitter junction is forward biased.

You may recall that these were the same conditions that existed in the active region of the common-base configuration. *The active region of the common-emitter configuration can be employed for voltage, current, or power amplification.*

The cutoff region for the common-emitter configuration is not as well defined as for the common-base configuration. Note on the collector characteristics of Fig. 4.16 that I_C is not equal to zero when I_B is zero. For the common-base configuration, *when the input current I_E was equal to zero, the collector current was equal only to the reverse saturation current I_{CO}*, so that the curve I_E = 0and the voltage axis were, for all practical purposes, one.

The reason for this difference in collector characteristics can be derived through the proper manipulation of equation

$$I_C = \alpha I_E + I_{CBO} \tag{4.9}$$

That is

$$I_C = \alpha (I_C + I_B) + I_{CBO} \qquad (\because I_E = I_C + I_B) \tag{4.10}$$

Rearranging yields

$$I_C = \frac{\alpha I_B}{1 - \alpha} + \frac{I_{CBO}}{1 - \alpha} \tag{4.11}$$

If we consider the case discussed above, where $I_B = 0$ A, and substitute a typical value of α such as 0.996, the resulting collector current is the following:

$$I_C = \frac{\alpha(0\ A)}{1 - \alpha} + \frac{I_{CBO}}{1 - 0.996}$$

$$= \frac{I_{CBO}}{0.004} = 250 I_{CBO} \tag{4.12}$$

If I_{CBO} were 1 μA, the resulting collector current with $I_B = 0$ A would be 250 (1 μA) = 0.25 mA, as reflected in the characteristics of Fig. 4.16.

For future reference, the collector current defined by the condition $I_B = 0$ μA will be assigned the notation indicated by Eq. (4.13)

$$I_{CEO} = \left.\frac{I_{CBO}}{1 - \alpha}\right|_{I_B} = 0\ \mu A \tag{4.13}$$

In Fig. 4.17 the conditions surrounding this newly defined current are demonstrated with its assigned reference direction.

For linear (least distortion) amplification purposes, cut-off for the common-emitter configuration will be defined by $I_C = I_{CEO}$.

In other words, the region below $I_B = 0$ μA is to be avoided if an undistorted output signal is required.

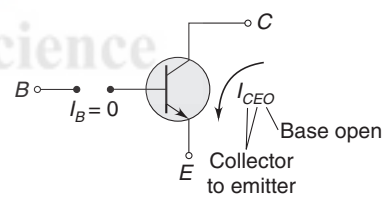


Fig. 4.17 *Circuit conditions related to I_{CEO}*

4.10 BETA (β)

In the dc mode, the levels of I_C and I_B are related by a quantity called beta and defined by the following relation

$$\beta_{dc} = \frac{I_C}{I_B}$$

where I_C and I_B are determined at a particular operating point on the characteristics. For practical devices the level of β typically ranges from about 50 to over 400, with most in the mid range. As for α, β certainly reveals the relative magnitude of one current to the other. For a device with a β of 200, the collector current is 200 times the magnitude of the base current.

On specification sheets β_{dc} is usually included as h_{FE} with the h derived from an ac hybrid equivalent circuit to be introduced in Chapter 6. The subscript FE are derived from forward-current amplification and common-emitter configuration, respectively.

For a.c. situations an a.c. beta (β) has been defined as follows:

$$\beta_{ac} = \left.\frac{\Delta I_C}{\Delta I_B}\right|_{V_{CE}} = \text{constant}$$

The formal name for β_{ac} is *common-emitter forward-current amplification factor*. On specification sheets β_{ac} is normally refered to as h_{fe}.

4.11 RELATIONSHIP BETWEEN β AND α

$$\beta = \frac{I_C}{I_B}, \qquad \alpha = \frac{I_C}{I_E}$$

$$I_E = I_C + I_B$$

$$\therefore \quad \frac{I_C}{\alpha} = I_C + \frac{I_C}{\beta}$$

$$\text{or} \quad \frac{1}{\alpha} = (I_C/I_C) + \left.\frac{I_C}{\beta}\right|I_C$$

$$\text{or} \quad \frac{1}{\alpha} = 1 + \frac{1}{\beta}$$

$$\text{or} \quad \beta = \alpha\beta + \alpha = (\beta + 1)\,\alpha$$

$$\text{so that} \quad \alpha = \frac{\beta}{\beta + 1} \tag{4.14}$$

$$\text{or} \quad \beta = \frac{\alpha}{1 - \alpha} \tag{4.15}$$

Further, we know that

$$I_{CEO} = \frac{I_{CBO}}{1 - \alpha} \tag{4.16}$$

From Eq. (4.14)

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

$$\text{or} \quad \frac{1}{1 - \alpha} = \beta + 1$$

$$\therefore \quad I_{CEO} = (\beta + 1)\, I_{CBO}$$

$$\text{or} \quad I_{CEO} \approx \beta\, I_{CBO} \tag{4.17}$$

Also, $$I_C = \beta I_B \quad (4.18)$$

and since $$I_E = I_C + I_B$$

$$= \beta I_B + I_B$$

so that $$I_E = (\beta + 1) I_B \quad (4.19)$$

Questions

4.1 Explain with diagrams, the operation of a *PNP* transistor.

4.2 Draw an *NPN* transistor in the *CB* configuration biased for operation in active region.

4.3 Sketch typical *CB* input characteristic curves for an *NPN* transistor. Label all variables.

4.4 Sketch typical output characteristic curves for a *PNP* transistor in *CB* configuration. Label all variables and indicate active, cut-off and saturation regions.

4.5 An *NPN* transistor is to be used in common-emitter configuration. Show how you will connect the external batteries so that the transistor works in the active region.

4.6 Sketch typical *CE* input characteristics for an *NPN* transistor. Label all variables.

4.7 Sketch typical *CE* output characteristic curves for an *NPN* transistor. Label all variables.

4.8 Derive the relationship between the beta (β) and alpha (α) of a transistor.

4.9 Explain the 'limits of operation" on transistor characteristics (i.e. distortion – free region of operation).

5 Transistor Biasing and Amplifiers

5.1 WHY BIAS A TRANSISTOR?

The purpose of d.c. biasing of a transistor is to obtain a certain d.c. collector current at a certain d.c. collector voltage. These values of current and voltage are expressed by the term *operating point* (or *quiescent point*). To obtain the operating point, we make use of some circuits; and these circuits are called "biasing circuits". Of course, while fixing the operating point, it has to be seen that it provides proper dc conditions so that the specific function of the circuit is achieved. The suitability of an operating point for the specific application of the circuit should be seen on the transistor characteristics. Generally, we discuss the suitability of the operating point with a view that the specific use of the circuit is in amplifiers.

5.2 THE LOAD LINE AND OPERATING POINT

A common-emitter amplifying stage is shown in Fig. 5.1. R_L is the load resistor into which the transistor is working. Now, writing in the conventional manner with terminal indicated by first subscript as positive, we have

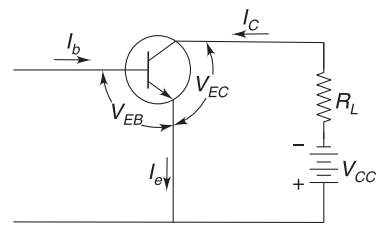


Fig. 5.1 *Simple common-emitter circuit*

$$V_{EC} = V_{CC} - I_C R_L \qquad (5.1)$$

Here V_{CC} and R_L are fixed values. Eqn. (5.1) is a first degree equation and can be represented by a straight line on the output characteristics

$$I_C = -\frac{V_{EC}}{R_L} + \frac{V_{CC}}{R_L}$$

$\left(\text{The intercept on } y\text{-axis is } \frac{V_{CC}}{R_L} \text{ and the slope of the line is } -\frac{1}{R_L}\right)$.

This is known as dc load line (Fig. 5.2). To draw the d.c. load line, we need two end points of the straight line. These can be located as follows:

(i) When collector current $I_C = 0$, the collector-emitter voltage is maximum and is equal to V_{CC}.

i.e., $V_{EC} = V_{CC} \quad (I_C = 0)$.

This gives the first point.

(ii) When collector-emitter voltage $V_{ec} = 0$, the collector current is maximum and is equal to V_{CC}/R_L.

i.e., $V_{EC} = V_{CC} - I_C R_L$,

$$0 = V_{CC} - I_C R_L$$

$$\therefore \quad I_C = \frac{V_{CC}}{R_L}$$

This gives the second point.

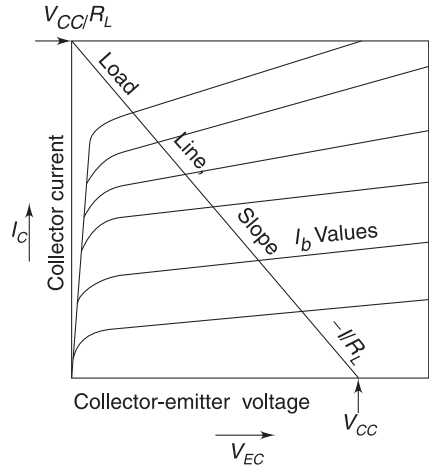


Fig. 5.2 *Load line plotted on output characteristics*

By joining these two points, d.c. load line may be drawn (Fig. 5.2).

Thus d.c. load line is a line on the output characteristics of a transistor circuit, which gives the values of I_C and V_{CE} corresponding to zero signal.

The zero signal values of collector current I_C and collector-emitter voltage V_{CE} are known as operating point. The point is called as operating point because the variations of I_C and V_{CE} take place about this point when signal is applied. The point is also named as *quiescent* (silent) *point* or Q-point as it is the point on $I_C - V_{CE}$ characteristics when the transistor is silent in the absence of signal (Fig. 5.3).

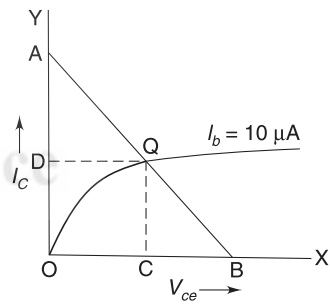


Fig. 5.3

5.3 SELECTION OF OPERATING POINT

In order that the circuit amplifies the signal properly, a judicious selection of the operation point is very necessary. The biasing arrangement should be such as to make the emitter-base junction forward-biased and the collector-base junction reverse-biased. Under such biasing, the transistor is said to operate in the *active region* of its characteristics. Various transistor ratings are to be kept in view while designing the biasing circuit. These ratings—specified by the manufacturer—limit the range of useful operation of the transistor and $V_{CE(\max)}$ is the maximum voltage that can be applied across it safely. In no case should these current and voltage limits be crossed.

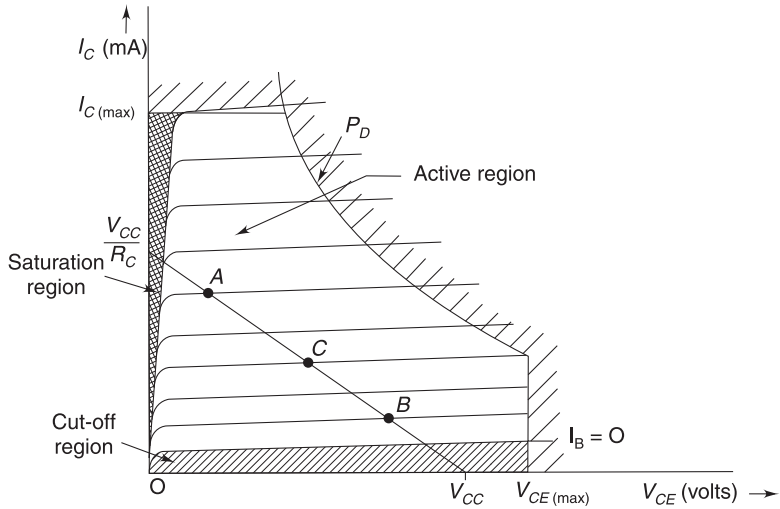


Fig. 5.4 *Output characteristics of a transistor in common-emitter configuration. Maximum current, voltage and power ratings are indicated*

If a transistor is to work as an amplifier, a load resistance R_C must be connected in the collector circuit. Only then the output a.c. signal voltage can develop across it. The d.c. load line corresponding to this resistance R_C and a given collector supply V_{CC} is shown in Fig. 5.4. The operating point will necessarily lie somewhere on this load line. Depending upon the base current, the operating point could be either at point A, B or C. Let us now consider which one of these is the most suitable operating point.

After the d.c. (or static) conditions are established in the circuit, an a.c. signal voltage is applied to the input. Due to this voltage, the base current varies from instant to instant. As a result of this, the collector current and the collector voltage also vary with time. That is how an amplified ac signal is available at the output. The variations in collector current and collector voltage corresponding to a given variation (which may be assumed sinusoidal) of base current can be seen on the output characteristics of the transistor.

These variations are shown in Figs. 5.5, 5.6 and 5.7 for the operating points A, B and C, respectively. In Fig. 5.5, point A is very near to the saturation region. Even though the base current is varying sinusoidally, the output current (and also output voltage) is seen to be clipped at the positive peaks. This results in *distortion* of the signal. At the positive peaks, the base current varies, but collector current remains constant at saturation value. Thus we see that point A is not a suitable operating point.

In Fig. 5.6, the point B is very near to the cut-off region. The output signal is now clipped at the negative peaks. Hence, this too is not a suitable operating point.

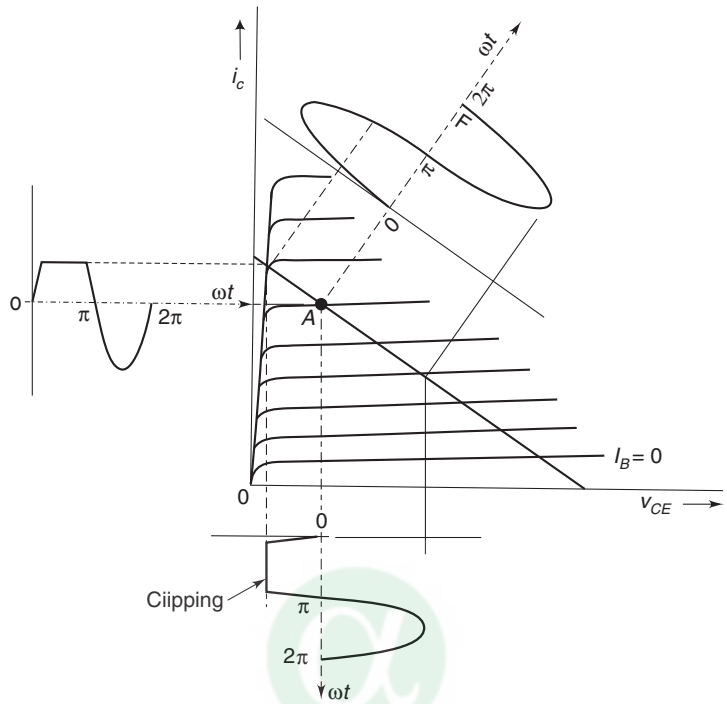


Fig. 5.5 *Operating point near saturation region given clipping at the positive peaks*

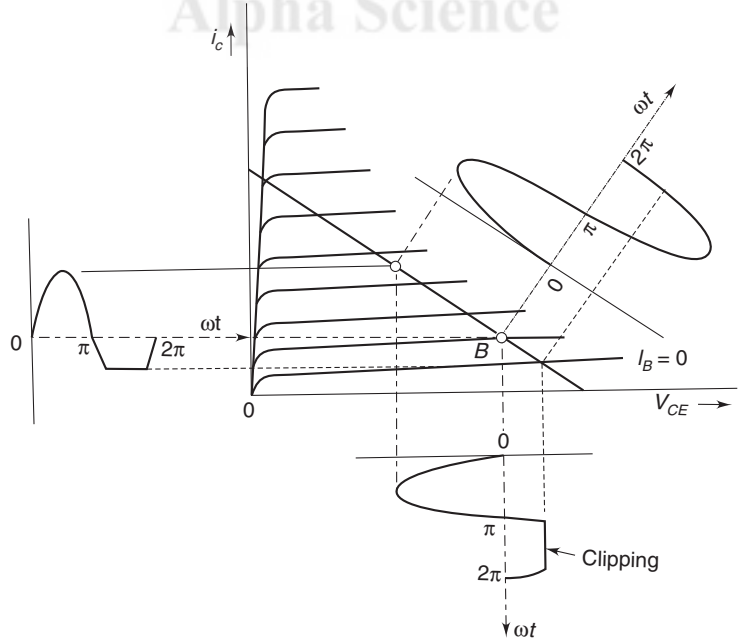


Fig. 5.6 *Operating point near cut-off region gives clipping at the negative peaks*

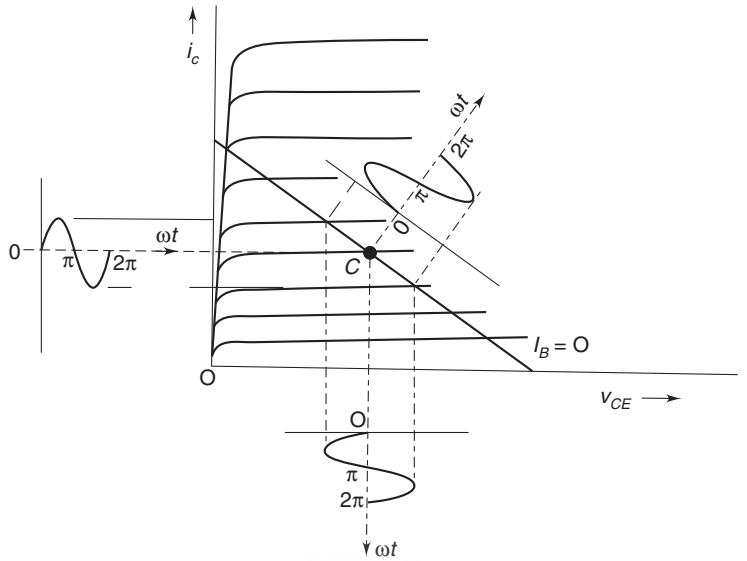


Fig. 5.7 *Operation point at the centre of active region is most suitable*

It is clear from Fig. 5.7 that the output signal is not at all distorted if point *C* is chosen as the operating point. A good amplifier amplifies signals without introducing distortion, as much as possible. Thus, point *C* is the most suitable operating point.

Even for the operating point *C*, distortion can occur in the amplifier if the input signal is large. Then output current and output voltage is clipped at both the positive and the negative peaks. Thus, the maximum signal that can be handled by an amplifier is decided by the choice of the operating point.

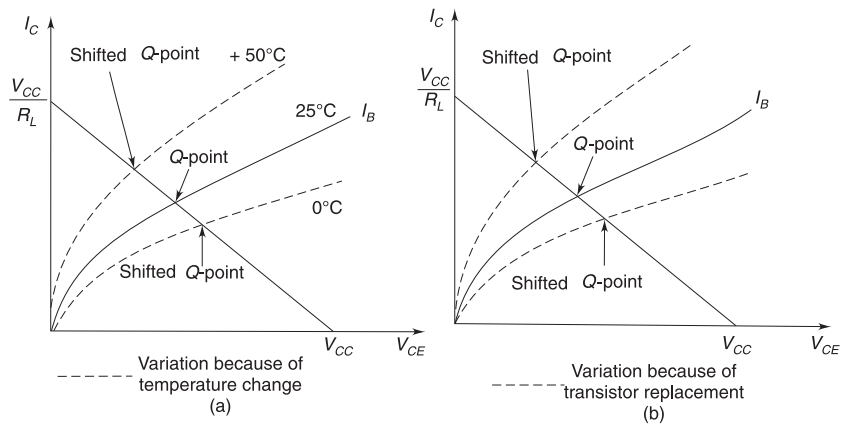


Fig. 5.8 *Shifting of the Q-point with (a) variations in temperature and (b) replacement of the transistor*

5.4 INHERENT VARIATIONS OF TRANSISTOR PARAMETERS

In practice, the transistor parameters such as β, V_{BE} are not the same for every transistor even of the same type. For example, *BC* 147 is a silicon *NPN* transistor with β varying from 100 to 600 i.e., β for one transistor may be 100 and for the other, it may be 600, although both of them are *BC* 147. The major reason for these variations is that manufacturing techniques yet are not much advanced. For instance, it has not been possible to control the base width and it may vary, although slightly, from one transistor to the other (even of the same type). Such small variations result in large change in transistor parameters such as β and V_{BE} etc.

The inherent variations of transistor parameters may change the operating point, resulting in unfaithful amplification. It is, therefore, very important that biasing network be so designed that it should be able to work with all transistors of one type whatever may be the spread in β or V_{BE}. In other words, the operating point should be independent of transistor parameters variations.

5.5 STABILIZATION

The collector current in a transistor changes rapidly when;

(i) the temperature changes

(ii) the transistor is replaced by another of the same type (This is due to inherent variations of transistor parameters).

When the temperature changes or the transistor is replaced, the operation point (i.e. zero signal I_C and V_{CE}) also changes (Figs. 5.8(a) and (b). However, for faithful amplification, it is essential that operation point remains fixed. This necessitates to make the operation point independent of these variations. This is known as stabilization.

The process of making operating point independent of temperature changes or variations in transistor parameters is known as stabilization. A good biasing circuit always ensures the stabilization of operating point.

Need for Stabilization Stabilization of the operating point is necessary due to the following factors:

(i) Temperature dependence of I_C

(ii) Individual variations.

(iii) Thermal runaway

(i) Temperature dependence of I_C The collector current I_C for *CE* circuit is given by

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

The collector leakage current I_{CBO} is greatly influenced (especially in Ge transistor) by temperature changes. A rise of 10°C doubles the collector leakage current which may be as high as 0.2 mA for low powered Ge transistors. As biasing conditions in such transistors are generally so set that zero signal I_C = 1 mA, therefore, the change in I_C due to temperature variations cannot be tolerated. This necessitates to stabilize the operating point i.e., to hold I_C constant inspite of temperature variations.

(ii) Individual Variations The value of β and V_{BE} are not exactly the same for any two transistors even of the same type. Further V_{BE} itself decreases when temperature increases. When a transistor is replaced by another of the same type, these variations change the operation point.

It is to be noted that

1. I_{CBO} approximately doubles for every 10°C rise in temperature for both Si and Ge transistors.
2. The current gain β of a transistor in CE mode increases with an increase in temperature. The β increases at the rate of approximately 0.6%/°C.
3. The base-emitter junction voltage V_{BE} of a transistor decreases with increasing temperature at the rate of 2.5 mV/°C (Linear relationship is obeyed over a wide temperature range.

Tables 5.1 and 5.2 illustrate the typical values of the transistor's parameters at different constant temperatures.

Table 5.1 *Si transistor parameters*

Temp (°C)	I_{CBO} (nA)	β	V_{BE} (V)
– 65°C	1.95×10^{-3}	25	0.78
+ 25°C	1.0	55	0.60
+175°C	33,000	100	0.225

(iii) Thermal Runaway The collector current for a CE configuration is given by

Table 5.2 *Ge transistor parameters*

Temp (°C)	I_{CBO} (μA)	β	V_{BE} (V)
– 65°C	1.95×10^{-3}	20	0.38
+ 25°C	1.0	55	0.20
+175°C	32	90	0.10

$$I_C = \beta I_B + (\beta + 1) I_{CBO} \tag{5.1}$$

The collector leakage current I_{CBO} is strongly dependent on temperature. The flow of collector current produces heat within the transistor. This raises the transistor temperature and if no stabilization is done, the collector leakage

current I_{CBO} also increases. It is clear from eqn. (1) that if I_{CBO} increases the collector current I_C increases by $(\beta + 1)\ I_{CBO}$. The increased I_C will raise the temperature of the transistor, which in turn will cause I_{CBO} to increase. This effect is cumulative and in a matter of seconds, the collector current may become very large, causing the transistor to burn out. This self-destruction of a unstabilized transistor is known as *thermal runaway*.

In order to avoid thermal runaway and consequent destruction of transistor, it is very essential that operation point is stabilized, i.e., I_C is kept constant. In practice, this is done by causing I_B to decrease automatically with temperature increase by circuit modification. Then decrease in $\beta\ I_B$ will compensate for the increase in $(\beta + 1)\ I_{CBO}$, keeping I_C nearly constant. In fact, it is always aimed at while designing a biasing circuit.

5.6 ESSENTIALS OF A TRANSISTOR BIASING CIRCUIT

We already know that transistor biasing is required for faithful amplification. The biasing should meet the following requirements:

(i) It should ensure proper zero signal collector current.

(ii) It should ensure that V_{CE} does not fall below 0.5 V for Ge transistors and 1 V for Si transistors at any instant.

(iii) It should ensure the stabilization of the operating point.

5.7 STABILITY FACTOR

It is desirable and necessary to keep I_C constant in the face of variations of I_{CBO} (sometimes represented as I_{CO}). The extent to which a biasing circuit is successful in achieving this goal is measured by stability factor S. It is defined as follows:

The rate of change of collector current I_C with respect to the collector leakage current I_{CO} at constant β and I_B is called *stability factor* i.e.,

$$\text{Stability factor} \qquad S = \frac{dI_C}{dI_{CO}} \quad \text{at constant } I_B \text{ and } \beta.$$

The stability factor indicates the change in collector current I_C due to the change in collector leakage current I_{CO}. In order to achieve greater thermal stability, it is desirable to have as low stability factor as possible. The ideal value of S is one but it is never possible to achieve it in practice. Experience shows that values of S exceeding 25 result in unsatisfactory performance.

The general expression of stability factor for a CE configuration can be obtained as follows:

$$I_C = \beta\ I_B + (\beta + 1)\ I_{CO} \tag{5.2}$$

(where I_{CO} is collector leakage current in CB arrangement).

Differentiating the above expression with respect to I_C, we get

$$1 = \beta \frac{dI_B}{dI_C} + (\beta + 1) \frac{dI_{CO}}{dI_C} \tag{5.3}$$

or

$$1 = \beta \frac{dI_B}{dI_C} + \frac{(\beta + 1)}{S} \tag{5.4}$$

or

$$S = \frac{\beta + 1}{1 - \beta \left(\frac{dI_B}{dI_C} \right)} \tag{5.5}$$

5.8 METHODS OF TRANSISTOR BIASING

In the transistor amplifier circuits, biasing can be due with the aid of a battery V_{BB} which is separate from the battery V_{CC} used in the output circuit. However, in the interest of simplicity and economy, it is desirable that transistor circuit should have a single source of supply – the one in the output circuit (i.e. V_{CC}). The following are the most commonly used methods of obtaining transistor biasing from one source of supply (i.e. V_{CC}):

(i) Base resistor method
(ii) Biasing with feedback resistor
(iii) Bias circuit with emitter resistor
(iv) Voltage-divider bias

In all these methods, the same basic principle is employed i.e., required value of base current (and hence I_C) is obtained from V_{CC} in the zero signal conditions. The value of collector load R_C is selected keeping in view that V_{CE} should not fall below 0.5 V for germanium transistors and 1 V for silicon transistors.

For example, if $\beta = 100$ and the zero signal collector current I_C is to be set at 1 mA, then I_B is mad equal to $I_C/\beta = 1/100 = 10$ μA. Thus, the biasing network should be so designed that a base current of 10 μA flows in the zero signal conditions.

5.9 BASE RESISTOR METHOD

In this method, a high resistance R_B (several hundred kΩ) is connected between the base and +ve end of supply for *NPN* transistor (Fig. 5.9) and between base and negative end of supply for *PNP* transistor. Here the required zero signal base current is provided by V_{CC} and it flows through R_B.It is because now base is positive w.r.t. emitter i.e., base – emitter junction is forward biased. The required value of zero signal base current I_B (and hence $I_C = \beta I_B$) can be mad to flow by selecting the proper value of base resistor R_B.

Circuit Analysis It is required to find the value of R_B so that required collector current flows in the zero signal conditions. Let I_C be the required zero signal collector current.

$$I_B = \frac{I_C}{\beta} \tag{5.6}$$

Considering the closed circuit *ABENA* and applying Kirchhoff's voltage law, we get

$$V_{CC} = I_B R_B + V_{BE} \tag{5.7}$$

or $$I_B R_B = V_{CC} - V_{BE} \tag{5.8}$$

$$\therefore \quad R_B = \frac{V_{CC} - V_{BE}}{I_B} \tag{5.9}$$

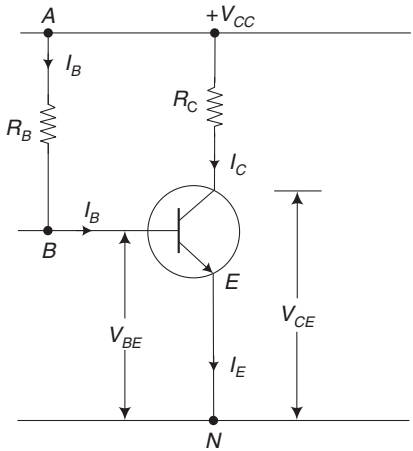


Fig. 5.9

As V_{CC} and I_B are known and V_{BE} can be seen from the transistor manual, therefore, value of R_B can be readily found from Eq. (5.9).

Since V_{BE} is generally quite small as compared to V_{CC}, the former can be neglected with little error. If then follows from Eq. (5.9) that

$$R_B = \frac{V_{CC}}{I_B} \tag{5.10}$$

It may be noted that V_{CC} is a fixed known quantity and I_B is chosen at some suitable value. Hence R_B can always be found directly, and for this season, this method is sometimes called *fixed bias method*.

The stability factor is given by

$$S = \frac{\beta + 1}{1 - \beta \left(\frac{dI_B}{dI_C} \right)} \tag{5.11}$$

In fixed-bias method of biasing, I_B is independent of I_C so that $dI_B/dI_C = 0$. Putting this in the above expression, we have

$$S = \beta + 1 \tag{5.12}$$

Thus the stability factor in a fixed bias is $(\beta + 1)$. This means that I_C changes $(\beta + 1)$ times as much as any change in I_{CO}.

Advantages

(i) This biasing circuit is very simple as only one resistance R_B is required.

(ii) Biasing conditions can easily be set and the calculations are simple.

(iii) There is no loading of the source by the biasing circuit since no resistor is employed across base-emitter junction.

Disadvantages

(i) This method provides poor stabilization. It is because there is no means to stop a self increase in collector current due to temperature rise and individual variations. For example, if β increases due to transistor replacement, then I_C also increases by the same factor as I_B is constant.

(ii) The stability factor is very high. Therefore, there are strong chances of thermal runaway. Due to these disadvantages, this method of biasing is rarely employed.

Example 5.1 *Stability of the fixed base bias circuit* (Fig. 5.10).

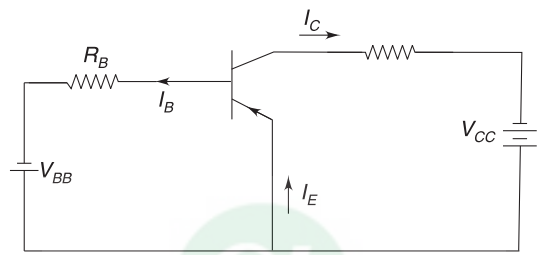


Fig. 5.10

Solution

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

Stability factor

$$S = \frac{\partial I_C}{\partial I_{CBO}} = (\beta + 1) \tag{ii}$$

I_{CBO} varies more widely with temperature than any other quantity, so stabilizing w.r.t. I_{CBO} will adequately stabilize against temperature variations in β. However, β may be subject to variations (irrespective of changes in I_{CBO}) e.g., when a transistor is replaced. Then, we define

$$S_h = \frac{\partial I_C}{\partial \beta} = I_B + I_{CBO}$$

$$= \frac{I_C - (\beta + 1) I_{CBO}}{\beta} + I_{CBO} \quad \text{(using Eq. (i))}$$

or

$$S_h = \frac{I_C - I_{CBO}}{\beta} \tag{iv}$$

Considering small changes ΔI_C and $\Delta\beta$ in I_C and β respectively, we may write

$$\frac{\Delta I_C}{\Delta \beta} = \frac{I_C}{\beta}\left(1 - \frac{I_{CBO}}{I_C}\right) \tag{v}$$

or $$\frac{\Delta I_C}{I_C} = \frac{\Delta\beta}{\beta}\left(1 - \frac{I_{CBO}}{I_C}\right) \qquad \text{(vi)}$$

5.10 BIASING WITH FEEDBACK RESISTOR

In this method, one end of R_B is connected to the base and the other end to the collector as shown in Fig. 5.11. Here, the required zero signal base current is determined not by V_{CC} but by the collector-base voltage V_{CB}. It is clear that V_{CB} forward biases the base-emitter junction and hence base current I_B flows through R_B. This causes the zero signal collector current to flow in the circuit.

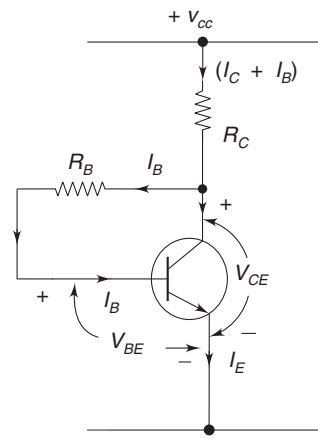


Fig. 5.11

Circuit Analysis The required value of R_B needed to give the zero signal current I_C can be determined as follows. Referring to Fig. 5.11,

$$V_{CC} = (I_B + I_C)\, R_C + I_B\, R_B + V_{BE} \qquad (5.13)$$

(Actual voltage drop across R_C is $(I_B + I_C)\, R_C$ but $I_B << I_C$, therefore it is $\approx I_C\, R_C$).

or $$R_B = \frac{V_{CC} - V_{BE} - I_C\, R_C}{I_B}$$

$$= \frac{V_{CC} - V_{BE} - \beta\, I_B\, R_C}{I_B}$$

Alternatively,

$$V_{CE} = V_{BE} + V_{CB}$$

or $$V_{CB} = V_{CE} - V_{BE}$$

$\therefore$ $$R_B = \frac{V_{CB}}{I_B} = \frac{V_{CE} - V_{BE}}{I_B}, \quad \left(\text{where } I_B = \frac{I_C}{\beta}\right) \qquad (5.14)$$

It can be shown mathematically that stability factor

(using $I_C = \beta\, I_B + (\beta + 1)\, I_{CBO}$) $< (\beta + 1)$ (5.15)

Let us now see what happens when the temperature rises.

Suppose the temperature increases, causing the leakage current (and also β) to increase. This increases the collector current (since $I_c + \beta I_B + I_{CEO}$). As the collector current increases, the voltage V_{CE} decreases (since $V_{CE} = V_{CC} - I_C\, R_C$). As can be seen from Eq. (5.14), the reduced V_{CE} causes decrease in base current I_B. The lowered base current in turn reduces the original increase in collector current. Thus, a mechanism exists in the circuit because of which the collector current is not allowed to increase rapidly. There is a tendency in the circuit to stabilize the operating point (Fig. 5.12).

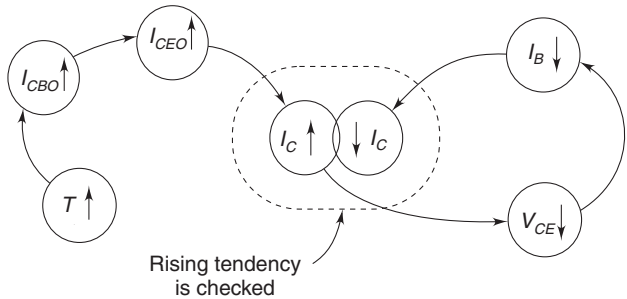


Fig. 5.12 *Collector-to-base bias circuit checks the rising tendency of collector current*

Note that the resistor R_B connects the collector (the output) with the base (the input). This means that a feedback exists in the circuit. The base current is dependent on the collector voltage. And this dependence is such as to nullify the changes in base current. That is why this circuit is also called a *voltage feedback* bias circuit.

Suppose the transistor in this circuit is replaced by another having different value of β. The shift in the operating point will not be as much as it occurs in case of fixed bias circuit. This can be seen as follows.

For determining the operating point, we substitute βI_β for I_C in Eq. (5.13) to get

$$V_{CC} = R_C\, \beta I_\beta + (R_C + R_B)\, I_B + V_{BE}$$

or
$$V_{CC} = V_{BE} + [R_B + (\beta + 1)\, R_C]\, I_B$$

or
$$I_B = \frac{V_{CC} - V_{BE}}{R_B + (\beta + 1)\, R_C} \approx \frac{V_{CC}}{R_B + \beta\, R_C} \qquad (5.16)$$

(as compared to $I_B = V_{CC} - V_{BE}/R_B$ for fixed bias circuit)

Why collector-to-base bias circuits is seldom used? This circuit has a tendency to stabilize the operating point against temperature and β variations. But the circuit is not used very much. The resistor R_B not only provides a d.c. feedback for the stabilization of operating point, but it also causes an ac feedback. This reduces the voltage gain of the amplifier. It is not desirable. After all, the biasing of a transistor was needed so that it could amplify the ac signals properly. Because of this drawback, the circuit is not very commonly used.

Example 5.2 *Given the device characteristic of Fig. 5.13(a), determine V_{CC}, R_B and R_C for the fixed-bias configuration of Fig. 5.13(b).*

Solution From the load line $V_{CC} = 20$ V

$$I_C = \left.\frac{V_{CC}}{R_C}\right|_{V_{CE} = 0\text{ V}}$$

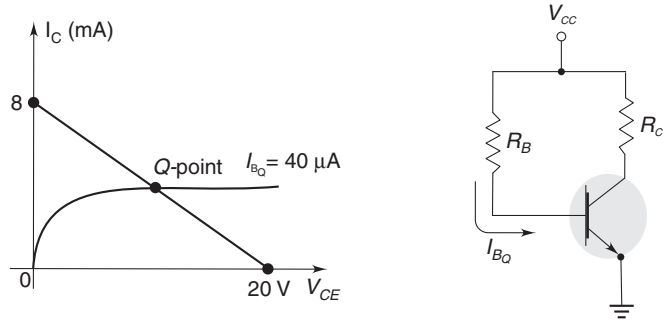


Fig. 5.13

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

$$I_B = \frac{V_{CC} - V_{BE}}{R_B}$$

with
$$R_B = \frac{V_{CC} - V_{BE}}{I_B}$$

$$= \frac{20\text{ V} - 0.7\text{ V}}{40\ \mu\text{A}} = \frac{19.3\text{ V}}{40\ \mu\text{A}}$$

$$= 482.5\text{ k}\Omega$$

Standard resistor values:

$$R_C = 2.5\text{ k}\Omega$$

$$R_B = 470\text{ k}\Omega$$

Using standard resistor values gives

$$I_B = 41.1\ \mu\text{A}$$

which is well within 5% of the value specified.

5.11 BIAS CIRCUIT WITH EMITTER RESISTOR

We can modify the fixed bias circuit by connecting a resistor to the emitter terminal (as shown in Fig. 5.14). In this circuit, we have three resistors R_C, R_B and R_E and a battery V_{CC}.

We shall now see what happens to the Q point when the temperature increases. For this, we write the loop equation for the input section of the circuit.

$$V_{CC} = R_B I_B + V_{BE} + I_E R_E \tag{5.17}$$

or $$I_B = \frac{V_{CC} - I_E R_E - V_{BE}}{R_E} \quad (5.18)$$

or $$I_B \approx \frac{V_{CC} - I_E R_E}{R_B}$$

(since V_{BE} is very small) (5.19)

As the temperature tends to increase, the following sequence of events occur (Fig. 5.15).

Because of the temperature rise, the leakage current increases. This increases the collector current as well as the emitter current. As a result, the voltage drop across resistor R_E also increases. This reduces the numerator of Eq. (5.19) and hence the current I_B also reduces. This results in reduction of the collector current. Thus we see that the collector current is not allowed to increase to the extent it would have been in the absence of the resistor R_E.

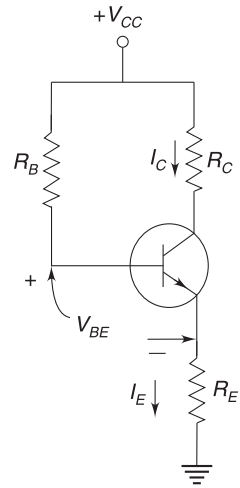


Fig. 5.14 *Bias circuit with emitter resistor*

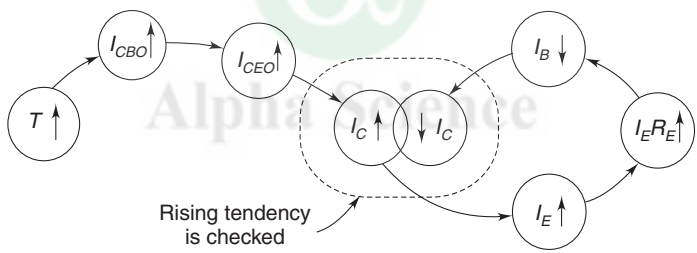


Fig. 5.15

In case the transistor is replaced by another of the same type (which may have different value of β), then also this current provides stabilization of the Q point, as is shown in Fig. 5.16.

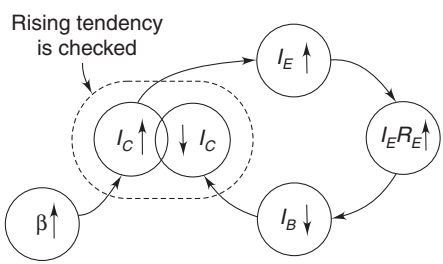


Fig. 5.16

Having seen that the operating point is stable in this circuit, let us determine the Q point. To do this, let us rewrite Eq. (5.17) as

$$V_{CC} = I_B R_B + V_{BE} + (\beta + 1) I_B R_E \tag{5.20}$$

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

or

$$I_B = \frac{V_{CC} - V_{BE}}{R_B + (\beta + 1) R_E} \approx \frac{V_{CC}}{R_B + \beta R_E} \tag{5.21}$$

We can calculate then collector current easily, since

$$I_C = \beta I_B = \frac{\beta V_{CC}}{R_B + \beta R_E} = \frac{V_{CC}}{R_E + (R_B/\beta)} \tag{5.22}$$

To find V_{CE}, we write the loop equation for the output section,

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

$$V_{CE} = V_{CC} - (R_C + R_E) I_C \quad (\text{since } I_C \approx I_E) \tag{5.23}$$

Operating point is thus determined by Eqs. (5.22) and (5.23).

The resistor R_E is present in the output side as well as in the input side of the circuit. A feedback occurs through this resistor. The feedback voltage is proportional to the emitter current. Hence, this circuit is also called *current feedback* biasing circuit. While the dc feedback helps in the stabilization of the Q point, the ac feedback reduces the voltage gains of the amplifier again, an undesirable feature. Of course, this drawback can be remedied by putting a capacitor C_E across the resistor R_E, as shown in Fig. 5.17. The capacitor C_E offers very low impedance to the ac current. The emitter is effectively placed at ground potential for the a.c. signal. The circuit provides dc feedback for the stabilization of the Q point, but does not give any a.c. feedback. The process of amplification of the ac signal remains unaffected.

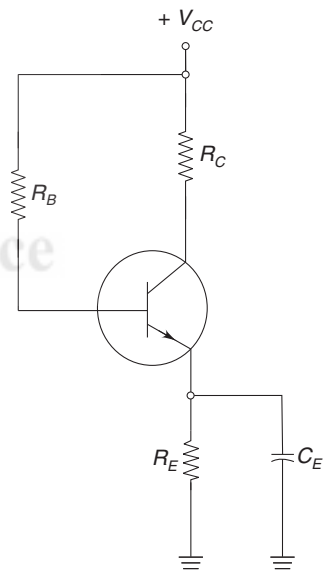


Fig. 5.17

Why this circuit is not used? The circuit in Fig. 5.14 does provide some stabilization of the Q point. But as you can see from Eq. (5.22), the denominator can be independent of β only if

$$R_E >> \frac{R_B}{\beta}$$

This means we should either have a very high value of R_E or a very low value of R_B. A high value of R_E will cause a large dc drop across it. To obtain a particular operating point under this condition, it will require a high dc source V_{CC}. On the other hand, if R_B is low, a separate low voltage supply has to be used in the base circuit. Both the alternatives are quite impractical. We should, therefore, look for some better circuit.

5.12 VOLTAGE DIVIDER BIAS

This is the most widely used method of providing biasing and stabilization to a transistor. In this method, two resistances R_1 and R_2 are connected across the supply voltage V_{CC} (Fig. 5.18), and provide biasing.

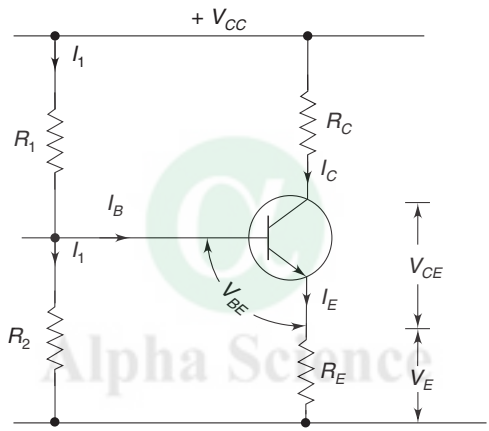


Fig. 5.18

The emitter resistance R_E provides stabilization. The name "voltage divider" comes from the voltage divider formed by R_1 and R_2. The voltage drop across R_2 forward biases the emitter junction. This causes the base current and hence collector current flow in the zero signal conditions.

Circuit Analysis: Suppose that the current flowing through resistance R_1 is I_1. As base current I_B is very small, therefore, it can be assumed with reasonable accuracy that current flowing through R_2 is also I_1.

(i) Collector current I_C

$$I_1 = \frac{V_{CC}}{R_1 + R_2}$$

$\therefore$ Voltage across R_2 is

$$V_2 = \left(\frac{V_{CC}}{R_1 + R_2}\right) R_2$$

Applying Kirchhoff's voltage law to the base circuit of Fig. 5.18, we have

$$V_2 = V_{BE} + V_E = V_{BE} + I_E R_E \tag{5.24}$$

or $$I_E = \frac{V_2 - V_{BE}}{R_E}$$

Since $I_E \approx I_C$, therefore,

$$I_C = \frac{V_2 - V_{BE}}{R_E} \tag{5.25}$$

It is clear from Eq. (5.25) above that I_C does not at all depend on β. Though I_C depends upon V_{BE} but in practice $V_2 >> V_{BE}$ so that I_C is practically independent of transistor parameters and hence good stabilization is ensured. It is due to this reason that potential divider bias has become universal method for providing transistor biasing.

(ii) Collector-emitter Voltage V_{CE}. Applying Kirchhoff's voltage law to the collector side,

$$\begin{aligned} V_{CC} &= I_C R_C + V_{CE} + I_E R_E \\ &= I_C R_C + V_{CE} + I_C R_E \quad (\because I_E \approx I_C) \\ &= I_C (R_C + R_E) + V_{CE} \end{aligned}$$

$$\therefore \quad V_{CE} = V_{CC} - I_C (R_C + R_E)$$

Stabilization In this circuit, excellent stabilization is provided by R_E. Consideration of Eq. (5.24) reveals this fact.

$$V_2 = V_{BE} + I_C R_E$$

Suppose the collector current I_C increases due to rise in temperature. This will cause the voltage drop across emitter resistance R_E to increase. As voltage drop across R_2 (i.e., V_2) is independent of I_C, therefore, V_{BE} decreases. This in turn causes I_B to decrease. The reduced value of I_B tends to restore I_C to the original value.

Stability Factor It can be shown mathematically that stability factor of the circuit is given by:

Stability factor,

$$S = \frac{(\beta + 1)(R_T + R_E)}{R_T + R_E + \beta R_C}$$

$$= (\beta + 1) \times \frac{1 + \dfrac{R_T}{R_E}}{\beta + 1 + \dfrac{R_T}{R_E}}$$

where $$R_T = \frac{R_1 R_2}{R_1 + R_2}$$

If the ratio R_T/R_E is very small, then R_T/R_E can be neglected as compared to 1 and the stability factor becomes:

$$\text{Stability factor} = (\beta + 1) \times \frac{1}{\beta + 1} = 1$$

This is the smallest possible value of S and leads to the maximum possible thermal stability. Due to design consideration, R_T/R_E has a value that cannot be neglected as compared to 1. In actual practice, the circuit may have stability factor around 10.

5.13 ANALYSIS OF VOLTAGE-DIVIDER BIAS USING THEVENIN THEOREM

The input side of the network of Fig. 5.18 can be redrawn as shown in Fig. 5.19 for the d.c. analysis. The Thevenin equivalent network for the network to the left of the base terminal can then be found as follows:

R_{Th} The voltage source is replaced by a short circuit equivalent as shown in Fig. 5.20.

$$R_{Th} = R_1 \parallel R_2$$

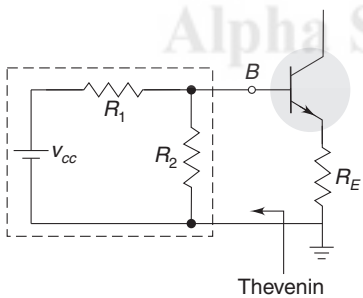


Fig. 5.19 *Redrawing the input side of the network of Fig. 5.18*

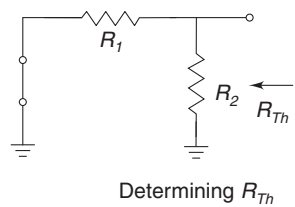


Fig. 5.20 *Determining R_{Th}*

E_{Th} The voltage source V_{CC} is returned to the network and the open-circuit Thevenin voltage of Fig. 5.21 determined as follows:

Applying the voltage-divider rule

$$E_{Th} = V_{R_2} = \frac{R_2 V_{CC}}{(R_1 + R_2)} \qquad (5.26)$$

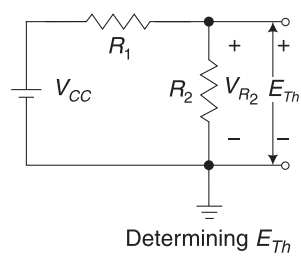


Fig. 5.21 *Determining E_{Th}*

The Thevenin network is then redrawn as shown in Fig. 5.22 and I_{B_2} can be determined by first applying the KVL in the clockwise direction for the loop indicated:

$$E_{Th} - I_B R_{Th} - V_{BE} - I_E R_E = 0 \qquad (5.27)$$

Substituting $I_E = I_C + I_B = (\beta + 1)\, I_B$ and solving for I_B yields

$$I_B = \frac{E_{Th} - V_{BE}}{R_{Th} + (\beta + 1)\, R_E} \qquad (5.28)$$

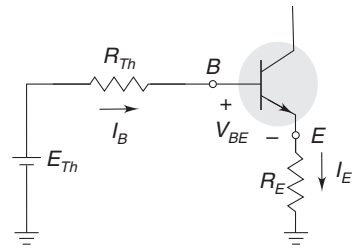


Fig. 5.22 *Inserting the Thevenin equivalent circuit*

The numerator in Eq. (5.28) is a difference of two voltage levels and the denominator is the base resistance plus the emitter resistor reflected by $(\beta + 1)$.

The other quantities of the network are same as for the emitter-bias configuration i.e.

$$V_{CE} = V_{CC} - I_C\,(R_C + R_E) \qquad (5.29)$$

$$V_E = I_E R_E \qquad (5.30)$$

$$V_C = V_{CE} + V_E \qquad (5.31)$$

or $$V_C = V_{CC} - I_C R_C \qquad (5.32)$$

$$V_B = V_{CC} - I_B R_B \qquad (5.33)$$

or $$V_B = V_{BE} + V_E \qquad (5.34)$$

Example 5.3 *Given that $I_{C_Q} = 2$ mA and $V_{CEQ} = 10$ V, determine R_1 and R_C for the network of Fig. 5.23.*

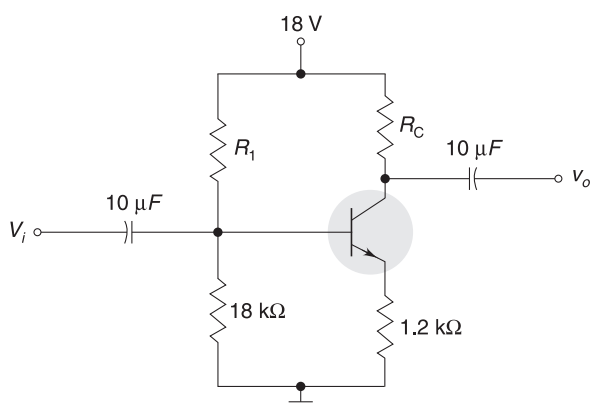


Fig. 5.23 *Example 5.3*

Solution

$$V_E = I_E R_E \approx I_C R_E$$
$$= (2 \text{ mA})\ (1.2 \text{ k}\Omega) = 2.4 \text{ V}$$
$$V_B = V_{BE} + V_E = 0.7 \text{ V} + 2.4 \text{ V} = 3.1 \text{ V}$$
$$V_B = \frac{R_2 V_{CC}}{R_1 + R_2} = 3.1 \text{ V}$$

and
$$\frac{(18 \text{ k}\Omega)\ (18 \text{ V})}{R_1 + 18 \text{ k}\Omega} = 3.1 \text{ V}$$
$$324 \text{ k}\Omega = 3.1 R_1 + 55.8 \text{ k}\Omega$$
$$3.1 R_1 = 268.2 \text{ k}\Omega$$
$$R_1 = 268.2 \text{ k}\Omega / 3.1 = 86.52 \text{ k}\Omega$$
$$R_C = \frac{V_{R_C}}{I_C} = \frac{V_{CC} - V_C}{I_C}$$

with
$$V_C = V_{CE} + V_E = 10 \text{ V} + 2.4 \text{ V} = 12.4 \text{ V}$$

and
$$R_C = \frac{18 \text{ V} - 12.4 \text{ V}}{2 \text{ mA}}$$
$$= 2.8 \text{ k}\Omega$$

The nearest standard commercial values to R_1 are 82 kΩ and 91 kΩ. However, using the series combination of standard values of 82 kΩ and 4.7 kΩ = 86.7 kΩ would result in a value very close to the design level.

Example 5.4 *For the voltage divider bias circuit shown in Fig. 5.24, find I_C and V_{CE} at the Q-point if β = 100, R_L = 2 kΩ, R_E = 470 Ω, R_1 = 27 kΩ, R_2 = 4.7 kΩ, assuming that V_{BE} = 0.7 V and V_{CC} = 10 V. Also find the relative change in I_C about the Q-point if β changes to 150 assuming that there is no noticeable change in V_{BE}.*

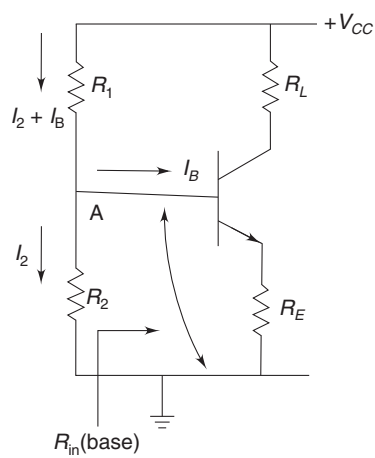


Fig. 5.24

Solution
$$V_{Th} = \frac{R_2}{R_1 + R_2} V_{CC}$$
$$= \frac{4.7 \times 10^3}{27 \times 10^3 + 4.7 \times 10^3} \times 10$$
$$= 1.48 \text{ V}$$

$$R_{Th} = R_1 \parallel R_2$$
$$= \frac{27 \times 10^3 \times 4.7 \times 10^3}{27 \times 10^3 + 4.7 \times 10^3}$$
$$= 4 \text{ k}\Omega$$
$$I_B = \frac{V_{Th} - V_{BE}}{R_{Th} + (1 + \beta) R_E}$$
$$= \frac{1.48 - 0.7}{4 \times 10^3 + 101 \times 0.47 \times 10^3}$$
$$= 15.15 \text{ mA}$$
$$I_C = \beta I_B = (100 \times 15.15) \mu\text{A} \approx 1.52 \text{ mA}$$

For $\beta = 150$, we can obtain

$$I_B = 10.4 \ \mu\text{A}, \quad I_C = 1.56 \quad \text{and} \quad V_{CE} = 6.15 \text{ V}$$

Therefore

$$\frac{\Delta I_C}{I_C} = \frac{1.56 - 1.52}{1.52} = 0.026 \text{ or } 2.6\%$$

Example 5.5 *Determine the dc bias voltage V_{CE} and the current I_C for the voltage-divider configuration of Fig. 5.25.*

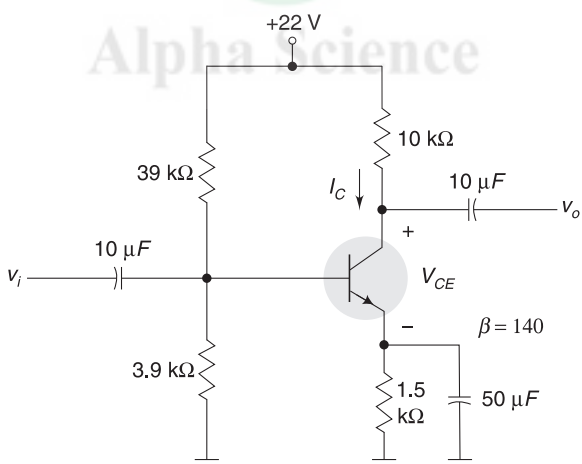


Fig. 5.25

Solution

$$R_{Th} = R_1 \parallel R_2$$
$$= \frac{(39 \text{ k}\Omega)(3.9 \text{ k}\Omega)}{39 \text{ k}\Omega + 3.9 \text{ k}\Omega} = 3.55 \text{ k}\Omega$$

$$E_{Th} = \frac{R_2\, V_{CC}}{R_1 + R_2}$$

$$= \frac{(3.9\ \text{k}\Omega)\ (22\ \text{V})}{39\ \text{k}\Omega + 3.9\ \text{k}\Omega} = 2\ \text{V}$$

$$I_B = \frac{E_{Th} - V_{BE}}{R_{Th} + (\beta + 1)\ R_E}$$

$$= \frac{2\ \text{V} - 0.7\ \text{V}}{3.55\ \text{k}\Omega + (141)\ (1.5\ \text{k}\Omega)}$$

$$= \frac{1.3\ \text{V}}{3.55\ \text{k}\Omega + 211.5\ \text{k}\Omega}$$

$$= 6.05\ \mu\text{A}$$

$$I_C = \beta\, I_B$$

$$= (140)\ (6.05\ \mu\text{A})$$

$$= 0.85\ \text{mA}$$

$$V_{CE} = V_{CC} - I_C\ (R_C + R_E)$$

$$= 22\ \text{V} - (0.85\ \text{mA})\ (10\ \text{k}\Omega + 1.5\ \text{k}\Omega)$$

$$= 22\ \text{V} - 9.78\ \text{V}$$

$$= 12.22\ \text{V}$$

5.14 STABILITY AND STABILITY FACTOR OF VOLTAGE DIVIDER BIAS CIRCUIT

The voltage divider bias circuit may be represented as shown in Fig. 5.26.

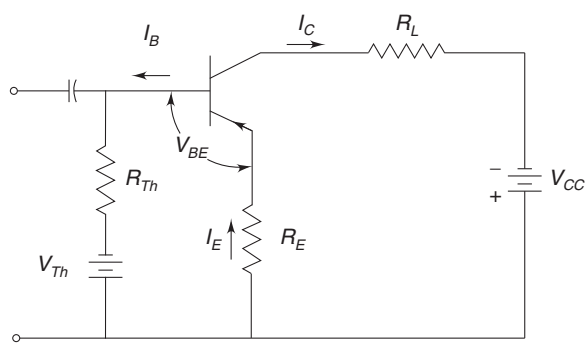


Fig. 5.26

The following equations apply to the circuit shown

$$I_C = I_E - I_B \tag{5.35}$$

$$I_C = \beta I_B + (\beta + 1) I_{CBO} \tag{5.36}$$

By mesh analysis of the base-to-emitter circuit,

$$V_{Th} = - I_E R_E - I_B R_{Th} + V_{BE} \tag{5.37}$$

where $$V_{Th} = \frac{R_2 V_{CC}}{R_1 + R_2}$$

and $$R_{Th} = \frac{R_1 R_2}{R_1 + R_2}$$

Substituting for I_E and I_B from Eqs. (5.35) and (5.36), Eq. (5.37) becomes

$$V_{BE} - V_{Th} = I_C \left[R_E + \frac{R_{Th} + R_E}{\beta} \right] - \left[\frac{\beta + 1}{\beta} \right] I_{CBO} (R_{Th} + R_E) \tag{5.38}$$

If the variations in V_{BE} and β are ignored, Eq. (5.38) may be differentiated w.r.t. I_{CBO} to find the stability factor $S \equiv \dfrac{\partial I_C}{\partial I_{CBO}}$ as

$$0 = S \left[R_E + \frac{R_{Th} + R_E}{\beta} \right] - \left[\frac{\beta + 1}{\beta} \right] (R_{Th} + R_E)$$

$$\therefore \quad S = \frac{(\beta + 1)(R_{Th} + R_E)}{(\beta + 1) R_E + R_{Th}}$$

$$= \frac{R_{Th} + R_E}{R_E + \dfrac{R_{Th}}{\beta + 1}}$$

$$= \frac{1 + R_{Th}/R_E}{1 + \dfrac{R_{Th}}{(\beta + 1) R_E}} \tag{5.39}$$

when $R_E \to 0$, $S \to (\beta + 1)$

when $R_{Th} \to 0$, $S \to 1$

(A high value of S indicates poor stability and a low value of S shows good stability).

In practice, a compromise between the two extreme values of S is chosen. If R_E is made very high, it will dissipate too much power from the collector supply and will create an excessive voltage drop between the emitter and the common

lead. If R_{Th} is made too low, it will act as a shunt to the input signal current and reduce the effective gain of the stage. For low power stages, where the rise in temperature is small, $S = 10$ can often be tolerated. For a typical value of $\beta = 50$, this requires that $R_{Th}/R_E \approx 11$ from Eq. (5.39).

The variation in I_C because of changes in β will now be investigated. Differentiating Eq. (5.38) w.r.t. β gives

$$0 = S_h R_E + (R_E + R_{Th}) \left[\frac{\beta S_h - I_C}{\beta^2} \right]$$

$$- I_{CBO} \frac{(R_E + R_{Th})}{\beta^2} \times [\beta - (\beta + 1)]$$

Thus

$$S_h = \frac{\partial I_C}{\partial \beta} = \frac{I_C - I_{CBO}}{\beta^2} \frac{R_E + R_{Th}}{(\beta R_E + R_E + R_{Th})/\beta}$$

$$= \frac{(I_C - I_{CBO})}{\beta} \frac{S}{(\beta + 1)}$$

$$\therefore \quad \frac{\Delta I_C}{I_C} = \frac{\Delta \beta}{\beta} \frac{S}{(\beta + 1)} \left(1 - \frac{I_{CBO}}{I_C} \right), \tag{5.40}$$

(for small changes in I_C and β)

This represents an improvement by the factor $S/(\beta + 1)$ over the performance of the fixed base bias circuit of Example 5.1.

5.15 PRACTICAL DESIGN APPROACH USING VOLTAGE-DIVIDER BIAS

The voltage-divider bias stabilization is widely used in discrete circuits. Because of this, three basic criteria for quick design have been evolved which give reasonable compromises between stability, gain and output voltage swing. The criteria can be stated as follows:

(a) The emitter-to-ground voltage V_E across R_E should be approximately one-tenth of the dc supply voltage V_{CC} i.e.,

$$V_E = I_E R_E = \frac{1}{10} V_{CC} \tag{5.41}$$

(b) The d.c. current in R_2 should be at least ten times the d.c. base bias current I_B. This is met if the input resistance looking into the base of the transistor βR_E is at least ten times R_2 i.e.,

$$\beta R_E = 10 R_2 \tag{5.42}$$

(c) The collector-emitter voltage V_{CE} at the Q-point should be approximately equal to one-half of the d.c. supply voltage V_{CC} i.e.,

$$V_{CE} \approx \frac{1}{2} V_{CC} \tag{5.43}$$

Let us now apply these criteria to the design of a voltage-divider bias circuit given that

$$V_{CC} = 12 \text{ V}, \quad \beta = 100, \quad V_{BE} = 0.7 \text{ V}$$

and the circuit is to be biased at $I_C = 1$ mA.

Assuming that at the Q-point I_C is referred to as $I_{C(Q)}$ and $I_{E(Q)} = I_{C(Q)}$ and using Eq. (5.41) we obtain

$$R_E = \frac{0.1 \times V_{CC}}{I_{C(Q)}} = \frac{0.1 \times 12}{1 \times 10^{-3}} = 1.2 \text{ k}\Omega \tag{5.44}$$

Using Eqs. (5.42) and (5.44), we get

$$R_2 = \frac{\beta R_E}{10} = \frac{100 \times 1.2 \times 10^3}{10} = 12 \text{ k}\Omega \tag{5.45}$$

Using Eq. (5.43) and reffering to Fig. , we write

$$\begin{aligned} V_{CE(Q)} &= V_{CE} \\ &= V_{CC} - I_{C(Q)} R_L - I_{C(Q)} R_E \\ &\approx \frac{1}{2} V_{CC} \end{aligned}$$

or

$$\begin{aligned} R_L &= \frac{V_{CC}}{2I_{C(Q)}} - R_E = \frac{12}{2 \times 1 \times 10^{-3}} - 1.2 \times 10^3 \\ &= 4.8 \text{ k}\Omega \end{aligned} \tag{5.46}$$

Also referring to Fig. 5.22, we have

$$I_{B(Q)} = \frac{V_{Th} - V_{BE}}{R_{Th} + \beta R_E} = \frac{I_{C(Q)}}{\beta} \tag{5.47}$$

where

$$V_{Th} = \frac{R_2}{R_1 + R_2} V_{CC}$$

$$R_{Th} = \frac{R_1 R_2}{R_1 + R_2} \tag{5.48}$$

Using Eqs. (5.47) and (5.48), we obtain

$$R_1 = \frac{V_{CC} - V_{BE} - I_{C(Q)} R_E}{(V_{BE}/R_2) + I_{C(Q)} \left(\frac{R_E}{R_2} + \frac{1}{\beta} \right)} \tag{5.49}$$

Substituting the values, we get

$$R_1 = \frac{12 - 0.7 - 1 \times 10^{-3} \times 1.2 \times 10^3}{\frac{0.7}{12 \times 10^3} + 1 \times 10^{-3}\left[\frac{1.2 \times 10^3}{12 \times 10^3} + \frac{1}{100}\right]}$$

$$= 60.12 \text{ k}\Omega \qquad (5.50)$$

Thus, we have obtained the circuit components as

$$R_E = 1.2 \text{ k}\Omega, \quad R_L = 4.8 \text{ k}\Omega, \quad R_1 = 60.12 \text{ k}\Omega, \quad R_2 = 12 \text{ k}\Omega$$

In practice, the available resistances close to the computed values are selected.

5.16 MID-POINT BIASING

When an amplifier circuit is so designed that operating point Q lies at the centre of d.c. load line, the amplifier is said to be mid-point biased. When the amplifier is mid-point biased, the Q-point provides values of I_C and V_{CE} that are one-half of their maximum possible values. This is illustrated in Fig. 5.27. Since the *Q*-point is centred on the load line.

$$I_C = \frac{1}{2} I_{C(\max)}; \qquad V_{CE} = \frac{V_{CC}}{2}$$

When a transistor is used as an amplifier, it is always designed for mid-point bias. The reason is that mid-point biasing allows optimum operation of the amplifier. In other words, midpoint biasing provides the largest possible output. This point is illustrated in Fig. 5.27 where *Q*-point is centered on the load line.

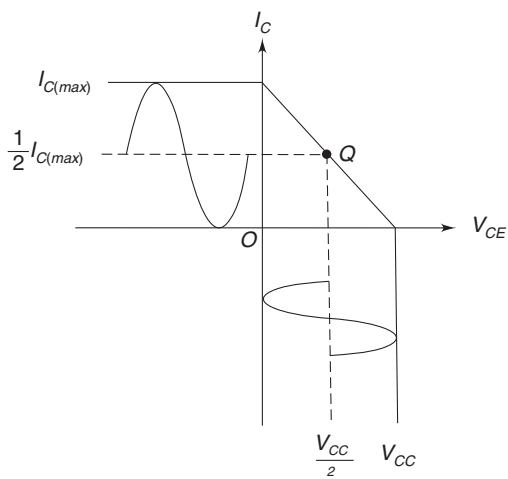


Fig. 5.27

When an a.c. signal is applied to the base of the transistor, collector current and collector-emitter voltage will both vary around their Q-point values. Since Q-point is centered, I_C and V_{CE} can both make the maximum possible transitions above and below their initial d.c. values. If Q-point is located above centre on the load line, the input may cause the transistor to saturate. As a result, a part of the output wave will be clipped off. Similarly, if Q-point is below mid-point on the load line, the input may cause the transistor to go into cut-off. This can also cause a portion of the output to be clipped. It follows, therefore, that mid-point biased amplifier circuit allows the best possible a.c. operation of the circuit.

5.17 PRACTICAL CIRCUIT OF TRANSISTOR AMPLIFIER

It is important to note that a transistor can accomplish faithful amplification only if proper associated circuitry is used with it. Figure 5.28 shows a practical single stage transistor amplifier. The various circuit elements and their function are as follows:

(i) Biasing circuit. The resistances R_1, R_2, and R_E form the biasing and stabilization circuit. The biasing circuit must establish a proper operating point otherwise a part of the negative half-cycle of the signal may be cut off in the output.

(ii) Input capacitor C_{in}. An electrolytic capacitor C_{in} ($\approx$ 10 μF) is used to couple the signal to the base of the transistor. If it is not used, the signal source resistance will come across R_2 and thus change the bias. The capacitor C_{in} allows only a.c. signal to flow but isolates the signal source from R_2.*

(iii) Emitter bypass capacitor C_E. An emitter bypass capacitor C_E ($\approx$ 100 μF) is used in parallel with R_E to provide a low reactance path to the amplified a.c. signal. If it is not used, then amplified a.c. signal flowing through R_E will cause a voltage drop across it, thereby reducing the output voltage.

(iv) Coupling capacitor C_C. The coupling capacitor C_C ($\approx$10 μF) couples one stage of amplification to the next stage. If it is not used, the bias conditions of the nest stage will be drastically changed due to the shunting effect of R_C. This is because of R_C will come in parallel with the upper resistance R_1 of the biasing network of the next stage, thereby altering the biasing conditions of the latter. In short, the coupling capacitor C_C isolates the d.c. of one stage from the next stage, but allows the passage of a.c. signal.

Various Circuit Current It is useful to mention the various currents in the complete amplifier circuit. These are shown in the circuit of Fig. 5.28.

---

* It may be noted that a capacitor offers infinite reactance to d.c. and blocks it completely whereas it allows a.c. to pass through it.

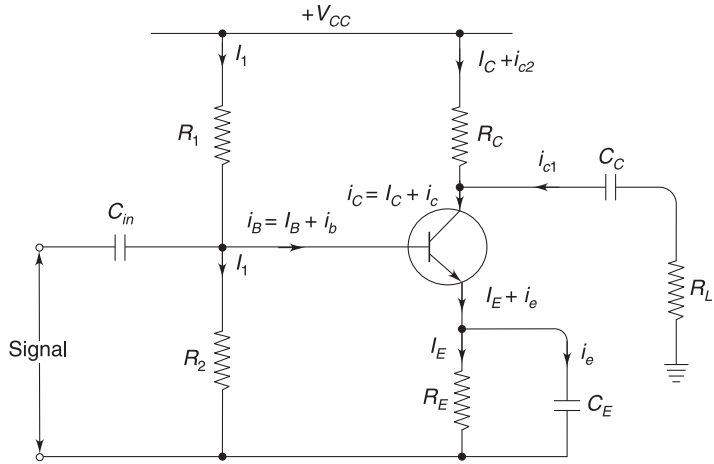


Fig. 5.28

(i) Base Current When no signal is applied in the base circuit, d.c. base current I_B flows due to biasing circuit. When a.c. signal is applied, a.c. base current i_b also flows. Therefore, with the application of signal, total base current i_B is given by

$$i_B = I_B + i_b$$

(ii) Collector Current When no signal is applied, a d.c. collector current I_C flows due to biasing circuit. When a.c. signal is applied, a.c. collector current i_c also flows. Therefore, the total collector current i_C is given by;

$$i_C = I_C + i_c$$

where $I_C = \beta I_B$ = zero signal collector current

$i_c = \beta i_b$ = collector current due to signal

(iii) Emitter Current When nonsignal is applied, a d.c. emitter current I_E flows. With the application of signal, total emitter current i_E is given by;

$$i_E = I_E + i_e$$

It is useful to keep in mind that:

$$I_E = I_B + I_C$$

$$i_e = i_b + i_c$$

Now base current is usually very small, therefore, as a reasonable approximation,

$$I_E \approx I_C \quad \text{and} \quad i_e \approx i_c$$

5.18 PHASE REVERSAL

In common emitter connection, when the input signal voltage increases in the positive sense, the output voltage increases in the negative direction and vice-versa. In other words, there is a phase difference of 180° between the input and output voltage in *CE* connection. This is called phase reversal.*

The phase difference of 180° between the signal voltage and output voltage in a common emitter amplifier is known as **phase reversal**.

Consider a common emitter amplifier circuit shown in Fig. 5.29. The signal is fed at the input terminals (i.e., between base and emitter) and output is taken from collector and emitter end of supply. The total instantaneous output voltage v_{CE} is given by;

$$^{**}v_{CE} = V_{CC} - i_C R_C \tag{5.51}$$

When the signal voltage increases in the positive half-cycle, the base current also increases. The result is that collector current and hence voltage drop $i_C R_C$ increases. As V_{CC} is constant, therefore, output voltage v_{CE} decreases. In other words, as the signal voltage is increasing in the positive half-cycle, the output voltage is increasing in the negative sense i.e., output is 180° out of phase with the input. It follows, therefore, that in a common-emitter amplifier, the positive half-cycle of the signal appears as amplified negative half-cycle in the output and vice versa. It may be noted that amplification is not affected by this phase reversal.

The fact of phase reversal can be readily proved mathematically. Thus differentiating exp. (1), we get

$$dv_{CE} = 0 - di_c R_C$$

or

$$dv_{CE} = - di_c R_C$$

The negative sign shows that output voltage is 180° out of phase with the input signal voltage.

Graphical Demonstration The fact of phase reversal in *CE* connection can be shows graphically with the help of output characteristics and load line (see Fig. 5.30).

In Fig. 5.30, *AB* is the load line. The base current fluctuates between, say ± 5μA with 10 μA as the zero signal base current. From the figure, it is clear that when the base current is maximum in the positive direction, v_{CE} becomes maximum in the negative direction (point *G* in Fig. 5.30). On the other hand, when the base current is maximum in the negative direction, v_{CE} is maximum in

* This is so if output is taken from collector and emitter end of supply as is always done. However, if the output is taken across R_C, it will be in phase with the input.

** Reactance of C_C (= 10 μF) is negligible at ordinary signal frequencies. Therefore, it can be considered a short for the signal.

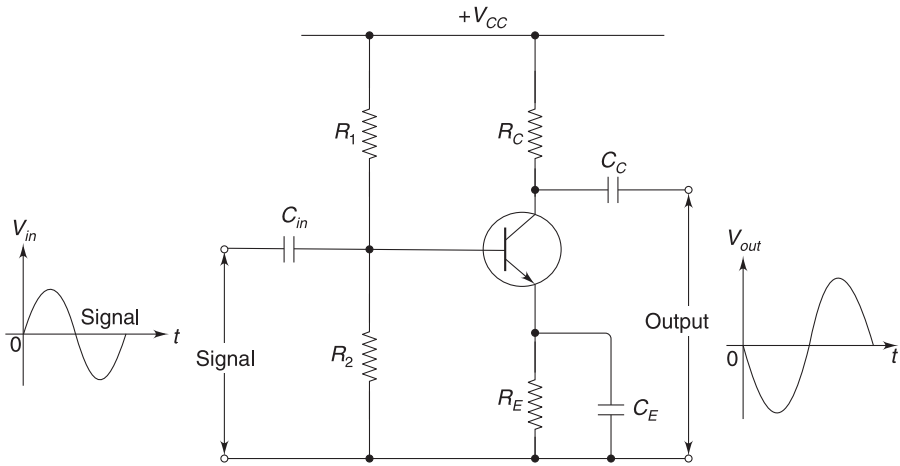


Fig. 5.29

the positive sense (point *H* in Fig. 5.30). Thus, the input and output voltages are in *phase opposition* or equivalently, the transistor is said to produce a 180° phase reversal of output voltage w.r.t. signal voltage.

Note: No phase reversal of voltage occurs in common base and common collector amplifier. The a.c. output voltage is in phase with the a.c. input signal. For all three amplifier configurations; input and output currents are in phase.

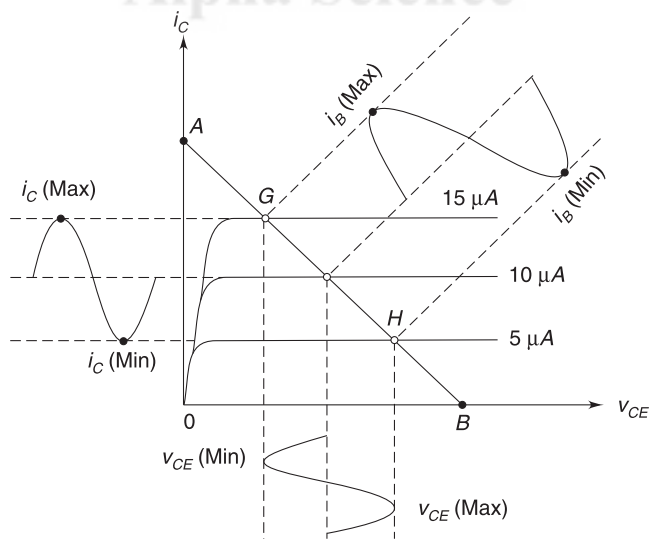


Fig. 5.30

5.19 D.C. AND A.C. EQUIVALENT CIRCUITS

In a transistor amplifier, both d.c. and a.c. conditions prevail. The d.c. sources set up d.c. currents and voltages whereas the a.c. source (i.e. signal) produces fluctuations in the transistor currents and voltages. Therefore, a simple way to analyse the action of a transistor is to split the analysis into two parts *viz*., a d.c. analysis and an a.c. analysis. In the d.c. analysis, we consider all the d.c. sources at the same time and work out the d.c. currents and voltages in the circuit. On the other hand, for a.c. analysis, we consider all the a.c. sources at the same time and work out the a.c. currents and voltages. By adding the d.c. and a.c. currents and voltages, we get the total currents and voltages in the circuit. For example, consider the amplifier circuit shown in Fig. 5.31. this circuit can be easily analysed by splitting it into d.c. *equivalent circuit* and a.c. *equivalent circuit*.

(i) D.C. Equivalent Circuit In the d.c. equivalent circuit of a transistor amplifier, only d.c. conditions are to be considered i.e., it is presumed that no signal is applied. As direct current cannot flow through a capacitor, therefore, *all the capacitors look like open circuits in the d.c. equivalent circuit*. It follows therefore, that in order to draw the equivalent d.c. circuit, the following two steps are applied to the transistor circuit:

(a) Reduce all a.c. sources to zero.

(b) Open all the capacitors.

Applying these two steps to the circuit shown in Fig. 5.31, we get the d.c. equivalent circuit shown in Fig. 5.32. We can easily calculate the d.c. currents and voltages from this circuit.

(ii) A.C. Equivalent Circuit In the a.c. equivalent circuit of a transistor amplifier, only a.c. conditions are to be considered. Obviosuly, the d.c. voltage is not important for such a circuit and may be considered zero. The capacitors are generally used to couple or bypass the a.c. signal. The designer intentionally selects capacitors that are large enough to appear as short-circuit to the a.c. signal. It follows, therefore, that in order to draw the a.c. equivalent circuit, the following two steps are applied to the transistor circuit:

(a) Reduce all d.c. sources to zero (i.e. $V_{CC} = 0$)

(b) Short all the capacitors.

Applying these two steps to the circuit shown in Fig. 5.31, we get the a.c. [*] equivalent circuit shown in Fig. 5.33. We can easily calculate the a.c. currents and voltages from this circuit.

[*] Note that R_1 is also in parallel with transistor input so far as signal is concerned. Since R_1 is connected from the base lead to V_{CC} and V_{CC} is at "a.c. ground", R_1 is effectively connected from the base lead to ground as far as signal is concerned.

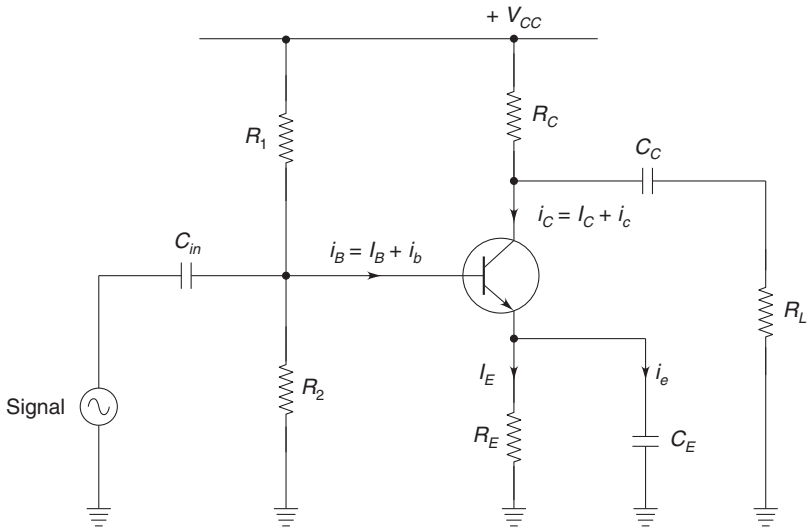


Fig. 5.31

It may be seen that total current in any branch is the sum of d.c. and a.c. currents through that branch. Similarly, the total voltage across any branch is the sum of d.c. and a.c. voltages across that branch.

Example 5.6 *For the transistor amplifier circuit shown in Fig. 5.31, determine:*

(i) d.c. load and a.c. load

(ii) maximum collector-emitter voltage and collector current under d.c. conditions

(iii) maximum collector-emitter voltage and collector current when a.c. signal is applied

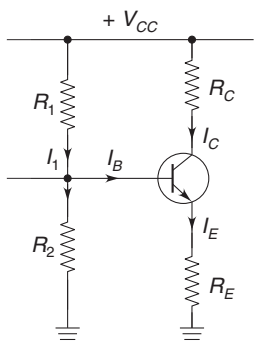


Fig. 5.32

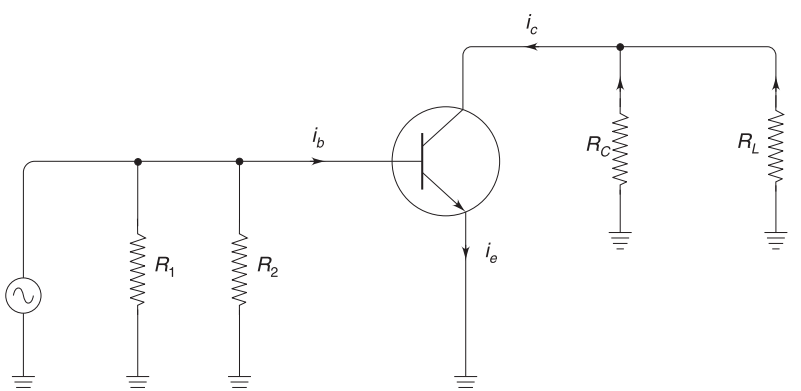


Fig. 5.33

Solution Refer back to the transistor amplifier circuit shown in Fig. 5.31.

(i) The d.c. load for the transistor is Thevenin's equivalent resistance as seen by the collector and emitter terminals. Thus referring to the d.c. equivalent circuit shown in Fig. 5.32. Thevenin's equivalent resistance can be found by shorting the voltage source (i.e. V_{CC}) as shown in Fig. 5.34.

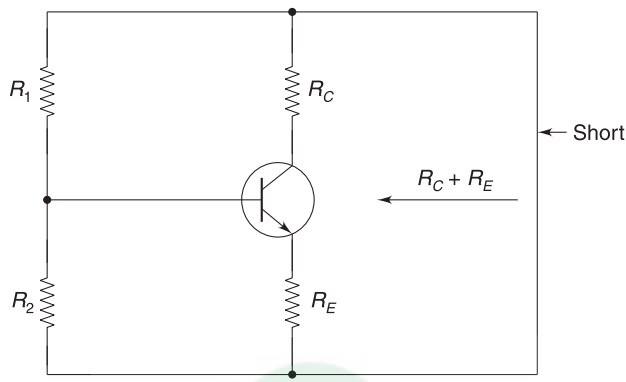


Fig. 5.34

Because a voltage source looks like a short, it will bypass all other resistances except R_C and R_E which will appear in series. Consequently, transistor amplifier will see a d.c. load of $R_C + R_E$ i.e.,

$$\text{D.C. load} = R_C + R_E$$

Referring to the a.c. equivalent circuit shown in Fig. 5.33, it is clear that as far as a.c. signal is concerned, resistance R_C appears in parallel with R_L. In other words, transistor amplifier sees an a.c. load equal to $R_C \parallel R_L$ i.e.,

$$\text{A.C. load,} \qquad R_{AC} = R_C \parallel R_L = R_C R_L / R_C + R_L$$

(ii) Referring to d.c. equivalent circuit of Fig. 5.32

$$V_{CC} = V_{CE} + I_C (R_C + R_E)$$

The maximum value of V_{CE} will occur when there is no collector current i.e., $I_C = 0$

$$\therefore \text{ Maximum} \qquad V_{CE} = V_{CC}$$

The maximum collector current will flow when $V_{CE} = 0$

$$\therefore \text{ Maximum} \qquad I_C = \frac{V_{CC}}{R_C + R_E}$$

(iii) When no signal is applied, V_{CE} and I_C are the collector-emitter voltage and collector current respectively. When a.c. signal is applied, it causes changes to take place above and below the operating point Q (i.e. V_{CE} and I_C).

Maximum collector current due to a.c. signal = $^{*}I_C$

$\therefore$ Maximum positive swing of a.c. collector-emitter voltage

$$= I_C \times R_{AC}$$

Total maximum collector-emitter voltage

$$= V_{CE} + I_C R_{AC}$$

Maximum positive swing of a.c. collector current

$$= V_{CE}/R_{AC}$$

$\therefore$ Total maximum collector current

$$= I_C + V_{CE}/R_{AC}$$

5.20 LOAD LINE ANALYSIS

The output characteristics are determined experimentally and indicate the relation between V_{CE} and I_C. However, the same information can be obtained in a much simpler way by representing the mathematical relation between I_C and V_{CE} graphically. As discussed before, the relationship between V_{CE} and I_C is linear so that it can be represented by a straight line on the output characteristics. This is known as a *load line*. The points lying on the load line give the possible values of V_{CE} and I_C in the output circuit. As in a transistor circuit both d.c. and a.c. conditions exist, therefore, there are two types of load lines, namely; d.c. load line and a.c. load line. The former determines the locus of I_C and V_{CE} in the zero signal conditions and the latter shows these values when the signal is applied.

(i) D.C. load line *It is the line on the output characteristics of a transistor circuit which gives the values of I_C and V_{CE} corresponding to zero signal or d.c. conditions.*

Consider the transistor amplifier shown in Fig. 5.35. In the absence of signal, d.c. conditions prevail in the circuit as shown in Fig. 5.36(a). Referring to this circuit and applying Kirchhoff's voltage law,

$$V_{CE} = V_{CC} - I_C R_C - I_E R_E$$

or
$$V_{CE} = V_{CC} - I_C (R_C + R_E) \qquad (\because I_E \approx I_C) \tag{5.52}$$

As for a given circuit, V_{CC} and $(R_C + R_E)$ are constant, therefore, it is a first degree **equation and can be represented by a straight line on the output characteristics. This is known as *d.c. load line* and determines the locus of V_{CE} and I_C points in the zero signal conditions. The d.c. load line can be readily plotted by locating two *end points* of the straight line.

---

* For faithful amplification

** This Equation is known as load line equation since it relates the collector-emitter voltage (V_{CE}) to the collector current I_C flowing through the load.

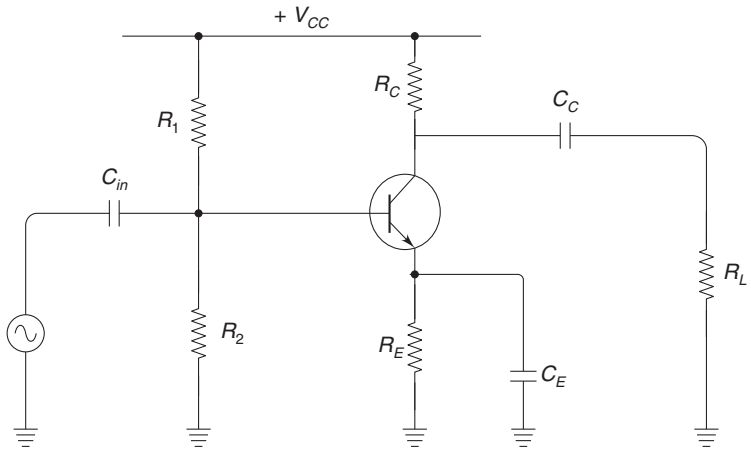


Fig. 5.35

The value of V_{CE} will be maximum when $I_C = 0$. Therefore, by putting $I_C = 0$ in Eq. (5.52), we get,

$$\text{max. } V_{CE} = V_{CC}$$

This locates the first point B ($OB = V_{CC}$) of the d.c. load line.

The value of I_C will be maximum when $V_{CE} = 0$.

$$\therefore \qquad \text{max. } I_C = \frac{V_{CC}}{R_C + R_E}$$

This locates the second point A ($OA = V_{CC}/R_C + R_E$) of the d.c. load line. By joining points A and B, d.c. load line AB is constructed (see Fig. 5.36(b)).

Alternatively The two end points of the d.c. load line can also be determined in another way.

$$V_{CE} + I_C (R_C + R_E) = V_{CC}$$

Dividing throughout by V_{CC}, we have,

$$\frac{V_{CE}}{V_{CC}} + \frac{I_C}{(V_{CC}/R_C + R_E)} = 1 \qquad (5.53)$$

The equation of a line having intercepts a and b on x-axis and y-axis respectively is given by;

$$\frac{x}{a} + \frac{y}{b} = 1 \qquad (5.54)$$

Comparing Eqs. (5.53) and (5.54), we have,

Intercept on x-axis = V_{CC}

$$\text{Intercept on } y\text{-axis} = \frac{V_{CC}}{R_C + R_E}$$

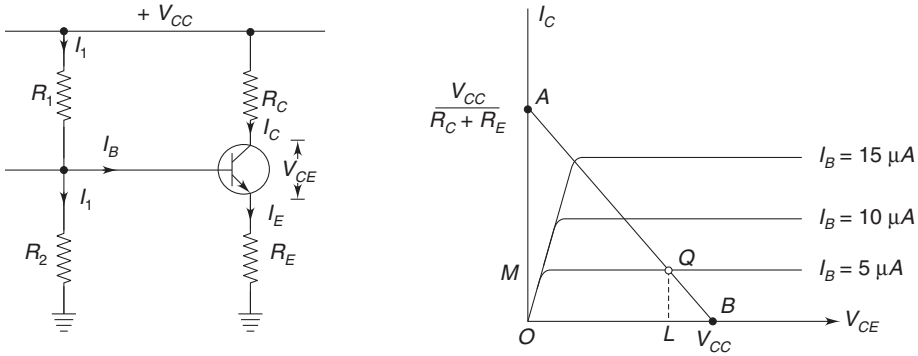


Fig. 5.36

With the construction of d.c. load line on the output characteristics, we get the complete information about the output circuit of transistor amplifier in the zero signal conditions. All the points showing zero signal I_C and V_{CE} will obviously lie on the d.c. load line. At the same time I_C and V_{CE} conditions in the circuit are also represented by the output characteristics. Therefore, actual operation conditions in the circuit will be represented by the point where d.c. load line intersects the base current curve under study. Thus, referring to Fig. 5.36(b), if I_B = 5 μA is set by the biasing circuit then Q (i.e. intersection of 5 μA curve and load line) is the operation point.

(ii) A.C. load line This is the line on the output characteristics of a transistor circuit which gives the value of i_C and v_{CE} when signal is applied.

Referring back to the transistor amplifier shown in Fig. 5.35, it's a.c. equivalent circuit as far as output circuit is concerned is as shown in Fig. 5.37(a). To add a.c. load line to the output characteristics, we again require two end points – one minimum collector-emitter voltage point and the other maximum

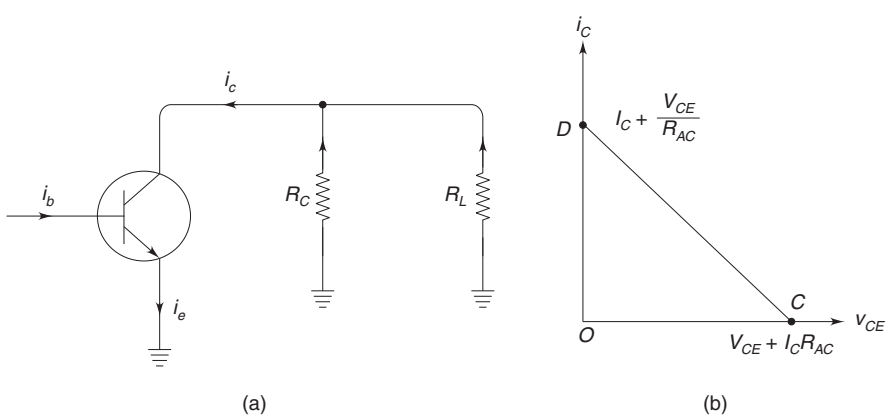


Fig. 5.37

collector current point. Under the application of a.c. signal, these values are (refer to Example 5.6).

Max. collector-emitter voltage = $V_{CE} + I_C R_{AC}$. This locates the point C of the a.c.load line on the collector-emitter voltage axis.

Maximum collector current

$$= I_C + \frac{V_{CE}}{R_{AC}}$$

where $$R_{AC} = R_C \parallel R_L = \frac{R_C R_L}{R_C + R_L}$$

This locates the point D of a.c. load line on the collector current axis. By joining points C and D, the a.c. load line CD is constructed (See Fig. 5.37(b)).

Example 5.7 *For the transistor amplifier shown in Fig. 5.38, R_1 = 10 kΩ, R_2 = 5 kΩ, R_C = 1 kΩ, R_E = 2 kΩ and R_L = 1 kΩ.*

(i) Draw d.c. load line (ii) Determine the operating point (iii) Draw a.c. load line.

Assume V_{BE} = 0.7 V.

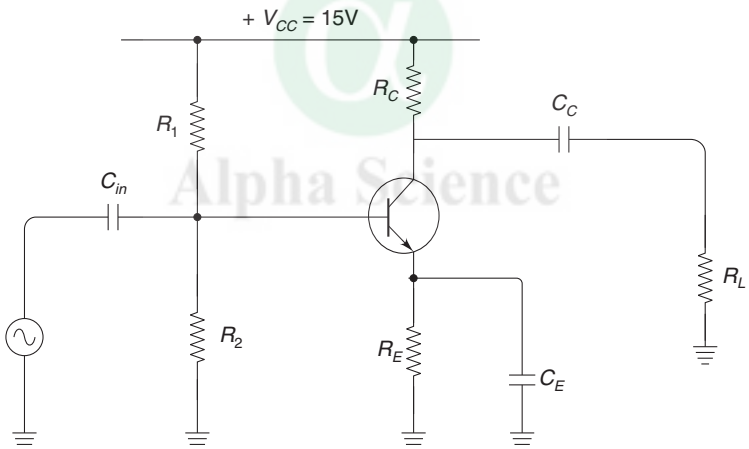


Fig. 5.38

Solution

(i) D.C. load line To draw d.c. load line, we require two end points viz. maximum V_{CE} point and maximum I_C point.

Maximum $V_{CE} = V_{CC} = 15$ V [See Art. 5.20]

This locates the point B (OB = 15 V) of the d.c. load line.

Maximum $$I_C = \frac{V_{CC}}{R_C + R_E} = \frac{15 \text{ V}}{(1 + 2) \text{ k}\Omega}$$

$$= 5\text{mA} \quad \text{(See Section 5.20)}$$

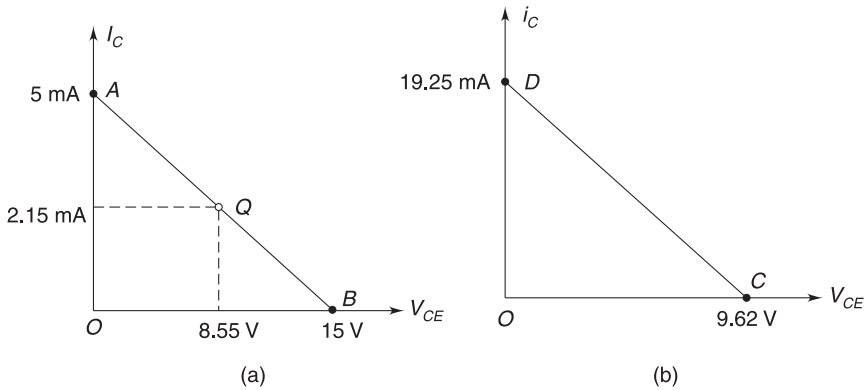


Fig. 5.39

This locates the point A (OA = 5 mA) of the d.c. load line. Figure 5.39(a) shows the d.c. load line AB.

(ii) Operating point Q The voltage across R_2 (= 5 kΩ) is 5 V i.e.,

$$V_2 = 5\text{V}.$$

Now, $$V_2 = V_{BE} + I_E R_E$$

$$\therefore \quad I_E = \frac{V_2 - V_{BE}}{R_E} = \frac{(5 - 0.7)\text{ V}}{2\text{ k}\Omega} = 2.15\text{ mA}$$

$$\therefore \quad I_C = I_E = 2.15\text{ mA}$$

Now $$V_{CE} = V_{CC} - I_C (R_C + R_E) = 15 - 2.15\text{ mA} \times 3\text{ k}\Omega$$
$$= 8.55\text{ V}$$

∴ Operating point Q is 8.55 V, 2.15 mA. This is shown on the d.c. load line.

(iii) A.C. load line To draw a.c. load line, we require two end points viz., maximum collector-emitter voltage point and maximum collector current point when signal is applied.

a.c. load, $$R_{AC} = R_C \parallel R_L = \frac{1 \times 1}{1 + 1} = 0.5\text{ k}\Omega$$

∴ Maximum collector-emitter voltage

$$= V_{CE} + I_C R_{AC} \text{ [See Example 5.6]}$$
$$= 8.55 + 2.15\text{ mA} \times 0.5\text{ k}\Omega = 9.62\text{ volts}$$

This becomes the points C (OC = 9.62 V) on the v_{CE} axis.

Maximum collector current = $I_C + V_{CE}/R_{AC}$

$$= 2.15 + (8.55\text{ V}/0.5\text{ k}\Omega) = 19.25\text{ mA}$$

This locates the point D (OD = 19.25 mA) on the i_C axis. By joining points C and D, a.c. load line CD is constructed (see Fig. 5.39(b)).

5.21 OUTPUT FROM TRANSISTOR AMPLIFIER

A transistor raises the strength of a weak signal and thus acts as an amplifier. Figure 5.40 shows the common-emitter amplifier. There are two ways of taking output from this transistor connection. The output can be taken either across R_C or across terminals 2-2. In either case, the magnitude of output is the same. This can be seen as follows.

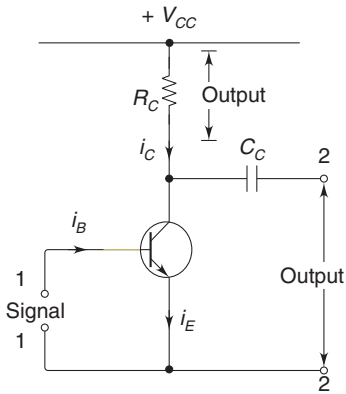


Fig. 5.40

(i) First method We can take the output directly by putting a load resistance R_C in the collector circuit, i.e.,

Output = voltage across

$$R_C = i_C R_C \qquad (5.55)$$

This method of taking output from collector load is used only in single stage amplification.

(ii) Second method The output can also be taken across terminals 2-2 i.e., from collector and emitter end of supply

$$\text{Output} = \text{Voltage across terminals 2-2.}$$
$$= V_{CC} - i_C R_L$$

As V_{CC} is a direct voltage and cannot pass through capacitor C_C, therefore only varying voltage $i_C R_C$ will appear across terminals 2-2.

$$\therefore \qquad \text{output} = - i_C R_C \qquad (5.56)$$

From Eqs. (5.55) and (5.56), it is clear that magnitude of output is the same whether we take output across collector load or terminals 1 and 2. The minus sign in Eq. (5.56) simply indicates the phase reversal. The second method of taking output is used in multistage of amplification.

Example 5.8 *Determine whether or not the circuit shown in Fig. 5.41(a) is mid-point biased.*

Solution Let us first construct the d.c. load line

$$I_{C\,(\max)} = \frac{V_{CC}}{R_C} = \frac{8\text{ V}}{2\text{ k}\Omega} = 4\text{ mA}$$

This locates the point A (OA = 4 mA) of the d.c. load line.

$$V_{CE\,(\max)} = V_{CC} = 8\text{ V}$$

This locates the point B (OB = 8 V) of the d.c. load line. By joining these two points, d.c. load line AB is constructed (see Fig. 5.41).

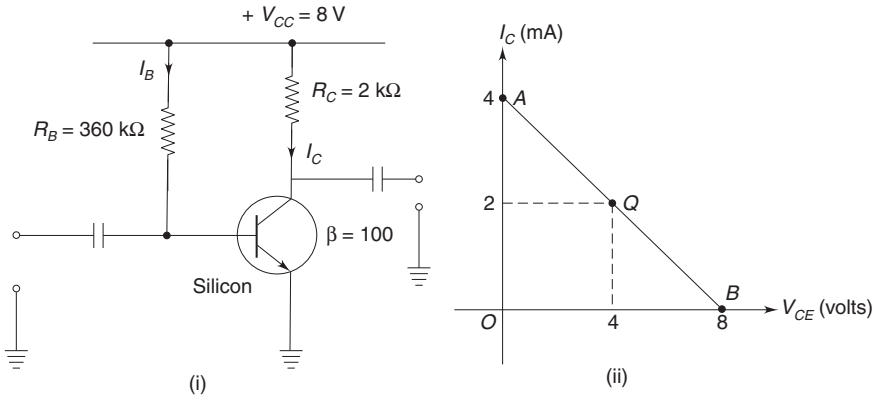


Fig. 5.41

Operating point. Referring to Fig. 5.41(a), we have,

$$I_B = \frac{V_{CC} - V_{BE}}{R_B} = \frac{8 \text{ V} - 0.7 \text{ V}}{360 \text{ k}\Omega} = 20.28 \ \mu\text{A}$$

$$\therefore \quad I_C = \beta = 100 \ (20.28 \ \mu\text{A}) = 2.028 \text{ mA}$$

Also
$$V_{CE} = V_{CC} - I_C R_C$$
$$= 8 \text{ V} - (2.028 \text{ mA}) \ (2 \text{ k}\Omega)$$
$$= 3.94 \text{ V}$$

Since V_{CE} is nearly one-half of V_{CC}, the amplifier circuit is mid-point biased.

Note: We can determine whether or not the circuit is mid-point biased without drawing the dc load line. By definition, a circuit is mid-point biased when the Q-point value of V_{CE} is one-half of V_{CC}. Therefore, all that you have do to is to find the operating point Q of the circuit. If the Q-point value of V_{CE} is one-half of V_{CC}, the circuit is mid-point biased.

Questions

5.1 For the fixed-bias configuration of Fig. 5.42, determine

(a) I_{B_Q} (b) I_{C_Q}
(c) V_{CE_Q} (d) V_C
(e) V_B (f) V_E

[**Ans.** I_{B_Q} = 32.55 μA, I_{C_Q} = 2.93 mA, V_{CE_Q} = 8.09 V, V_C = 8.09 V, V_B = 0.7 V, V_E = 0 V]

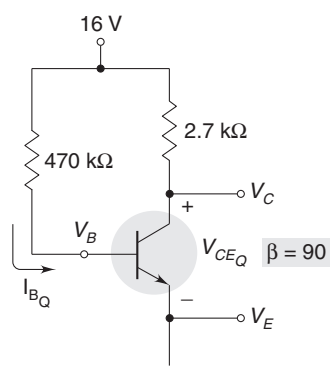


Fig. 5.42

5.2 Given the information appearing in Fig. 5.43, determine:

(a) I_C (b) V_{CC}

(c) β (d) R_B

[**Ans.** I_C = 3.98 mA, V_{CC} = 15.96 V, β = 199, R_B = 763 kΩ]

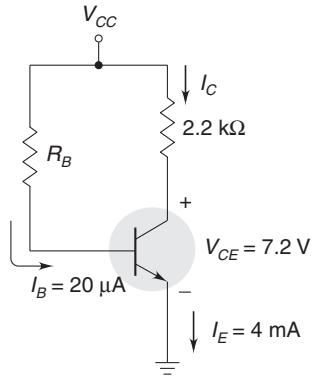


Fig. 5.43

5.3 What do you understand by transistor biasing? Why it is needed?

5.4 What do you understand by stabilization of operating point?

5.5 Mention the essentials of a biasing circuit.

5.6 Describe the various methods used for transistor biasing. State their advantages and disadvantages.

5.7 Describe the potential divider method in detail. How stabilization of operating point is achieved by this method?

5.8 Explain thermal runaway in a transistor circuit.

5.9 Explain why the battery connected between the emitter and base terminals requires a high resistance in series with it.

5.10 Draw a simple circuit in which why only one battery is used and biasing is achieved by fixing the base current.

5.11 Prove mathematically that the operating point in a potential-divider biasing circuit is independent of β. Make relevant assumptions.

5.12 Explain why the fixed bias circuit, inspite of its simplicity, is not much used in amplifiers.

5.13 Explain the function of the emitter resistor R_E in the potential divider biasing circuit.

5.14 Explain why operating point is fixed in the centre of the active region of transistor characteristics in a good voltage amplifier.

5.15 Calculate the collector current and collector-to-emitter voltage for the *PNP* silicon transistor used in voltage divider bias method assuming the following data:

R_1 = 100 kΩ, V_{CC} = 12 V

R_2 = 27 kΩ, V_{BE} = 0.751 V

R_C = 2 kΩ, β = 75

R_E = 1 kΩ [**Ans.** $I_C \approx I_E$ = 1.8 mA, V_{CE} = 6.6 V]

5.16 Derive the expression for Q-point in the voltage divider bias circuit using Thevenin's theorem.

6 Two Port Networks and Amplifiers

6.1 TWO PORT NETWORK

A linear circuit that gives the same response at the output and input ports as does a second network is said to be *equivalent* to the second network. By equating the parameters of a network to the measured parameters of the transistor, a two port equivalent circuit can be made to act as does the transistor in the circuit. The concept of equivalence contributes greatly to the understanding of circuits containing transistors since the transistor as a circuit element is given parameters that are compatible in our usual circuit analysis.

To ensure transistor linearity adequate for active-circuit models requires that the transistors be operated with small-signal amplitudes and over small regions of quasi-linear characteristics.

We will now provide a brief review of applicable two port network theory and then apply this theory to develop circuit models for transistors. Then the circuit models can be used in determination of the characteristics of the basic amplifier circuits.

Figure 6.1(a) shows a network box which is assumed to contain a linear network, either active or passive. Conductors can be internally connected to arbitrary nodes and brough out as terminals; any pair of terminals is a *port*. Currents at the terminals are considered as positive inward, so reversal of a box will not alter an analysis.

It is the terminal quantities v_1, i_1, v_2 and i_2 by which the two port network responds to external forcing functions, and specification of these quantities is equivalent to specification of the network response. It is possible to apply network relations to transistors that do not have linear volt-ampere characteristics by restriction of the operation to regions of quasi-linearity. Linearity is illustrated in Fig. 6.1, where the slope is constant; in Fig. 6.1(d) is a piecewise linear approximation to the non-linear relation of Fig. 6.1(c).

Any pair of the terminal variables, v_1, i_1, v_2 and i_2 may be arbitrarily chosen as independent, leading to two equations that may be solved for the other two

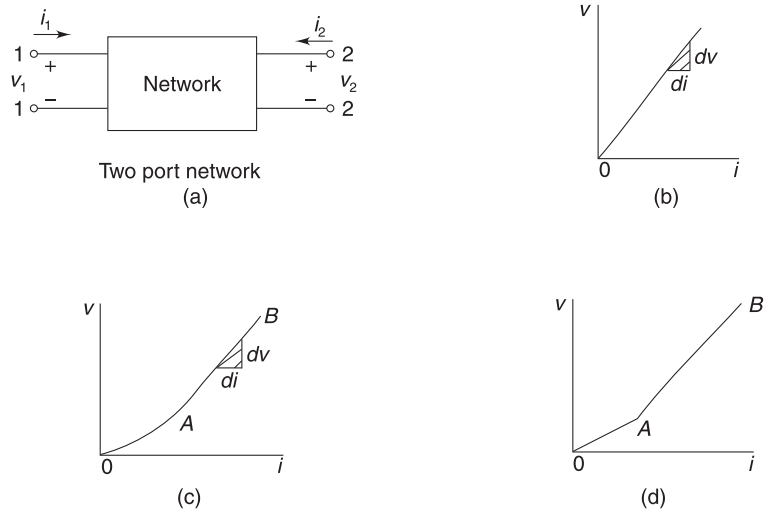


Fig. 6.1 *(a) Volt-ampere curve for a linear circuit, (b) non-linear device, (c) piecewise linear approximation*

dependent variables. Choice of three of the possible pairs of independent variables as v_2 and i_1, v_1 and v_2, and i_1 and i_2 gives us three sets of circuit parameters which have been found useful in electronic circuit analysis.

In choosing i_1 and v_2 as independent variables, we imply that

$$v_1 = f_1\,(i_1, v_2)$$
$$i_2 = f_2\,(i_1, v_2)$$

Our circuits are operated with a.c. signals, and the effect of changes in the thermal quantities can be determined by taking the total differentials as

$$dv_1 = \frac{\partial v_1}{\partial i_1}\,di_1 + \frac{\partial v_1}{\partial v_2}\,dv_2 \tag{6.1}$$

$$di_2 = \frac{\partial i_2}{\partial i_1}\,di_1 + \frac{\partial i_2}{\partial v_2}\,dv_2 \tag{6.2}$$

With operation over linear regions of the device curves with constant slopes, the partial derivatives become constants. Writing the above equations for sinusoidal changes, we have

$$V_1 = h_i\,I_1 + h_r\,V_2 \tag{6.3}$$

$$I_2 = h_f\,I_1 + h_o\,V_2 \tag{6.4}$$

which in matrix form is

$$\begin{bmatrix} V_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} h_i & h_r \\ h_f & h_o \end{bmatrix} \begin{bmatrix} I_1 \\ V_2 \end{bmatrix}$$

An advantage of matrix notation is demonstrated, in that it separates the independent variables I_1 and V_2 on the right, the circuit function, and the dependent variables V_1 and I_2 on the left.

The *h* parameters may be correlated with a given network by measurements made at the terminals of the network, with assigned open- or short-circuit terminations. With a short-circuit at 2,2 we have $V_2 = 0$. Applying this condition in Eqs. (6.3) and (6.4) we define two *h* parameters by currents and voltages as

$$\left.\begin{aligned} h_i &= \frac{V_1}{I_1} = \text{short-circuit input impedance; } V_2 = 0 \\ h_f &= \frac{I_2}{I_1} = \text{short-circuit forward current gain; } V_2 = 0 \end{aligned}\right\} \quad (6.5)$$

With an open-circuit at 1,1 we have $I_1 = 0$, and using this condition in Eqs. (6.3) and (6.4) yields definitions:

$$\left.\begin{aligned} h_r &= \frac{V_1}{V_2} = \text{open-circuit reverse voltage gain; } I_1 = 0 \\ h_o &= \frac{I_2}{V_2} = \text{open-circuit output admittance; } I_1 = 0 \end{aligned}\right\} \quad (6.6)$$

The *h* coefficients are known as the *hybrid parameters*, since both open- and short-circuit terminations are used in defining them. It should be noted that one parameter is an impedance, one an admittance, and two are dimensionless ratios.

Example 6.1 *Find the values of the h parameters for the circuit of Fig. 6.2.*

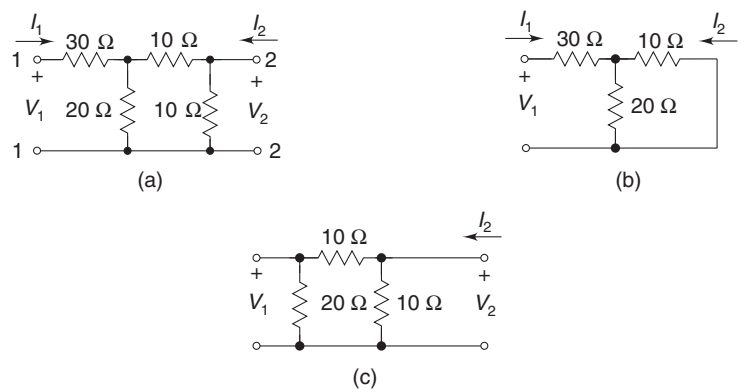


Fig. 6.2 *Circuits for the example*

Solution We define h_i as

$$h_i = \frac{V_1}{I_1} \quad (V_2 = 0)$$

With a short-circuit at 2, 2 the circuit is changed to that at (b). Then

$$h_i = 30 + \frac{10 \times 20}{10 + 20} = 30 + \frac{200}{30}$$
$$= 36.7 \text{ ohms}$$

For h_f we have

$$h_f = \frac{I_2}{I_1} \quad (V_2 = 0)$$

Again using the circuit at (b), we can use current division to write

$$-I_2 = \frac{20}{10 + 20} I_1$$

$$\frac{I_2}{I_1} = -\frac{20}{30} = -0.67$$

For an open-circuit at 1,1 we arrive at the circuit at (c). Then

$$h_r = \frac{V_1}{V_2} \quad (I_1 = 0)$$

By voltage division we have

$$V_1 = \frac{20}{10 + 20} V_2$$

$$h_r = \frac{V_2}{V_1} = \frac{20}{30} = 0.067$$

With the open-circuit termination we obtain the definition for h_o as

$$h_o = \frac{I_2}{V_2} \quad (I_1 = 0)$$

The impedance seen at 2, 2 is

$$Z = \frac{20 \times 30}{20 + 10 + 10} = \frac{300}{40} = 7.5 \text{ ohms}$$

Then

$$V_2 = 7.5\, I_2$$

$$h_o = \frac{I_2}{V_2} = \frac{1}{7.5} = 0.133 \text{ mho}$$

Summarizing

$$h_i = 36.7 \text{ ohms} \qquad h_r = 0.67$$
$$h_f = -0.67 \qquad h_o = 0.133 \text{ mho}$$

6.2 CONTROLLED SOURCES

An independent ideal source in a network has its strength arbitrarily defined. It delivers its specified voltage or current regardless of the system to which

it is connected. A significant property of the *ideal controlled source* is that it is *unilateral*; a signal applied at the control terminals produces a proportional output, but a signal at the output terminals does not affect the control input.

Transistors operate as control devices and amplifiers; they are sources of energy whose output is under control of a voltage or current elsewhere in the system. Thus they are not independently specified and are *dependent*. As an example, the defining relations in Figs. 6.3(a) and (c).

$$i_c = h_{fe}\, i_b$$

$$v_o = \mu_m\, v_i$$

provide for transmission from control to output, but no means for transmission to occur from output back to the control terminals. The circuits in (b) and (d) show two more controlled relations, μ is known as *amplification factor*, g_m is known as *transconductance* and r_m is known as the *transresistance*.

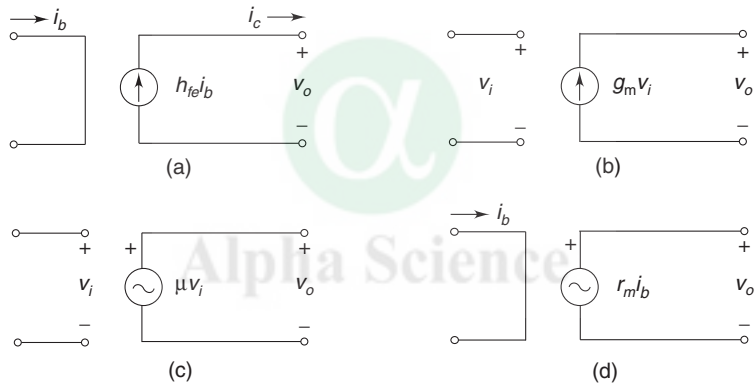


Fig. 6.3 *Ideal controlled sources*

Input terminals may not always be explicitly shown. Instead the control is derived from a variable of the system. Our actual sources will be treated as ideal, and internal resistances and capacitances will be added externally.

Such models are illustrated in Fig. 6.4, and can be recognized as Thevenin and Norton sources but with *control* of the generated voltage or current.

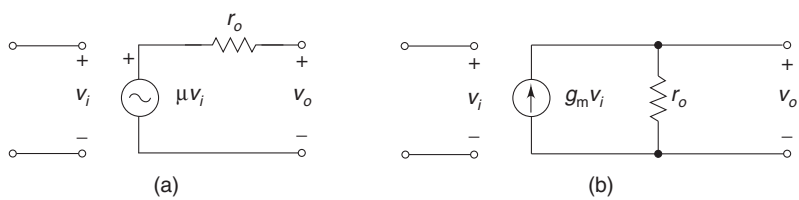


Fig. 6.4 *Modified sources*

6.3 HYBRID-PARAMETER EQUIVALENT CIRCUIT

An equivalent circuit may be developed for a linear active devices, such as transistor; the *h*-parameter equations are widely used in a model for such devices. The model for the network box of Fig. 6.1 requires

$$V_1 = h_i\, I_1 + h_r\, V_2 \tag{6.7}$$

$$I_2 = h_f\, I_1 + h_o\, V_2 \tag{6.8}$$

The first equation is a voltage summation, indicating a series arrangement of an impedance h_i and a transfer or controlled voltage generator $h_r\ V_2$ which feeds back to the input circuit a voltage proportional to the output voltage V_2. The left hand circuit of Fig. 6.5 is representative of Eq. (6.7). The second equation is a current summation at a node, suggesting a parallel arrangement of an admittance h_o or impedance $1/h_o$ and a transfer current source $h_f\, I_1$, which is controlled by the input current I_1. The right hand circuit of Fig. 6.5 is representive of Eq. (6.8).

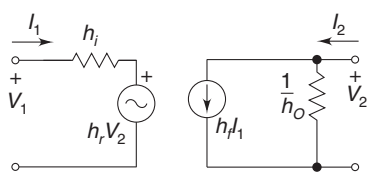


Fig. 6.5 *Hybrid-parameter equivalent circuit*

The complete circuit acts in accordance with the *h*-parameter equations derived from the arbitrary network, and is *equivalent* to that network box. By assigning *h* values as measured from a transistor, the circuit may be used to replace that transistor in a circuit analysis.

6.4 AN EQUIVALENT CIRCUIT FOR THE BJT

The active circuit model of Fig. 6.6(a) has two ports or four terminals. The bipolar junction transistor of Fig. 6.6(b) has three internal elements, and to represent the transistor by the *h*-parameter circuit requires that one of the transistor elements be made common to the input and output ports. Three possible circuit configurations result, as shown in Fig. 6.7. These yield diverse characteristics and operating properties. The conventional signal currents and voltages are defined in the figure.

To indicate the use of the *common-emitter* (CE) circuit, the *common base* (CB) circuit, or the *common-collector* (CC) circuit, a second subscript is used with the *h* parameters: h_{fe}, h_{ib}, and h_{oc}.Values of the *h*-parameters may be taken from manufacturer's data or obtained by measurement. Those for the common-emitter circuit are most generally available but conversion equations for all three configurations are given in Section 6.11.

The general *h*-parameter equivalent circuit is shown in Fig. 6.8(a). This circuit is derived from the following two equations on next page

$$v_1 = h_{11}\, i_1 + h_{12}\, v_2 \tag{6.9}$$

$$i_2 = h_{21}\, i_1 + h_{22}\, v_2 \tag{6.10}$$

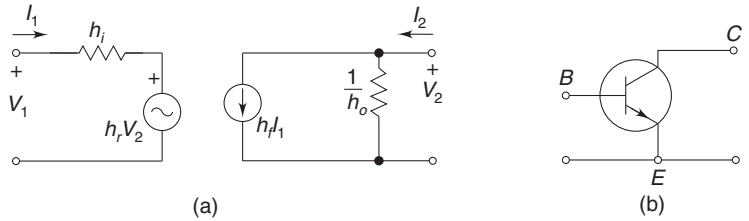


Fig. 6.6 *(a) h-parameter equivalent circuit; (b) transistor as a two port network*

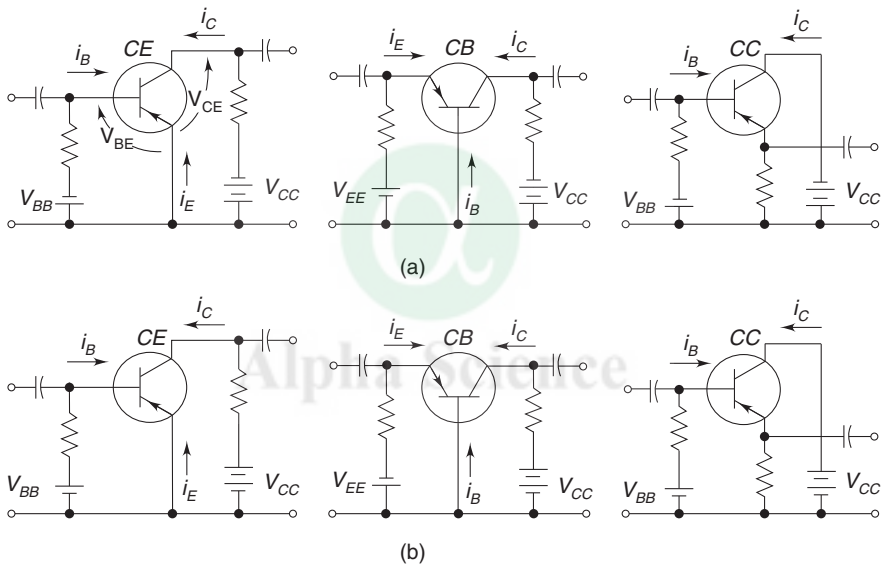


Fig. 6.7 *Defined currents and voltages: (a) PNP transistors; (b) NPN transistors*

The input circuit (Fig. 6.8) is derived with the help of Eq. (6.9). It appears as a resistance h_{11} in series with a voltage generator $h_{12}\, v_2$. The output circuit is derived with the help of Eq. (6.10). It involves a current generator $h_{21}\, i_1$ and a shunt resistance h_{22}. The hybrid equivalent circuit for the three orientations are shown in Fig. 6.8(b). The circuit is called as hybrid equivalent because its input port is a Thevenin equivalent or voltage generator with series resistance while output circuit is Norton equivalent i.e., current generator with shunt resistance. In the equivalent circuit, the effect of output upon input is represented by equivalent generator $h_{12}\, v_2$ and its value depends upon output voltage. Similarly, the effect of input upon output is represented by current generator $h_{21}\, i_1$ and its value depends upon input current.

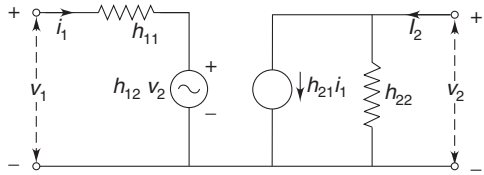


Fig. 6.8 *(a) h-parameter equivalent circuit*

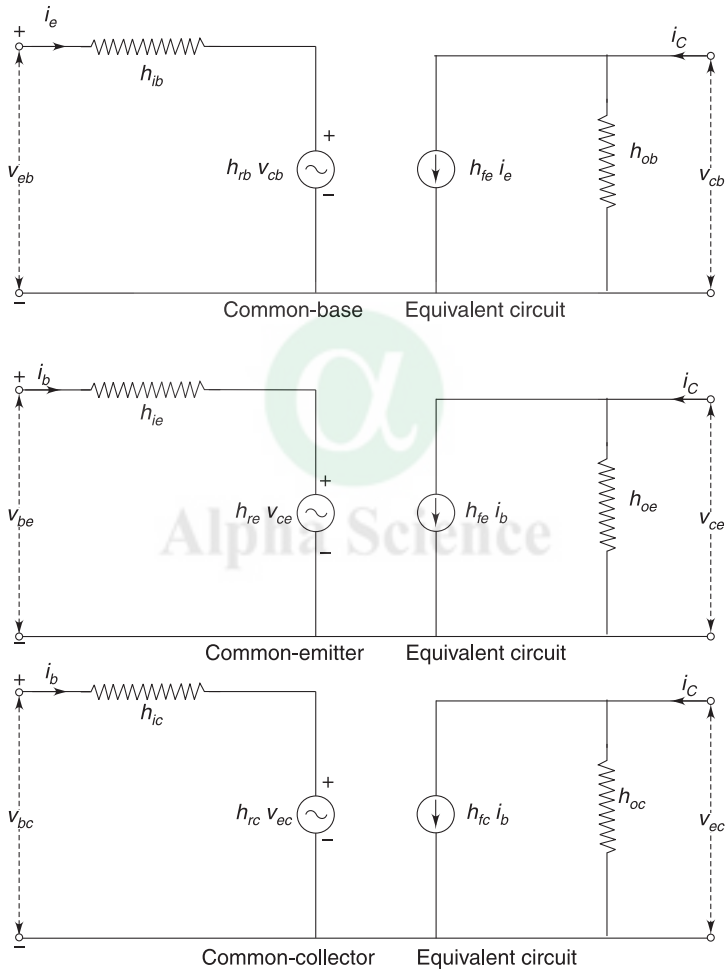


Fig. 6.8(b)

The three equivalent circuits of Fig. 6.8(b) correspond to the following three sets of equation:

$$\left.\begin{aligned} v_e &= h_{ib}\ i_e + h_{rb}\ v_c \\ i_c &= h_{fb}\ i_e + h_{ob}\ v_c \end{aligned}\right\} \quad \text{(CB)}$$

$$\left.\begin{aligned} v_b &= h_{ie}\, i_b + h_{re}\, v_c \\ i_c &= h_{fe}\, i_b + h_{oe}\, v_c \end{aligned}\right\} \quad \text{(CE)}$$

$$\left.\begin{aligned} v_b &= h_{ic}\, i_b + h_{rc}\, v_e \\ i_e &= h_{fc}\, i_b + h_{oc}\, v_e \end{aligned}\right\} \quad \text{(CC)}$$

6.5 GRAPHICAL DETERMINATION OF H-PARAMETERS

To illustrate the graphical method for determining the *h*-parameters, let us consider a transistor in CE mode. The characteristic of the transistor in CE mode are shown in Fig. 6.9.

From the V_{BE} versus I_B characteristic curves shown in Fig. 6.9(a), h_{ie} is determined at the *Q*-point as

$$h_{ie} = \left.\frac{\Delta V_{BE}}{\Delta I_B}\right|_{V_{CE} = \text{constant}}$$

Note that requiring the ac quantity v_{ce} to be zero is the same as requiring that the dc quantity V_{CE} to be held constant.

Referring to the V_{CE} versus I_C characteristic curve shown in Fig. 6.9(b), ΔI_C and ΔI_B are computed as marked on the curve in Fig. 6.9(b). From these two measurements h_{fe} is computed as

$$h_{fe} = \left.\frac{\Delta I_C}{\Delta I_B}\right|_{V_{CE} = \text{constant}}$$

Furthermore, from Fig. 6.9(c), ΔV_{BE} and ΔV_{CE} are computed as indicated and hence h_{re} is found from

$$h_{re} = \left.\frac{\Delta V_{BE}}{\Delta V_{CE}}\right|_{I_B = \text{constant}}$$

Referring to Fig. 6.9(d), h_{oe} is computed from the equation

$$h_{oe} = \left.\frac{\Delta I_C}{\Delta V_{CE}}\right|_{I_B = \text{constant}}$$

6.6 EXPERIMENTAL DETERMINATION OF *h*-PARAMETERS

1. Measurement of h_{ie} and h_{fe}

The circuit for the measurement of h_{ie} and h_{fe} is shown in Fig. 6.10. Here R_B is a large resistance that together with V_{CC} determines I_B. The resistance R_C is used to establish the desired d.c. voltage at the collector, and a very small resistance

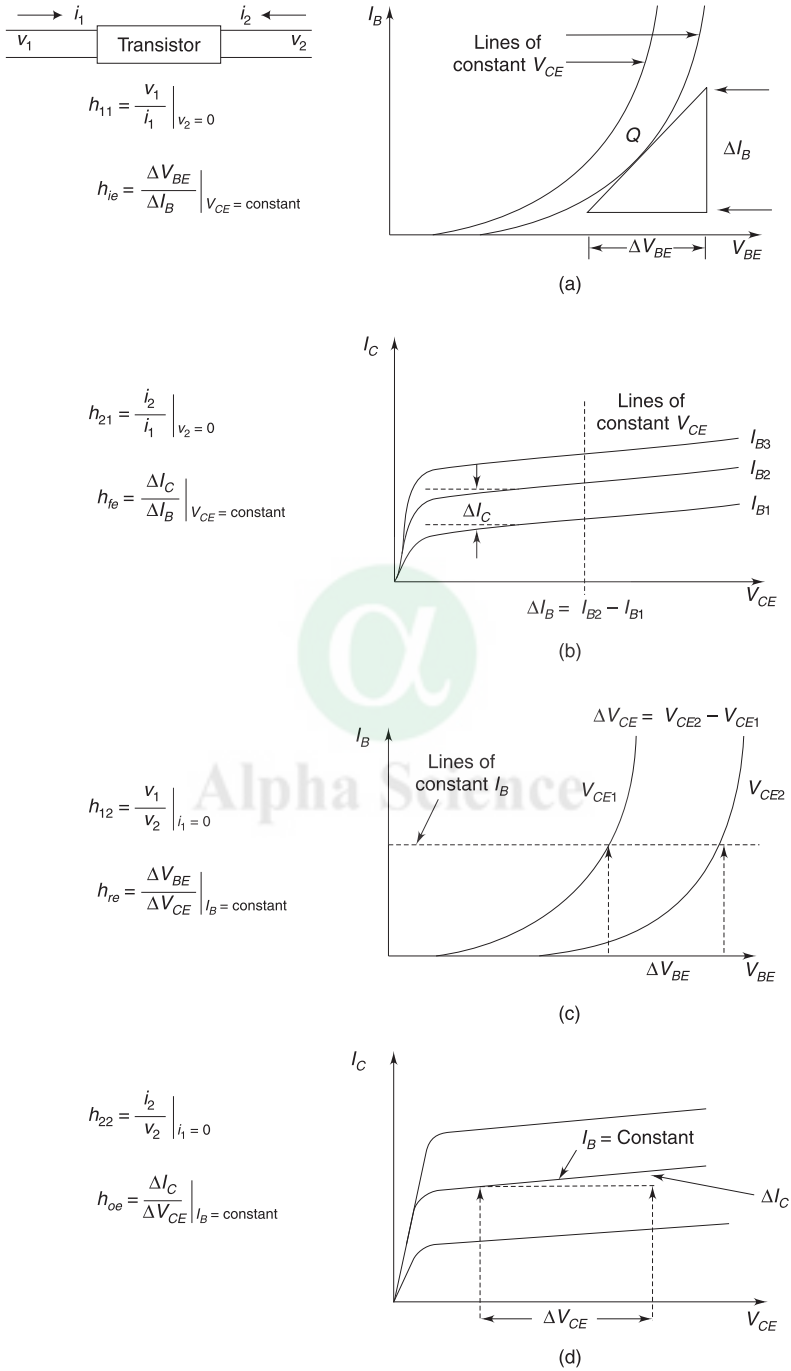


Fig. 6.9 *Graphical determination of common-emitter h-parameters*

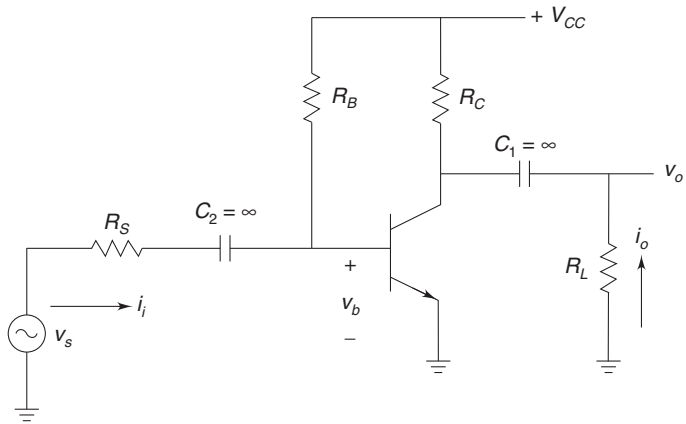


Fig. 6.10 *Circuit for measuring h_{ie} and h_{fe}*

R_L is used to enable measuring the signal current in the collector. Large value coupling capacitors are used to prevent d.c. currents to flow through R_L and R_S. Since R_L is small and the capacitor C_1 is large, the collector is effectively a.c. short-circuited to ground and thus

$$i_c \approx i_o \approx -\frac{v_o}{R_L} \tag{6.11}$$

The input signal current i_i is determined by measuring the voltage v across a known resistance R_S. If R_B is large and C_2 is also large, then

$$i_b \approx i_i \approx \frac{v}{R_S} \tag{6.12}$$

The input signal current v_b as labeled in Fig. 6.10 can be measured directly at the base. Using these measured values, we can compose h_{ie} and h_{fe}, using Eqs. (6.11) and (6.12) as

$$h_{ie} = \frac{v_b}{i_b} \quad \text{and} \quad h_{fe} = \frac{i_c}{i_b} \tag{6.13}$$

2. Measurement of h_{re}

To measure h_{re}, we use the circuit as shown in Fig. 6.11. Note that a signal source is used to excite the circuit at the collector side through a large value capacitor C_1. A high impedance ac voltmeter is used to measure the resulting a.c. voltage v_b across the base-emitter junction of the transistor. The bias resistor R_B should be kept much greater than the base-emitter a.c. resistance of the transistor. The value of h_{re} may then be determined as

$$h_{re} = \frac{v_b}{v_c} \tag{6.14}$$

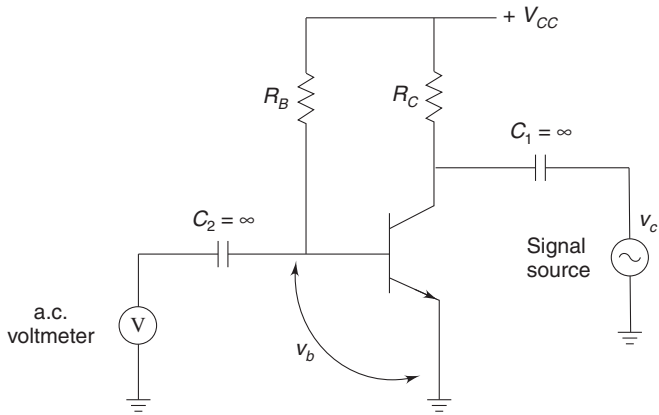


Fig. 6.11 *Circuit for measuring h_{re}*

3. Measurement of h_{oe}

We note that h_{oe} is the output conductance with the base open-circuited, which by definition is the slope of the V_{CE} versus I_C characteristic curves depicted in Fig. 6.8(d). The circuit in Fig. 6.11 can be used for measuring h_{oe}. Insert a small resistance in series with the signal source v_c and measure the a.c. current through it by measuring the a.c. voltage drop across it. Knowing v_c and i_c, we can compute the value of h_{oe} as

$$h_{oe} = \frac{i_c}{v_c} \tag{6.15}$$

6.7 PERFORMANCE OF A LINEAR CIRCUIT IN TERMS OF *h*-PARAMETERS

Consider a linear circuit with a load resistance R_L across the output terminals as shown in Fig. 6.12. In terms of h-parameters, we have

$$v_1 = h_{11}\, i_1 + h_{12}\, v_2 \tag{6.16}$$

$$i_2 = h_{21}\, i_1 + h_{22}\, v_2 \tag{6.17}$$

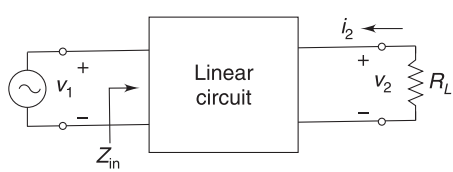


Fig. 6.12

From Fig. 6.12, $$i_2 = -\frac{v_2}{R_L} \tag{6.18}$$

Here negative sign is used because the load current is opposite to the direction of i_2.

1. Input Impedance

The input impedance Z_{in} of the circuit is the ratio of input voltage to the input current i.e.,

$$Z_{in} = \frac{v_1}{i_1} \tag{6.19}$$

From Eqs. (6.16) and (6.19)

$$Z_{in} = \frac{h_{11}\, i_1 + h_{12}\, v_2}{i_1} = \frac{h_{11} + h_{12}\, v_2}{i_1} \tag{6.20}$$

Substituting the value of i_2 from Eq. (6.18) in Eq. (6.17):

$$-\frac{v_2}{R_L} = h_{21}\, i_1 + h_{22}\, v_2$$

Solving, we get

$$\frac{v_2}{i_1} = \frac{-h_{21}}{\left(h_{22} + \frac{1}{R_L}\right)} \tag{6.21}$$

Substituting the value of v_2/i_1 from Eq. (6.21) in Eq. (6.20)

$$Z_{\text{in}} = (h_{11}) - \left[\frac{h_{12}\, h_{21}}{h_{22} + \left(\frac{1}{R_L}\right)}\right] \tag{6.22}$$

2. Current Gain

The current gain is given by $A_i = i_2/i_1$.

From Eqs. (6.17) and (6.18),

$$i_2 = h_{21}\, i_1 + h_{22}\, (-\, i_2\, R_L)$$

or $$i_2\, (1 + h_{22}\, R_L) = h_{21}\, i_1$$

$\therefore$ $$\frac{i_2}{i_1} = \frac{h_{21}}{1 + h_{22}\, R_L}$$

or $$A_i = \frac{h_{21}}{1 + h_{22}\, R_L} \tag{6.23}$$

3. Voltage Gain

The voltage gain is given by $A_v = v_2/v_1$.

Now, $$A_v = \frac{v_2}{v_1} = \frac{v_2}{i_1 Z_{in}}$$

From Eq. (6.21) $$\frac{v_2}{v_1} = \frac{-h_{21}}{\left(h_{22} + \frac{1}{R_L}\right)}$$

Therefore,

$$A_v = \frac{-h_{21}}{Z_{in}\left[h_{22} + \frac{1}{R_L}\right]} \tag{6.24}$$

6.8 MATHEMATICAL ANALYSIS OF COMMON-EMITTER AMPLIFIER USING HYBRID PARAMETERS

The common-emitter amplifier circuit is shown in Fig. 6.13(a) and the *h*-parameter equivalent circuit in Fig. 6.13(b).

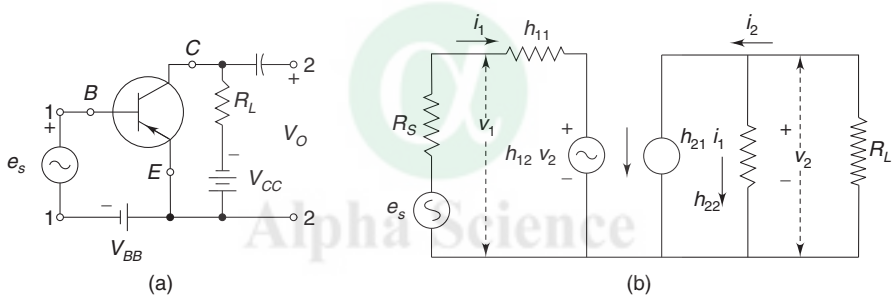


Fig. 6.13

Since the output blocking capacitor is of negligible reactance and since no signal voltage appears there, it is removed from the circuit of Fig. 6.13(b). it is assumed here that the input signals are small in accordance with the linearity requirements for the transistor.

Now, we know that the hybrid equations are

$$v_1 = h_{11}\, i_1 + h_{12}\, v_2 \tag{6.25}$$

$$i_2 = h_{21}\, i_1 + h_{22}\, v_2 \tag{6.26}$$

or

$$v_1 = h_{ie}\, i_1 + h_{re}\, v_2 \tag{6.27}$$

$$i_2 = h_{fe}\, i_1 + h_{oe}\, v_2 \tag{6.28}$$

Also, $$v_1 = e_s - i_1 R_S \tag{6.29}$$

and $$v_2 = -i_2 R_L \tag{6.30}$$

1. Current Gain

Substituting the value of v_2 from Eq. (6.30) in Eq. (6.28), we get

$$i_2 = h_{fe}\, i_1 - h_{oe}\, i_2\, R_L$$

or $$i_2\,(1 + h_{oe}\, R_L) = h_{fe}\, i_1$$

$$\therefore \quad \frac{i_2}{i_1} = \frac{h_{fe}}{1 + h_{oe}\, R_L}$$

and current gain

$$A_{ie} = \left(-\frac{i_2}{i_1}\right) = \frac{-\,h_{fe}}{(1 + h_{oe}\, R_L)} \tag{6.31}$$

2. Input Resistance

The input resistance is the resistance at the input terminals and is given by

$$R_{ie} = v_1/i_1$$

We know that

$$v_1 = h_{ie}\, i_1 + h_{re}\, v_2$$

or $$v_1 = h_{ie}\, i_1 - h_{re}\, i_2\, R_L \quad \text{(using Eq. (6.30))}$$

$$= h_{ie}\, i_1 + A_{ie}\, i_1\, h_{re}\, R_L \qquad \left(\because A_{ie} = -\frac{i_2}{i_1}\right)$$

$$\therefore \quad R_{ie} = \frac{v_1}{i_1} = h_{ie} + h_{re}\, A_{ie}\, R_L$$

or $$R_{ie} = h_{ie} - [h_{re}\, A_{ie}\, R_L\, /(1 + h_{oe}\, R_L)] \tag{6.32}$$

3. Voltage Gain

The voltage gain A_{ve} is the ratio of the output voltage v_2 to the input voltage v_1 i.e.,

$$A_{\text{ve}} = (v_2/v_1)$$

We know that

$$v_1 = h_{ie}\, i_1 + h_{re}\, v_2 \tag{6.33}$$

But $$i_1 = i_2\,(1 + h_{oe}\, R_L)/h_{fe} \quad \text{(from Eq. (6.31))}$$

or $$i_1 = -\,v_2\,(1 + h_{oe}\, R_L)/(h_{fe}\, R_L) \tag{6.34}$$

$$\left(\because i_2 = -\frac{v_2}{R_L}\right)$$

Substituting the value of i_1 from Eq. (6.34) in Eq. (6.33),

$$v_1 = h_{ie} - \left[\frac{v_2\,(1 + h_{oe}\,R_L)}{h_{fe}\,R_L}\right] + h_{re}\,v_2$$

$$= v_2\left[h_{re} - \frac{h_{ie}\,(1 + h_{oe}\,R_L)}{h_{fe}\,R_L}\right]$$

$$= v_2\left[\frac{h_{re}\,h_{fe}\,R_L - h_{ie}\,(1 + h_{oe}\,R_L)}{h_{fe}\,R_L}\right]$$

$$= v_2\left[\frac{h_{re}\,h_{fe}\,R_L - h_{ie} - h_{ie}\,h_{oe}\,R_L}{h_{fe}\,R_L}\right]$$

$$= -\,v_2\left[\frac{h_{ie} + (h_{ie}\,h_{oc} - h_{re}\,h_{fe})\,R_L}{h_{fe}\,R_L}\right]$$

or
$$A_{ve} = \frac{v_2}{v_1} = \frac{-\,h_{fe}\,R_L}{h_{ie} + (h_{ie}\,h_{oe} - h_{re}\,h_{fe})\,R_L} \tag{6.35a}$$

Voltage Gain in Terms of Current Gain

$$A_{ve} = \frac{v_2}{v_1} = \frac{-\,i_2\,R_L}{v_1} = \frac{A_{ie}\,i_1\,R_L}{v_1}$$

or
$$A_{ve} = \frac{A_{ie}\,R_L}{R_{ie}}$$

From Eq. (6.35a)

$$A_{ve} = \frac{-\,h_{fe}\,R_L}{h_{ie} + \Delta R_L} \tag{6.35b}$$

where
$$\Delta \equiv h_{ie}\,h_{oe} - h_{re}\,h_{fe}$$

4. Power Gain

It is given by

$$A_{be} = A_{ie}\,A_{ve}$$

Now,
$$A_{ie} = \frac{-\,h_{fe}}{(1 + h_{oe}\,R_L)}$$

and
$$A_{ve} = \frac{-\,h_{fe}\,R_L}{h_{ie} + \Delta R_L}$$

$\therefore$
$$A_{be} = \frac{h_{fe}^2\,R_L}{(1 + h_{oe}\,R_L)\,(h_{ie} + \Delta R_L)} \tag{6.36}$$

5. Output Resistance

For calculation of output resistance, we have to transfer the signal generator to the output terminals Fig. 6.14)

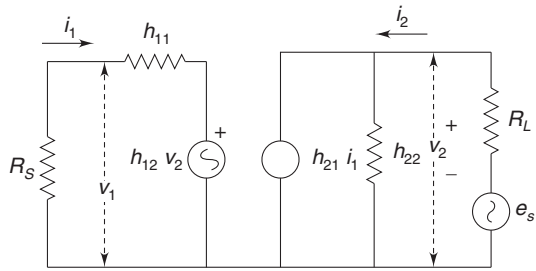


Fig. 6.14

The equations are

$$0 = (R_S + h_{11})\ i_1 + h_{12}\ v_2 \tag{6.37}$$

$$i_2 = h_{21}\ i_1 + h_{22}\ v_2 \tag{6.38}$$

From Eq. (6.37),

$$i_1 = \frac{-\ h_{12}\ v_2}{R_S + h_{11}} \tag{6.39}$$

Substituting this in Eq. (6.38), we get

$$i_2 = \frac{-\ h_{12}\ h_{21}\ v_2}{R_S + h_{11}} + h_{22}\ v_2$$

or

$$R_{oe} = \frac{v_2}{i_2} = \frac{R_S + h_{11}}{(R_S + h_{11})\ h_{22} - h_{12}\ h_{21}}$$

$$= \frac{R_S + h_{ie}}{R_S\ h_{oe} + \Delta} \tag{6.40a}$$

$$= \left[h_{oe} - \frac{h_{fe}\ h_{re}}{h_{ie} + R_S}\right]^{-1} \tag{6.40b}$$

6. Voltage Gain Considering Source Resistance R_S

The actual voltage applied to the input of the amplifier is e_s. Hence the overall voltage gain is given by

$$(A_{ve})_s = \frac{v_2}{e_s} = \frac{v_2}{v_1} \cdot \frac{v_1}{e_s} = A_{ve}\ \frac{v_1}{e_s} \tag{6.41}$$

Corresponding to the input impedance R_{ie}, the value of v_1 is given by

$$v_1 = \frac{e_s \cdot R_{ie}}{R_S + R_{ie}} \tag{6.42}$$

Consequently

$$(A_{ve})_s = A_{ve}\left(\frac{R_{ie}}{R_S + R_{ie}}\right) \tag{6.43}$$

or

$$(A_{ve})_s = \frac{-h_{fe} R_L}{h_{ie} + \Delta R_L}\left(\frac{R_{ie}}{R_S + R_{ie}}\right) \tag{6.44}$$

If $R_S = 0$, $(A_{ve})_s = A_{ve}$

7. Current Gain Considering the Source Resistance R_S

Let a current source i_S with internal resistance R_S be at the input side, then overall current gain is given by

$$(A_{ie})_S = \frac{i_2}{i_1} \cdot \frac{i_1}{i_S} = A_{ie} \cdot \frac{i_1}{i_S}$$

Now

$$i_1 = \frac{i_S R_S}{R_{ie} + R_S}$$

$\therefore$

$$(A_{ie})_S = A_{ie} \cdot \frac{R_S}{R_S + R_{ie}} \tag{6.45}$$

If $R_S = \infty$, then $(A_{ie})_S = A_{ie}$.

The subscript e affixed to each parameters indicates that these parameters belong to a common emitter transistor.

Table 6.1 *Important formulae for common emitter transistor amplifier using h-parameters*

Quantity ↓	Expression ↓
Current gain	$A_{ie} = -h_{fe}/(h_{oe} R_L + 1)$
Input resistance	$R_{ie} = h_{ie} + h_{re} A_{ie} R_L$
Voltage gain	$A_{ve} = A_{ie} R_L/R_{ie}$
Power gain	$A_{pe} = A_{ve} A_{ie}$
Output resistance	$R_{oe} = [h_{oe} - h_{fe} h_{re}/(h_{ie} + R_S)]^{-1}$

6.9 MATHEMATICAL ANALYSIS OF COMMON BASE AMPLIFIER USING HYBRID PARAMETERS

Hybrid equivalent circuit of a common base transistor amplifier is shown in Fig. 6.15.

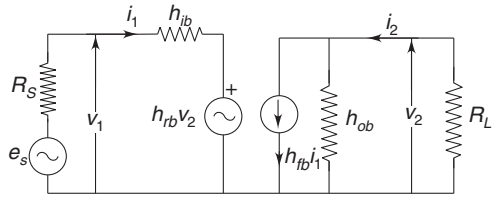


Fig. 6.15

The equations are

$$v_1 = h_{ib}\, i_1 + h_{rb}\, v_2$$

$$i_2 = h_{fb}\, i_1 + h_{ob}\, v_2 \tag{6.46}$$

The second two equations needed are

$$v_1 = e_S - i_1\, R_S \tag{6.47}$$

$$v_2 = - i_2\, R_L \tag{6.48}$$

Substituting the value of v_2 from Eq. (6.48) in Eq. (6.46), we get

$$i_2 = h_{fb}\, i_1 - h_{ob}\, i_2\, R_L$$

or $$i_2 + h_{ob}\, i_2\, R_L = h_{fb}\, i_1$$

$\therefore$ $$i_2\,(1 + h_{ob}\, R_L) = h_{fb}\, i_1 \tag{6.49}$$

1. Current Gain

$$A_{ib} = - \frac{i_2}{i_1} = \frac{- h_{fb}}{(1 + h_{ob}\, R_L)} \tag{6.50}$$

2. Input Resistance

We know that

$$v_1 = h_{ib}\, i_1 + h_{re}\, v_2$$

or $$v_1 = h_{ib}\, i_1 - h_{rb}\, i_2\, R_L \qquad \because v_2 = - i_2\, R_L$$

or $$v_1 = h_{ib}\, i_1 + A_{ib}\, i_1\, R_L\, h_{rb} \qquad \because A_{ib} = - i_2/i_1$$

$\therefore$ $$R_{ib} = \frac{v_1}{i_1} = h_{ib} + h_{rb}\, A_{ib}\, R_L$$

or $$R_{ib} = h_{ib} + h_{rb} \left(- \frac{h_{fb}}{1 + h_{ob}\, R_L} \right) R_L$$

$$= h_{ib} - \frac{h_{rb}\, h_{fb}\, R_L}{(1 + h_{ob}\, R_L)} \tag{6.51}$$

3. Voltage Gain

$$A_{vb} = \frac{v_2}{v_1} = \frac{-i_2 R_L}{v_1} = \frac{A_{ib}\, i_1 R_L}{v_1}$$

or $$A_{vb} = \frac{A_{ib} R_L}{R_{ib}}$$

$\therefore$ $$A_{vb} = \frac{-h_{fb}}{(1 + h_{ob} R_L)} \cdot \frac{R_L}{[h_{ib} - \{h_{rb} h_{fb} R_L/(1 + h_{ob} R_L)\}]}$$

$$= \frac{-h_{fb} R_L}{h_{ib}(1 + h_{ob} R_L) - h_{rb} h_{fb} R_L}$$

$$= \frac{-h_{fb} R_L}{h_{ib} + (h_{ib} h_{ob} - h_{rb} h_{fb}) R_L}$$

$$= \frac{-h_{fb} R_L}{(h_{ib} + R_L \Delta)} \qquad (6.52)$$

where $$\Delta = h_{ib} h_{ob} - h_{rb} h_{fb}$$

4. Power Gain

It is given by

$$A_{pb} = A_{ib} A_{vb} = \frac{h_{fb}^2 R_L}{(1 + h_{ob} R_L)(h_{ib} + R_L \Delta)} \qquad (6.53)$$

5. Output Resistance

For the calculation of output resistance, we have to transfer the signal generator to the output terminals. At the same time, we place only R_S across the input terminals Fig. 6.16.

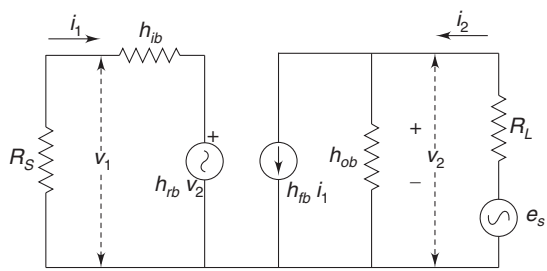


Fig. 6.16

The equations are

$$= (R_S + h_{ib})\, i_1 + h_{rb}\, v_2$$

and $$i_2 = h_{fb}\, i_1 + h_{ob}\, v_2$$

Putting i_1 from first equation into second, we get

$$i_2 = -\frac{h_{fb}\, h_{rb}}{(R_S + h_{ib})}\, v_2 + h_{ob}\, v_2$$

or
$$R_{ob} = \frac{v_2}{i_2} = \frac{R_S + h_{ib}}{(R_S + h_{ib})\, h_{ob} - h_{fb}\, h_{rb}}$$
$$= \frac{R_S + h_{ib}}{R_S\, h_{ob} + h_{ib}\, h_{ob} - h_{fb}\, h_{rb}}$$
$$= \frac{R_S + h_{ib}}{R_S\, h_{ob} + \Delta}$$

where
$$\Delta = h_{ib}\, h_{ob} - h_{fb}\, h_{rb} \tag{6.54a}$$

or
$$R_{ob} = \left[h_{ob} - \frac{h_{fb}\, h_{rb}}{h_{ib} + R_S}\right]^{-1} \tag{6.54b}$$

6. Voltage Gain Considering Source Resistance R_S (Overall Voltage Gain)

Overall voltage gain of the amplifier is given by

$$(A_{vb})_S = \frac{v_2}{e_S} = \frac{v_2}{v_1} \cdot \frac{v_1}{e_S} = A_{vb}\, \frac{v_1}{e_S}$$

corresponding to input impedance R_{ib}, the value of v_1 is given by

$$v_1 = \frac{e_S\, R_{ib}}{R_S + R_{ib}}$$

Hence
$$(A_{vb})_S = A_{vb}\, \frac{R_{ib}}{R_S + R_{ib}} \tag{6.55}$$

If $R_S = 0$
$$(A_{vb})_S = A_{vb}$$

7. Current Gain Considering the Source Resistance R_S (Overall current Gain)

Let a current source i_S with internal resistance R_S be at the input side, then overall current gain is given by

$$(A_{vb})_S = \frac{i_2}{i_1} \cdot \frac{i_1}{i_S} = A_{ib} \cdot \frac{i_1}{i_S}$$

Now
$$i_1 = \frac{i_S\, R_S}{R_{ib} + R_S}$$

$\therefore$
$$(A_{ib})_S = A_{ib} \cdot \frac{R_S}{R_S + R_{ib}} \tag{6.56}$$

If $(A_{ib})_S = \infty$, then $(A_{ib})_S = A_{ib}$.

Table 6.2 *Important formulae for common base transistor amplifier using h- parameters*

Quantity ↓	Expressions ↓
Current gain	$A_{ib} = -h_{if}/(h_{ob} R_L + 1)$
Input resistance	$R_{ib} = h_{ib} + h_{rb} A_{ib} R_L$
Voltage gain	$A_{vb} = A_{ib} R_L/R_{ib}$
Power gain	$A_{pb} = A_{vb} \cdot A_{ib}$
Output resistance	$R_{ob} = [h_{ob} - h_{fb} h_{rb}/(h_{ib} + R_S)]^{-1}$
Overall voltage gain	$(A_{vb})_S = A_{vb} \cdot R_{ib}/(R_S + R_{ib})$
Overall current gain	$(A_{ib})_S = A_{ib} \cdot R_S/(R_S + R_{ib})$

6.10 MATHEMATICAL ANALYSIS OF COMMON COLLECTOR TRANSISTOR USING HYBRID PARAMETERS

Hybrid equivalent circuit of common collector transistor amplifier is shown in Fig. 6.17.

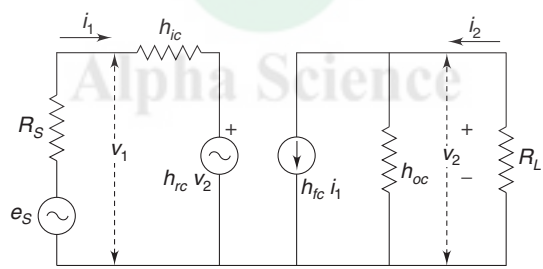


Fig. 6.17

The equations are

$$v_1 = h_{ic}\, i_1 + h_{rc}\, i_2 \tag{6.57}$$

$$i_2 = h_{fc}\, i_1 + h_{oc}\, r_2 \tag{6.58}$$

The second two equations needed are

$$v_1 = e_S - i_1\, R_S \tag{6.59}$$

$$v_2 = -\, i_2\, R_L \tag{6.60}$$

Substituting the value of v_2 from Eq. (6.60) in Eq. (6.58), we get

$$i_2 = h_{fc}\, f_1 - h_{oc}\, i_2\, R_L \quad \text{or} \quad i_2 + h_{oc}\, i_2\, R_L = h_{fc}\, i_1$$

$$i_2\,(1 + h_{oc}\, R_L) = h_{fc}\, i_1 \tag{6.61}$$

1. Current Gain

$$A_{ic} = -\frac{i_2}{i_1} = -\frac{-h_{fc}}{(1 + h_{oc} R_L)} \tag{6.62}$$

2. Input Resistance

We know that

$$v_1 = h_{ic}\, i_1 + h_{rc}\, v_2$$

or $$v_1 = h_{ic}\, i_1 - h_{rc}\, i_2 R_L \qquad \because v_2 = -i_2 R_L$$

or $$v_1 = h_{ic}\, i_1 + A_{ic}\, i_1 R_L h_{rc} \qquad \because A_{ic} = -i_2/i_1$$

$\therefore$ $$R_{ic} = \frac{v_1}{i_1} = h_{ic} + h_{rc} A_{ic} R_L$$

$$= h_{ic} - \frac{h_{rc} h_{fc} R_L}{(1 + h_{oc} R_L)} \tag{6.63}$$

3. Voltage Gain

Voltage gain in terms of current gain

$$A_{vc} = \frac{v_2}{v_1} = \frac{-i_2 R_L}{v_1} = \frac{A_{ic}\, i_1 R_L}{v_1}$$

or $$A_{vc} = \frac{A_{ic} R_L}{R_{ic}}$$

$\therefore$ $$A_{vc} = \left(\frac{-h_{fc} R_L}{1 + h_{oc} R_L}\right)\left[\frac{1}{h_{ic} - \{h_{rc} h_{fc} R_L/(1 + h_{oc} R_L)\}}\right]$$

$$= \frac{-h_{fc} R_L}{h_{ic} (1 + h_{oc} R_L) - h_{rc} h_{fc} R_L}$$

$$= \frac{-h_{fc} R_L}{h_{ic} + (h_{ic} h_{oc} - h_{rc} h_{fc}) R_L}$$

$$= \frac{-h_{fc} R_L}{(h_{ic} + R_L \Delta)} \tag{6.64}$$

where $$\Delta = h_{ic} h_{oc} - h_{rc} h_{fc}$$

4. Power Gain

It is given by

$$A_{pc} = A_{ic} A_{vc} = \frac{h_{fc}^2 R_L}{(1 + h_{oc} R_L)(h_{ic} + + R_L \Delta)} \tag{6.65}$$

5. Output Resistance

For calculation of the output resistance, we have to transfer the signal generator to the output terminals. At the same time, we place only R_S across the input terminals (Fig. 6.18). The equation are

$$0 = (R_S + h_{ic})\, i_1 + h_{rc}\, v_2$$

$$i_2 = h_{fc}\, i_1 + h_{oc}\, v_2$$

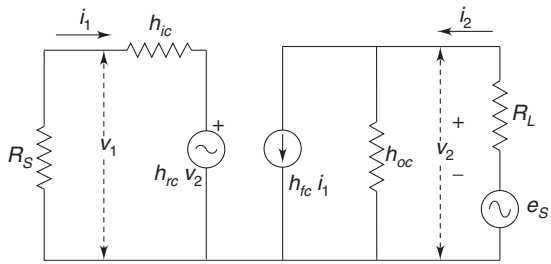


Fig. 6.18

Putting i_1 from first equation into second, we get

$$i_2 = -\frac{h_{fc}\, h_{rc}\, v_2}{(R_S + h_{ic})} + h_{rc}\, v_2$$

or
$$R_{oc} = \frac{v_2}{i_2}$$

$$= \frac{R_S + h_{ic}}{(R_S + h_{ic})\, h_{rc} - h_{fc}\, h_{rc}}$$

or
$$R_{oc} = \frac{R_S + h_{ic}}{R_S + h_{oc} + \Delta} \qquad (6.66a)$$

or
$$R_{oc} = \left[h_{oc} - \frac{h_{fc}\, h_{rc}}{h_{ic} + R_S}\right]^{-1} \qquad (6.66b)$$

6. Voltage Gain Considering Source Resistance (Overall Voltage Gain)

The overall voltage gain of the amplifier is given by

$$(A_{vc})_S = \frac{v_2}{e_S} = \frac{v_2}{v_1} \cdot \frac{v_1}{e_S} = A_{vc}\, \frac{v_1}{e_S}$$

Corresponding to input impedance R_{ic}, the value of v_1 is given by

$$v_1 = \frac{e_S\, R_{ic}}{R_S + R_{ic}}.$$

Hence $$(A_{vc})_S = A_{vc} \frac{R_{ic}}{R_S + R_{ic}} \tag{6.67}$$

If $R_S = 0$, $(A_{vc})_S = A_{vc}$.

7. Current Gain Corresponding the Source Resistance *R* (Overall Current Gain)

Let a current source i_S with internal resistance R_S be at the input side, then overall current gain is given by

$$(A_{ic})_S = \frac{i_2}{i_1} \cdot \frac{i_1}{i_S} = A_{ic} \frac{i_1}{i_S}.$$

Now $$i_1 = \frac{i_S R_S}{R_{ic} + R_S}$$

$\therefore$ $$(A_{iv})_S = A_{ic} \frac{R_S}{R_S + R_{ic}} \tag{6.68}$$

If $R_S = \infty$, $(A_{ic})_S = A_{ic}$.

Table 6.3 *Important formulae for common collector transistor amplifier using h-parameter*

Quantity ↓	Expressions ↓
Current gain	$A_{ic} = -h_{fc}/(h_{oc} R_L + 1)$
Input resistance	$R_{ic} = h_{ic} + h_{rc} A_{ic} R_L$
Voltage gain	$A_{vc} = A_{ic} R_L/R_{ic}$
Power gain	$A_{pc} = A_{vc} \cdot A_{ic}$
Output resistance	$R_{oc} = [h_{oc} - h_{fc} h_{rc}/(h_{ic} + R_S)]^{-1}$
Overall voltage gain	$(A_{vc})_S = A_{vc} \cdot R_{ic}/(R_S + R_{ic})$
Overall current gain	$(A_{ic})_S = A_{ic} \cdot R_S/(R_S + R_{ic})$

6.11 CONVERSION OF THE *h*-PARAMETERS

Transistor manufacturers usually publish the *h*-parameters of a transistor in the CE form. For those cases where the data are given in CB or CC form, we present conversion relations. The rigorous forms of many of these equations are difficult to use, and the approximate forms are obtained by application of the usual magnitude assumptions.

$$h_{ie} \approx \frac{h_{ib}}{1 + h_{fb}} \qquad h_{ib} \approx \frac{h_{ie}}{1 + h_{fe}} \qquad h_{ic} = h_{ie}$$

$$h_{oe} \approx \frac{h_{ob}}{1 + h_{fb}} \qquad h_{ob} \approx \frac{h_{oe}}{1 + h_{fe}} \qquad h_{oc} = h_{oe}$$

$$h_{fe} \approx \frac{-h_{fb}}{1 + h_{fb}} \qquad h_{fb} \approx \frac{-h_{fe}}{1 + h_{fe}} \qquad h_{fc} = -(1 + h_{fe})$$

$$h_{re} \approx \frac{\Delta_{hb} - h_{rb}}{1 + h_{fb}} \qquad h_{rb} \approx \frac{\Delta_{he} - h_{ce}}{1 + h_{fe}} \qquad h_{rc} = 1$$

6.12 VALUES OF *h*-PARAMETERS OF A TYPICAL BJT UNDER CE, CB AND CC CONFIGURATIONS

Table 6.4 gives the values of *h*-parameters.

Table 6.4 *Values of h-parameters of a typical junction transistor*

Parameters ↓	CE ↓	CB ↓	CC ↓
$h_{11} = h_i$	1100 Ω	22 Ω	1100 Ω
$h_{12} = h_r$	2.5×10^{-4}	3×10^{-4}	≈ 1
$h_{21} = h_f$	50	− 0.98	− 51
$h_{22} = h_e$	25 μA/V	0.5 μA/V	25 μA/V
$\frac{1}{h_o}$	40 kΩ	2 MΩ	40 kΩ

6.13 CLASSIFICATION OF AMPLIFIERS

An amplifier is a circuit meant to amplify a signal with a minimum of distortion, so as to make it more useful. The classification of amplifiers is somewhat involved. A complete classification must include information about the following:

(i) Active device used.
(ii) Frequency range of operation.
(iii) Coupling scheme used.
(iv) Ultimate purpose of the circuit.
(v) Condition of the bias and magnitude of signal.

An amplifier may use either a semiconductor device (such as BJT or FET) or a vacuum-tube device (such as triode, pentode, or beam power tube).

Based on frequency range of operation, the amplifiers may be classified as follows:

1. D.C. amplifiers (from zero to about 10 Hz)
2. Audio amplifiers (30 Hz to about 15 kHz)
3. Video or wide-band amplifiers (up to a few MHz)
4. RF amplifiers (a few kHz to hundreds of MHz)

Usually, in an amplifier system, a number of stages are used. These stages may be cascaded by either direct coupling, *RC* coupling, or transformer coupling. Sometimes, *LC* (inductance capacitance) coupling is also used. Accordingly, the amplifiers are classified as:

1. Direct-coupled amplifiers
2. *RC*-coupled amplifiers
3. Transformer-coupled amplifiers
4. *LC*-coupled amplifiers

Depending upon the ultimate purpose of an amplifier, it may be broadly classified as either voltage (small-signal), or power (large-signal) amplifier. Till now, we had considered the voltage amplifiers. In the next unit, we shall discuss the power amplifiers.

The amplifiers may also be classified according to where the quiescent point is fixed and how much the magnitude of the input signal is. Accordingly, four classes of operation for either transistor- or tube-amplifiers are defined as follows:

Class A In class A operation, the transistor stays in the active region throughout the ac cycle. The *Q* point and the input signal are such as to make the output current flow for 360°, as shown in Fig. 6.19(a).

Class B In class B operation, the transistor stays in the active region only for half the cycle. The *Q* point is fixed at the cut-off point of the characteristics. The power drawn from the dc power supply, by the circuit, under quiescent conditions is small. The output current flows only for 180° (see Fig. 6.19(b)).

Class AB This operation is between class A and B. The transistor is in the active region for more than half the cycle, but less than the whole cycle. The output current flows for more than 180° but less than 360° (see Fig. 6.19(c)).

Class C In a class C amplifier, the *Q* point is fixed beyond the extreme end of the characteristics. The transistor is in the active region for less than half cycle. The output current remains zero for more than half cycle, as shown in Fig. 6.19(d). The dc current drawn from the power supply is very small.

In case of a vacuum-tube amplifier, sometimes a suffix 1 or 2 may be added to the class of operation. The suffix 1 indicates the absence of grid current, whereas suffix 2 indicates that during some part of the cycle, grid current flows. Thus, the designation class AB_1 will mean that the amplifier operates in class AB condition and that no grid current flows during any part of the cycle.

In this unit, as well as in the previous unit, we have considered small-signal amplifiers under class A operation.

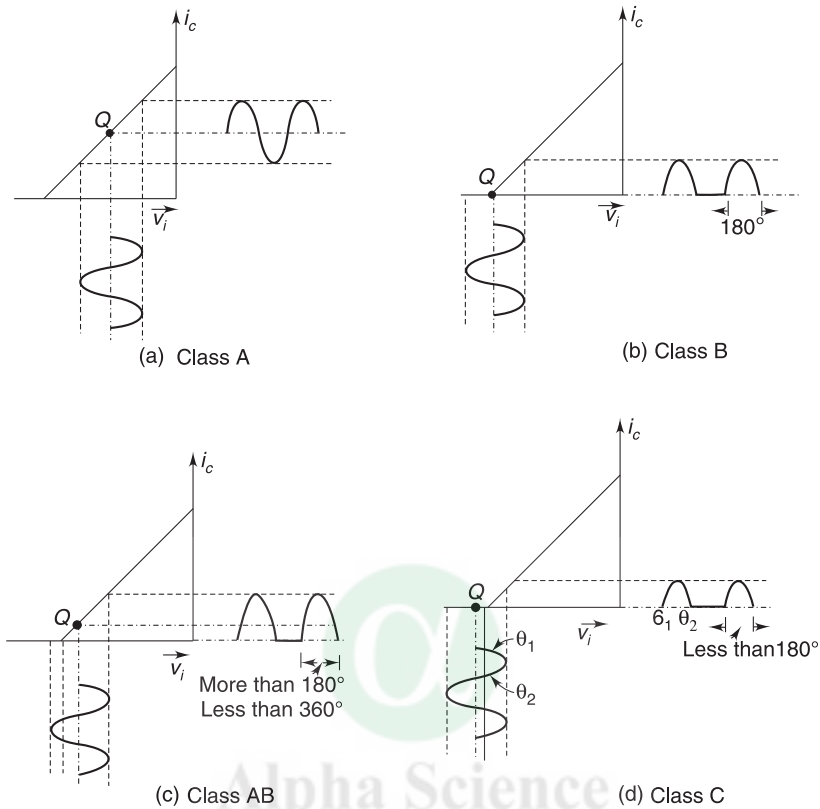


Fig. 6.19 *Classification of amplifiers based on the biasing condition*

(For details of class A, class B and class C operation, the reader is referred to "Electronic Principles" by Albert Paul Malvino, Tata McGraw-Hill, N. Delhi (1984)).

6.14 MULTISTAGE AMPLIFIERS

The performance of a single stage amplifier has already been discussed in section 5.17. Although the gain of an amplifier does depend on the parameters of the device and circuit components, there exits an upper theoretical limit for the gain obtainable from a single stage. Thus in actual voltage practice, in order to obtain the desired voltage amplification or power gain or frequency response, we use two or more or frequency response, we use two or more stages in *cascade*, called multistage amplifier. In a multistage amplifier, the output of first stage is combined to the next stage through a coupling device (This process is known as *cascading*). The coupling device is used to:

(i) Transfer the a.c. output of one stage to the input of the next stage and

(ii) Block the d.c. to pass from one stage to the next stage.

(iii) Although some voltage loss of signal cannot be avoided in the coupling network but this loss should be minimum (just negligible).

(iv) The coupling network should offer equal impedance to the various frequencies of signal wave. In other words, the network impedance should not be frequency dependent.

Unfortunately, there is no coupling network which fulfills all the above demands. Depending on the requirements, There are four basic methods of coupling viz., *R-C* coupled amplifier, transformer coupled amplifier, and direct coupled amplifier.

The coupling network not only couples two stages, it also forms a part of the load impedance of the preceding stage.

RC coupling is the most commonly used coupling between the two stages of a cascaded or multistage amplifier because it is cheaper in cost and very compact circuit and provides excellent frequency response.

The most suitable transistor configuration for cascading is CE configuration because the voltage gain of common emitter amplifier is greater than unity while CC configuration has voltage gain less than unity and the voltage-gain of CB configuration using cascading is also less than unity (Note: No doubt, the voltage gain of a single CB amplifier is more than unity, but the overall voltage gain of multistage amplifier using CB configuration is low, almost equal to the voltage gain of the last stage alone). It is also noteworthy point that for input stage, the consideration is not the maximum voltage gain but the impedance matching of the source with the input impedance of the input stage. In certain cases, choice of configuration for input stage is the minimization of noise and maximization of signal to noise power ratio.

6.14.1 Gain of *n*-Cascaded Stages

A multistage amplifier can be represented by a block diagram, as shown in Fig. 6.20. It is to be noted that the output of the first stage makes the input for the second stage, the output of second stage makes the input for third stage and so on. The signal voltage V_s is applied to the input of the first stage and the final output V_{out} is available at the output terminals of the last stage.

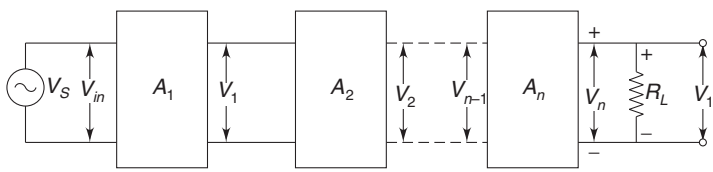


Fig. 6.20 *n-stage amplifier*

Input to the first stage,

$$V_{\text{in}} = \text{signal voltage } V_s$$

Output to the first stage or input to the second stage

$$V_1 = A_{v_1} V_{\text{in}}$$

where A_{v_1} is the voltage gain of first stage

Output of second stage or input to the third stage

$$V_2 = A_{v_2} V_1$$

where A_{v_2} is the voltage gain of the second stage

Similarly the output of nth stage (or final output)

$$V_{\text{out}} = V_n = A_{v_n} V_{n-1}$$

where A_{v_n} is the voltage gain of the last stage

Overall voltage gain of the amplifier is given as

$$A = \frac{V_{\text{out}}}{V_S}$$

(visualizing the multistage amplifier as a single amplifier with input voltage V_s and output voltage V_{out})

$$= \frac{V_1}{V_S} \times \frac{V_2}{V_1} \times \frac{V_3}{V_2} \times \ldots \frac{V_{n-1}}{V_{n-2}} \times \frac{V_n}{V_{n-1}}$$

$$= A_{v_1} \times A_{v_2} \times \ldots \times A_{v_{n-1}} \times A_{v_n} \tag{6.69}$$

i.e., the gain of a multistage amplifier is equal to the product of gains of individual stages. It is worthwhile to mention her that in practice total gain A is less than $A_{v_1} \times A_{v_2} \times \ldots \times A_{v_{n-1}} \times A_{v_n}$ due to the loading effects of the following stages.

When the gains are expressed in dB, the overall gain of a multistage amplifier is given as the sum of gains of individual stages in decibels (dB).

Taking logarithm (to the base 10) of Eq. (6.69) and then multiplying each term by 20 we have

$$20 \log_{10} A_v = 20 \log_{10} A_{v_1} + 20 \log_{10} A_{v_2} + \ldots + 20 \log_{10} A_{v_{n-1}} + 20 \log_{10} A_{v_n}$$

In the above equation, the term to the left is the overall gain of the multistage amplifier expressed in decibels. The terms on the right denote the gains of the individual stages expressed in decibels. Thus

$$A_{v\,\text{dB}} = A_{v\,\text{dB}_1} + A_{v\,\text{dB}_2} + \ldots A_{v\,\text{dB}_n} \tag{6.70}$$

Example 6.2 *A multistage amplifier consists of three stages. The voltage gains of stages are 60, 100, 160. Calculate the overall voltage gain in dB.*

Solution Voltage gain of first stage in dB = 20 $\log_{10}$ 60 = 35.563 dB

Voltage gain of second stage in dB = 20 $\log_{10}$ 100 = 40 dB

Voltage gain of third stage in dB = 20 $\log_{10}$ 160 = 44.082 dB

Overall voltage gain of the amplifier = 35.536 + 40 + 44.082 = 119.645 db

6.14.2 RC Coupled, *n*-Stage Cascaded Amplifier

Several amplifier stages are usually cascaded to increase the overall voltage gain of the amplifier. However, sometimes cascading is done to obtain the desired output and input impedance for specific applications. Block diagram of an *n*-stage cascaded amplifier is given in Fig. 6.20. The first stage is driven by a voltage source V_s (or a current source I_s). The output of first stage is fed to the input of the second stage, the output of second stage is supplied to the input of the third stage and so on. The output of the *n*th or last stage is fed to the load R_L. Actual voltage available at the input of the first stage is V_{in} (V_{in} equals signal voltage V_s if source resistance R_S is negligible) and the voltage available at the output terminals of the last stage is V_{out}. Then the ratio V_{out}/V_{in} gives the voltage gain of the *n*-stage cascaded amplifier. Thus by designing properly a cascaded amplifier, a weak input signal V_{in} of just a few microvolts can be amplified giving output voltage V_{out} of several volts.

Here we take up *R-C* coupled amplifier, being the most popular amplifier employed for audio-frequency amplification. A typical two-stage *R-C* coupled amplifier using *NPN* transistors in common emitter configuration (the most suitable transistor configuration for cascading) is given in Fig. 6.21. Detailed discussion of *R-C* coupled amplifier will be taken up in Section 6.15. Typical values of circuit components are given in the brackets.

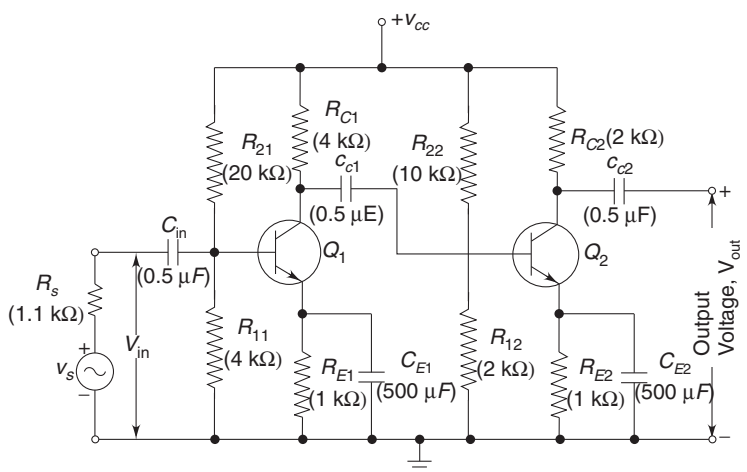


Fig. 6.21 *Two stage R-C coupled transistor amplifier*

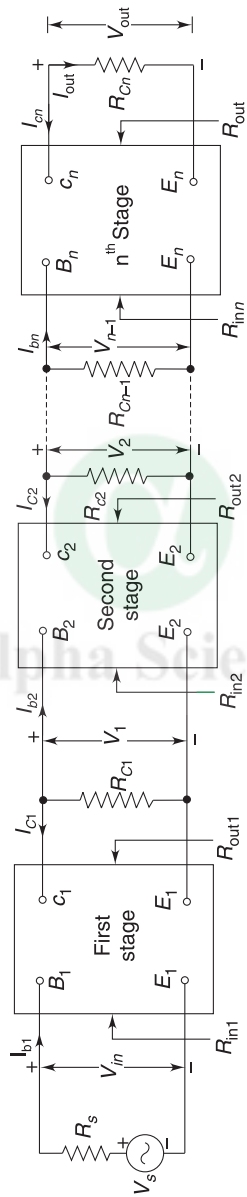


Fig. 6.22 *n-stage CE cascaded amplifier*

The block diagram of a two stage *R-C* coupled transistor amplifier shown in Fig. 6.21 indicating the various voltages currents and resistances involved is given in Fig. 6.22. The biasing arrangements and coupling capacitors have been omitted for sake of simplicity and clarity.

1. Voltage Gain

The overall voltage gain of the amplifier is given by the product of the voltage gains of the individual stages. This is proved as below:

Voltage gain of first stage,

$$A_V = \frac{V_1}{V_{\text{in}}} = \frac{\text{Output voltage of first stage}}{\text{Input voltage of first stage}}$$

$$= A_{v_1} \angle\theta_1 \tag{6.71}$$

where A_{v_1} is the magnitude of the voltage gain of the first stage and θ_1 is the phase angle between output and input voltages of this stage.

Similarly, the voltage gain of n^{th} stage,

$$A_{Vn} = \frac{\text{Output voltage of } n\text{th stage}}{\text{Input voltage of } n\text{th stage}}$$

$$= A_{v_n} \angle\theta_n \tag{6.72}$$

Thus the overall voltage gain of the complete *n*-stage cascaded amplifier is given as

$$A_V = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{\text{Output voltage of the } n^{\text{th}} \text{ stage}}{\text{Input voltage to the first stage}}$$

$$= A_v \angle\theta \tag{6.73}$$

where A_v is the magnitude of the voltage gain and θ is the phase angle between the output and input voltages of the amplifier.

Since $$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{V_1}{V_{\text{in}}} \times \frac{V_2}{V_1} \times \frac{V_3}{V_2} \times \ldots \times \frac{V_{\text{out}}}{V_{n-1}}$$

So $$A_V = A_{v_1} \times A_{v_2} \times A_{v_3} \times \ldots \times A_{Vn} \tag{6.74}$$

or $$A_V \angle\theta = A_{v_1} \,.\, A_{v_2} \,.\, A_{v_3} \,.\, \ldots \,.\, A_{V_n} \angle\theta_1 + \theta_2 + \theta_3 + \ldots + \theta_n \tag{6.75}$$

Hence, $$A_V = A_{v_1} \,.\, A_{v_2} \,.\, A_{v_3} \,.\, \ldots \,.\, A_{V_n} \tag{6.76}$$

and $$\theta = \theta_1 + \theta_2 + \theta_3 + \ldots + \theta_n \tag{6.77}$$

Thus it is concluded that the magnitude of the overall voltage gain equals the product of the magnitudes of the voltage gains of the individual stages and the resultant phase shift of the amplifier equals the sum of the phase shifts introduced by individual stages.

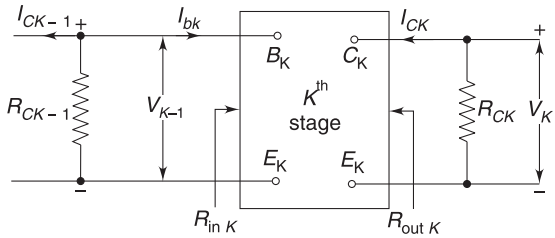


Fig. 6.23 *k^{th} stage of an n-stage CE Cascaded amplifier*

The voltage gain of k^{th} stage (an intermediate stage) of an n-stage CE cascaded amplifier shown in Fig. 6.23 is given as

$$A_{vk} = \frac{A_{i\,\text{k}}\, R_{L\,\text{k}}}{R_{in\,\text{k}}} \tag{6.78}$$

where R_{Lk} is the effective load at the collector of the kth stage.

The quantities is above equation (i.e. $A_{i\,\text{k}}$, $R_{L\,\text{K}}$ and $R_{\text{in k}}$) are evaluated by starting with the last stage and proceeding backward. Thus for the nth stage, current gain A_{in} and input resistance $R_{\text{in n}}$ are given as

$$A_{\text{in}} = \frac{-\,h_{fe}}{1 + h_{oe}\, R_{L\,\text{n}}} \tag{6.79}$$

and

$$R_{in\,\text{n}} = h_{ie} + h_{re}\, A_{in}\, R_{L\,\text{n}} \tag{6.80}$$

where $R_{L\,\text{n}}$ is the effective load impedance of the final (i.e., n^{th} stage and equals R_{cn}).

For the $(n-1)$th stage, effective load impedance $R_{in\,n-1}$ equals $R_{c\,n-1} \parallel R_{in\,n}$. So

$$R_{L\,n-1} = \frac{R_{cn-1} \cdot R_{in\,n}}{R_{cn-1} + R_{in\,n}} \tag{6.81}$$

Now the current gain $A_{i\,(n-1)}$, of the last but one stage [i.e. $(n-1)$th stage] is obtained using Eq. (6.79) by replacing $R_{L\,n}$ by $R_{L\,n-1}$. Similarly, the input impedance $R_{in\ n-1}$ of the $(n-1)$th stage is determined from Eq. (6.80) by replacing A_{in} by $R_{i\,n-1}$ and $R_{L\,n}$ by $R_{L\,n-1}$. Thus proceeding backward, base-to-collector current gains and input impedances of every stage including the first can be determined. Then the voltage gain of each stage can be determined by using Eq. (6.78).

Overall voltage gain of the n-stage cascaded amplifier may then be calculated from Eq. (6.74).

2. Current Gain

Overall voltage gain of an n-stage cascaded amplifier can be obtained, without determining the voltage gain of individual stage, from the following equation

$$A_v = A_i \frac{R_{cn}}{R_{in\ 1}} \tag{6.82}$$

where A_i is the overall current gain of the n-stage cascaded amplifier and

$$A_i = \frac{\text{Output current of the } n^{\text{th}} \text{ stage}}{\text{Input current (or base current of first stage)}}$$

$$= \frac{I_{\text{out}}}{I_{\text{in}}} = \frac{-I_{cn}}{I_{bi}} = -\frac{I_n}{I_{bi}} \tag{6.83}$$

where $I_{cn} \equiv I_n$, the collector current of the n^{th} stage.

Now let us obtain equations so as to calculate current gain A_i in terms of circuit parameters.

Since $$\frac{I_n}{I_{bi}} = \frac{I_1}{I_{bi}} \cdot \frac{I_2}{I_1} \cdot \ldots \cdot \frac{I_{n-1}}{I_{n-2}} \cdot \frac{I_n}{I_{n-1}}$$

then $$A_i = A_{i_1} \cdot A'_{i_2} \cdot A'_{i_3} \cdot \ldots A'_{i_{n-1}} \cdot A_{in}{}' \tag{6.84}$$

where A_{i_1} is the base-to-collector current gain of the first stage and is equal to $\frac{-I_{ci}}{I_{bi}}$ while A'_{i_2}, A'_{i_3} etc., are the collector-to-collector current gains of second, third etc. stages.

For k^{th} stage, collector-to-collector current gain A'_{ik} is given as

$$A'_{ik} = \frac{I_{ck}}{I_{ck-1}} \tag{6.85}$$

Similarly, base-to-collector current gain of k^{th} stage A_{ik} is given as

$$A_{ik} = \frac{-I_{ck}}{I_{bk}} \tag{6.85}$$

Now from Fig. 6.23,

$$I_{bk} = -I_{c\,k-1} \frac{R_{ck-1}}{R_{ck-1} + R_{in\,k}} \tag{6.87}$$

where $R_{in\,k}$ is the input impedance of the k^{th} stage.

Hence $$A'_{ik} = \frac{I_{ck}}{I_{ck-1}} = \frac{I_{ck}}{I_{bk}} \cdot \frac{I_{bk}}{I_{ck-1}}$$

$$= A_{ik} \cdot \frac{R_{ck-1}}{R_{ck-1} + R_{in\,k}} \tag{6.88}$$

Base-to-collector current gain A_{ik} is determined by starting with the output stage and proceeding backward to the k^{th} stage, as indicated in connection with Eqs. (6.79), (6.80) and (6.81). The collector-to-collector current gains are then determined by using Eq. (6.88) and the overall current gain of the n-stage amplifier is determined by using Eq. (6.84).

3. Input Impedance

The input impedance of the complete cascaded amplifier is determined, as discussed above, the starting with the last stage and proceeding backward.

4. Output Impedance

The output impedance of each transistor stage and of the overall amplifier is determined starting with the first stage and using equation

$$Y_{\text{out 1}} = h_{oe} - \frac{h_{fe}\, h_{re}}{h_{ie} + R_S} \tag{6.89}$$

Then $\frac{1}{Y_{\text{out 1}}}$ gives the corresponding output impedance $R_{\text{out 1}}$.

The output impedance $R'_{\text{out } k}$ of the k^{th} stage is the parallel combination of the output impedance $R_{\text{out } k}$ of k^{th} stage and R_{ck}. The effective source impedance of the $(k + 1)^{\text{th}}$ stage is also $R'_{\text{out } k}$.

5. Power Gain

The overall power gain of the n-stage cascaded amplifier is given as

$$A_p = \frac{\text{Output power}}{\text{Input power}} = \frac{V_{\text{out}}\, I_n}{V_{\text{in}}\, I_{bi}} = A_v\, A_i \tag{6.90}$$

or

$$A_p = (A_i)^2 \frac{R_{cn}}{R_{\text{in1}}} \tag{6.91}$$

6.14.3 Frequency Response of Coupled Amplifiers

The curve drawn between the voltage gain and signal frequency of an amplifier is known as the *frequency response*. The performance of an amplifier is judged to a considerable extent by its frequency response. In the design of an amplifier, appropriate steps are taken to ensure that gain is essentially uniform over some specified range.

Video amplifiers are almost invariably of the *R-C* coupled type. For such a stage the frequency characteristics may be divided into three regions: Midband-frequency region, low-frequency region and high-frequencies region.

In midband-frequency region, the amplification remains reasonably constant and equal to A_{vm}. For the present discussion midband gain may be assumed to be normalized to unity, i.e., $A_{vm} = 1$.

Low-Frequency Response In the low-frequency region, below the midband, an amplifier stage behaves like the simple high-pass circuit (Fig. 6.24) of time constant $\tau_1 = R_1\, C_1$.

The current through the circuit is given by

$$I = \frac{V_{in}}{R_1 - jX_1}$$

and output voltage,

$$\begin{aligned} V_{out} &= \text{Voltage drop across } R_1 \\ &= IR_1 \\ &= \frac{R_1}{R_1 - jX_1} V_{in} \\ &= \frac{R_1}{R_1 - \dfrac{j}{2\pi f C_1}} V_{in} \\ &= \frac{V_{in}}{1 - \dfrac{j}{2\pi f C_1 R_1}} \end{aligned} \tag{6.92}$$

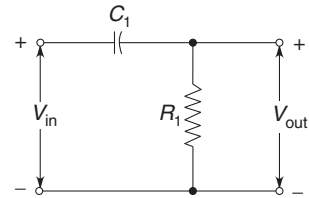


Fig. 6.24 *A high-pass R-C circuit for determination of low-frequency response of an amplifier*

The voltage gain at low frequencies A_{vl}, is defined as the ratio of the output voltage V_{out}, to the input voltage V_{in}.

i.e., $$A_{vl} = \frac{V_{out}}{V_{in}} = \frac{1}{1 - \dfrac{j}{2\pi f C_1 R_1}} = \frac{1}{1 - \dfrac{jf_1}{f}} \tag{6.93}$$

where f_1 is the cut-off frequency and $= \dfrac{1}{2\pi R_1 C_1}$ (6.94)

Now magnitude of A_{vl} and phase angle θ are given by

$$|A_{vl}| = \frac{1}{\sqrt{1 - (f_1/f)^2}} \tag{6.95}$$

and phase angle $$\theta_1 = \tan^{-1} \frac{f_1}{f} \tag{6.96}$$

At the frequency $f = f_1$, $A_{vl} = \dfrac{1}{\sqrt{2}} = 0.707$ whereas in the midband region ($f >> f_1$), $A_{vl} \to 1$. Hence f_1 is the frequency at which the gain has fallen to 0.707 times its midband value A_{vm}. This drop in signal level (assuming equal

input and output impedances) corresponds to a decibel reduction* of 20 log $\frac{1}{\sqrt{2}}$ or – 3 dB. Accordingly, f_1 is referred to as the *lower 3 dB frequency*. From Eq. (6.94) it is observed that f_1 is that frequency for which the resistance R_1 equals the capacitive reactance i.e.,

$$R_1 = \frac{1}{2\pi f_1 C_1}$$

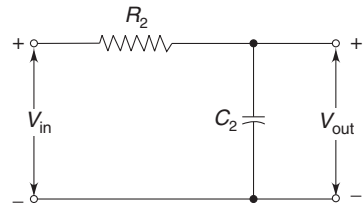


Fig. 6.25 *A low-pass R-C circuit for determination of high-frequency response of an amplifier*

High-Frequency Response In the high-frequency region, the stage behaves like a simple low-pass circuit shown in Fig. 6.25.

Proceeding as above, we obtain for the magnitude $|A_{vh}|$ and phase angle θ_2 of the gain.

$$|A_{vh}| = \frac{1}{\sqrt{1 + (f/f_2)^2}} \tag{6.97}$$

$$\theta_2 = \tan^{-1} \frac{f}{f_2} \tag{6.98}$$

where

$$f_2 = \frac{1}{2\pi R_2 C_2} \tag{6.99}$$

Since at $f = f_2$ the gain is reduced to $\frac{1}{\sqrt{2}}$ times its midband value, then f_2 is called the *upper 3 dB frequency*. It also represents that frequency for which the resistance R_2 is equal to the capacitive reactance $\frac{1}{2\pi f_2 C_2}$.

In the above expression θ_1 and θ_2 represent the angle by which the output lags the input, neglecting the initial 180° phase shift through the amplifier. The frequency dependence of the gains in the high- and low-frequency range is to be seen in Fig. 6.26.

Bandwidth The frequency range from f_1 to f_2 is called the *bandwidth* (BW) of the amplifier stage.

---

* Voltage gain in dB $= 10 \log_{10} \left[\frac{V_{out}^2/R_L}{V_{in}^2/R_{in}} \right]$

$= 10 \log_{10} \left[\left(\frac{V_{out}}{V_{in}} \right)^2 \times \frac{R_{in}}{R_L} \right]$

$= 20 \log \frac{V_{out}}{V_{in}}$ (for $R_L = R_{in}$)

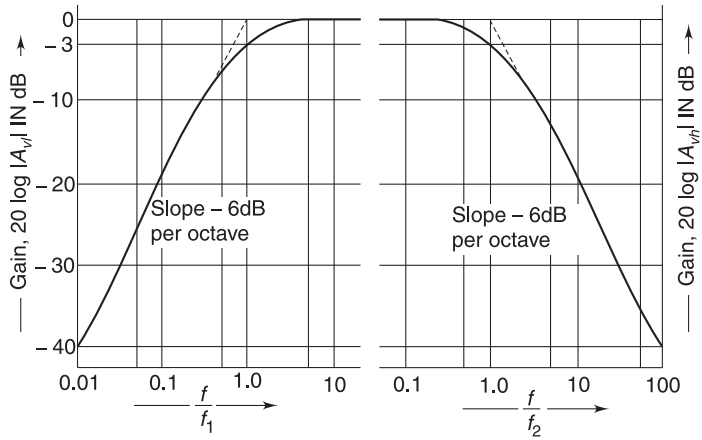


Fig. 6.26 *A log-log plot of the amplitude frequency-response characteristic of an R-C coupled amplifier*

6.15 R-C (RESISTANCE-CAPACITANCE) COUPLED TRANSISTOR AMPLIFIERS

A two-stage *R-C* coupled amplifier using *NPN* transistor in CE configuration is shown in Fig. 6.27. The two transistors used are identical and use a common power supply V_{CC}. The resistors R_1, R_2 and R_E form the biasing and stabilization network. In this arrangement, the signal developed across collector resistor R_C of the first stage is coupled to the base of the second stage through the coupling capacitor C_C. As the coupling from one stage to the next is obtained by a coupling capacitor followed by a connection to a shunt resistor, therefore, such amplifiers

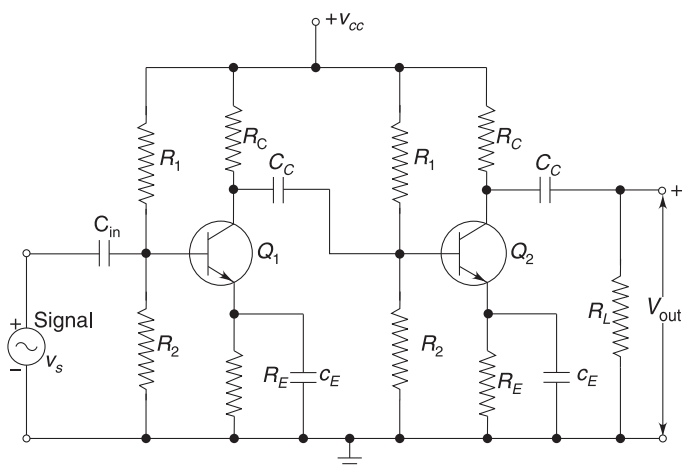


Fig. 6.27 *Two-stage R-C coupled transistor amplifier*

are called *resistance-capacitance coupled* or *R-C coupled amplifiers*. The input capacitor C_{in} couples ac signal voltage to the base of transistor Q_1. In the absence of C_{in} the signal source will be in parallel with resistor R_2 and the bias voltage of the base will be affected. Thus the function of C_{in} is to allow only the alternating current from signal source to flow into the input circuit.

The emitter bypass capacitor C_E, offers low reactance path to the signal. If it is not present, then the voltage drop across R_E will reduce the effective voltage available across the base-emitter terminals (the input voltage) and thus reduces the gain.

The coupling capacitor C_C transmits a.c. signal but blocks the d.c. voltage of the first stage from reaching the base of the second stage. Thus the d.c. biasing of the next stage is not interfered with. For this reason, the coupling capacitor C_C is also called the *blocking capacitor*.

6.15.1 Operation

When a.c. signal is applied to the base of the first amplifier, it appears in the amplified form across collector load R_C. The amplified signal developed across R_C is transmitted to the base of nest stage of the amplifier through coupling capacitor C_C. This is further amplified by the next stage and so on. Thus, the cascade stages amplify the signal and thus the overall gain is considerably increased. The phase of output is the same as that of input because the phase is reversed twice by two transistors as they are in CE configuration.

It may be mentioned here that the overall gain is less than the product of the gains of individual stages. This is because when a second stage is made to follow the first one, the *effective load resistance* of the first stage is reduced because of the shunting effect of the input resistance of the second stage. This reduces the gain of the stage which is loaded by the next stage. For example, in a four-stage amplifier the gains of first three stages will be reduced due to loading effect of the subsequent stage. However, the gain of final or last stage (4^{th} stage in this case) which has no loading effect of subsequent stage, remains unaffected. Thus the overall gain shall be less than the product of the gain of four stages.

6.15.2 Analysis

For drawing approximate model of the circuit shown in Fig. 6.27 following assumptions are made.

1. h_{re} is so small that the voltage across source $h_{re}\ V_{out}$ can be neglected.
2. $1/h_{oe}$ is so large that it can be considered as an open-circuit.
3. The bias resistors R_1 and R_2 are very large in comparison to h_{ie}.
4. The reactance of emitter bypass capacitor C_E for any given input frequency is so small that parallel combination of R_E and C_E can be considered as a short-circuit.

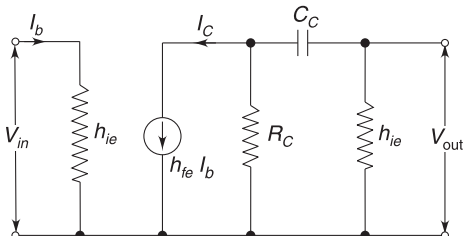


Fig. 6.28 *Approximate model of an R-C coupled transistor amplifier*

The approximate model drawn with the above assumptions is shown in Fig. 6.28.

For the purpose of analysis the entire frequency range may be divided into the following three categories.

(a) Mid Frequency Range At mid frequencies, the impedance offered by the coupling capacitor C_C is so small that it can be effectively considered as a short-circuit and its effect can be neglected. Equivalent circuit for mid frequency range is given in Fig. 6.29 and its Thevenin's equivalent circuit is given in Fig. 6.30.

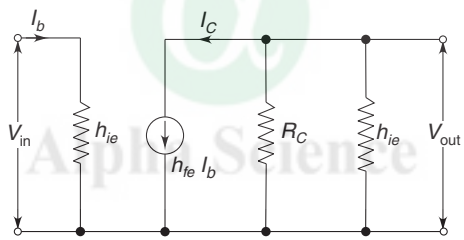


Fig. 6.29 *Equivalent circuit for mid frequency range*

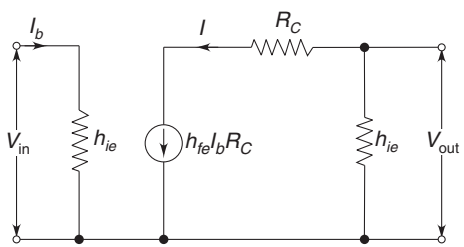


Fig. 6.30 *Thevenin's equivalent circuit*

From circuit shown in Fig. 6.30.

Current, $$I = \frac{h_{fe}\ I_b\ R_C}{R_C + h_{ie}} \qquad (6.100)$$

So current gain,

$$A_{im} = \frac{1}{I_b} = \frac{h_{fe}\, R_C}{R_C + h_{ie}} \tag{6.101}$$

Output voltage,

$$V_{\text{out}} = h_{ie}\, I = \frac{h_{ie}\, h_{fe}\, I_b\, R_C}{R_C + h_{ie}} \tag{6.102}$$

Input voltage,

$$V_{\text{in}} = h_{ie}\, I_b$$

So voltage gain

$$A_{v\,m} = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{\dfrac{h_{ie}\, h_{fe}\, I_b\, R_C}{(R_C + h_{ie})}}{h_{ie}\, I_b}$$

$$= \frac{h_{fe}\, R_C}{R_C + h_{ie}} \tag{6.103}$$

From Eqs. (6.101) and (6.103) it is obvious that current and voltage gains are equal.

(b) Low Frequency Range In low frequency range, the impedance offered by coupling capacitor C_C is comparable to the collector resistance R_C. It largely affects the current amplification. Therefore, it becomes necessary to include it in the equivalent circui, as shown in Fig. 6.28.

Thevenin' equivalent circuit is given in Fig. 6.31. From circuit shown in Fig. 6.31

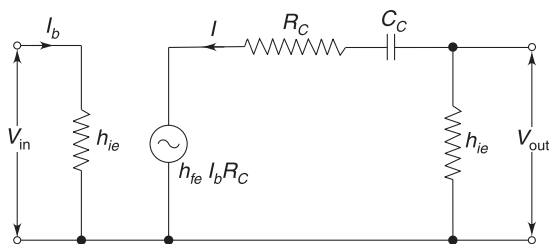


Fig. 6.31 *Thevenin's equivalent circuit for low frequency range*

Current,
$$I = \frac{h_{fe}\, I_b\, R_C}{h_{ie} + R_C - j/\omega C_C} \tag{6.104}$$

So current gain,

$$A_{i\,l} = \frac{1}{I_b} = \frac{h_{fe}\, R_C}{h_{ie} + R_C - j/\omega C_C} \tag{6.105}$$

Output voltage,

$$V_{\text{out}} = h_{ie}\, I = \frac{h_{ie}\, h_{fe}\, I_b\, R_C}{h_{ie} + R_C - j/\omega C_C} \tag{6.106}$$

Input voltage,

$$V_{\text{in}} = h_{ie}\, I_b$$

So voltage gain

$$A_{v\,m} = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{\dfrac{h_{ie}\, h_{fe}\, I_b\, R_C}{h_{ie} + R_C - j/\omega C_C}}{h_{ie}\, I_b}$$

$$= \frac{h_{fe}\, R_C}{h_{ie} + R_C - j/\omega C_C} \tag{6.107}$$

The magnitude of A_{vl} is given by

$$|A_{vl}| = \frac{h_{fe}\, R_C}{\sqrt{(h_{fe} + R_C)^2 + \left(\dfrac{1}{2\pi f C_C}\right)^2}} \tag{6.108}$$

From Eq. (6.108), it is obvious that in the low frequency range voltage gain decreases with the decrease in frequency.

(c) High Frequency Range In high frequency range, the reactance offered by coupling capacitor C_C is very small and it may be considered as a short-circuit.

The important factor that comes into picture at high frequencies is the interelectrode capacitances. These capacitances are due to formation of depletion layers at the junctions. The interelectrode capacitances are shown by dotted lines in Fig. 6.32. The capacitance C_{bc} between the base and collector connects the output with the input. Because of this, negative feedback takes place in the circuit and gain is reduced. The feedback effect increases with the increase in frequency because with the increase in frequency reactive impedance of the capacitor decreases. The capacitance C_{be}, between the base and emitter, offers a low impedance path at the input side at high frequencies. This reduces the input impedance of the device and consequently effective input signal is reduced. Thus gain falls. Similarly, the capacitance C_{ce} produces a shunting effect at high frequencies on the output side. It is noteworthy point that C_{bc} is the most important capacitance because feedback takes place from output circuit to input circuit through this capacitance. This is known as *Miller effect*. Besides these capacitances, there are wiring capacitances C_{W_1} and C_{W_2}, as shown in Fig. 6.32. These are the capacitances between the connecting wires of the circuit and ground.

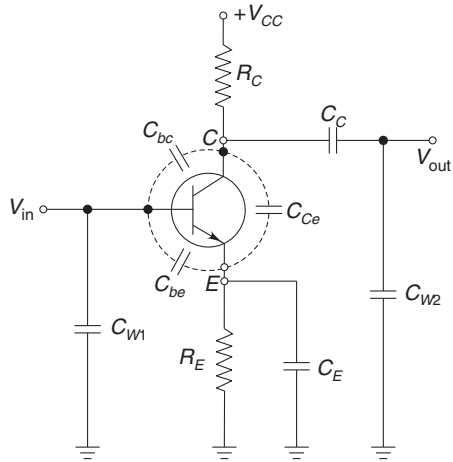


Fig. 6.32 *R-C coupled amplifier at high frequencies*

It can be shown that C_{be} and C_{be} may be replaced with a single capacitor C_d across the input resistance h_{ie} of the transistor. The value of shunt capacitance C_d in the input circuit of the first stage is small because it depends on the output impedance of the first transistor, which is small. But in the output circuit of the first stage C_d is increased by stray capacitance of the wiring. The reactance $\frac{1}{\omega C_d}$ will have appreciable shunting effect on R_2 and h_{ie}. The equivalent circuits are given in Figs. 6.33 and 6.34.

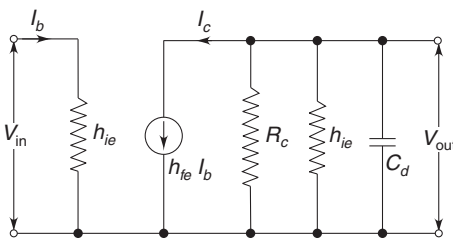


Fig. 6.33 *Equivalent circuit for high frequency range*

From Thevenin's equivalent circuit shown in Fig. 6.34.

$$\text{Current,} \qquad I = \frac{h_{fe}\, I_b \dfrac{R_C\, h_{ie}}{R_C + h_{ie}}}{\dfrac{R_C\, h_{ie}}{R_C + h_{ie}} + \dfrac{1}{j\omega C_d}} + \frac{h_{fe}\, I_b\, R_C\, h_{ie}}{R_C\, h_{ie} + \dfrac{1}{j\omega C_d}(R_C + h_{ie})} \qquad (6.109)$$

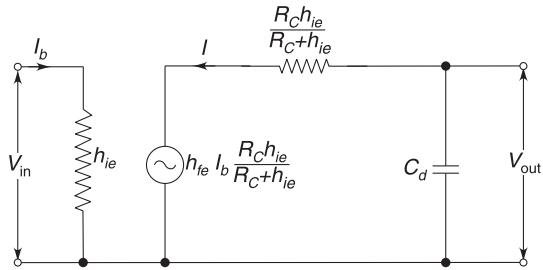


Fig. 6.34 *Thevenin's equivalent circuit*

So current gain,

$$A_{ih} = \frac{I}{I_b} = \frac{h_{fe}\, h_{ie}\, R_C}{h_{ie}\, R_C + \dfrac{1}{j\omega C_d}(h_{ie} + R_C)} \tag{6.110}$$

Output voltage,

$$V_{\text{out}} = 1 \times \frac{1}{j\omega C_d}$$

$$= \frac{\dfrac{h_{ie}\, h_{fe}\, I_b\, R_C}{(h_{ie} + R_C)}}{\dfrac{h_{ie}\, R_C}{h_{ie} + R_C} + \dfrac{1}{j\omega C_d}} \times \frac{1}{j\omega C_d}$$

$$= \frac{h_{ie}\, h_{fe}\, I_b\, R_C}{j\omega C_d\, h_{ie}\, R_C + h_{ie} + R_C}$$

Input voltage,

$$V_{\text{in}} = h_{ie}\, I_b$$

$\therefore$ Voltage gain,

$$A_{vh} = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{h_{fe}\, R_C}{j\omega C_d\, h_{ie}\, R_C + h_{ie} + R_C} \tag{6.111}$$

The magnitude of A_{vh} is given by

$$|A_{vh}| = \frac{h_{fe}\, R_C}{\sqrt{(h_{ie} + R_C)^2 + (2\pi f C_d\, h_{ie}\, R_C)^2}} \tag{6.112}$$

From Eq. (6.112), it is obvious that with the increase in input frequency the magnitude of voltage gain falls off.

6.15.3 Lower Cut-off Frequency

The lower cut-off frequency, f_1 may be defined as the frequency at which the magnitude of the voltage gain in the low frequency range fall off to $\frac{1}{\sqrt{2}}$ or 0.707 times the maximum gain in the mid frequency range. Thus, at $f = f_1$

$$|A_{vl}| = \frac{1}{\sqrt{2}} |A_{vm}|$$

From Eq. (6.108), the magnitude of A_{vl},

$$|A_{vl}| = \frac{h_{fe} R_C}{\sqrt{(h_{ie} + R_C)^2 + \left(\frac{1}{2\pi f C_C}\right)^2}}$$

$$= \frac{h_{fe} R_C}{(h_{ie} + R_C)\left[1 + \left\{\frac{1}{2\pi f C_C (h_{ie} + R_C)}\right\}^2\right]^{1/2}} \quad (6.113)$$

and from Eq. (6.103), the magnitude of $A_{v\,m}$,

$$|A_{vm}| = \frac{h_{fe} R_C}{h_{ie} + R_C} \quad (6.114)$$

From Eqs. (6.113) and (6.114)

$$\frac{A_{vl}}{A_{vm}} = \frac{1}{\sqrt{1 + \left\{\frac{1}{2\pi f C_C (h_{ie} + R_C)}\right\}^2}} \quad (6.115)$$

If f_1 is the lower cut-off frequency, then

$$\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{1 + \left\{\frac{1}{2\pi f_1 C_C (h_{ie} + R_C)}\right\}^2}}$$

or $$1 + \left[\frac{1}{2\pi f_1 C_C (h_{ie} + R_C)}\right]^2 = 2$$

or $$f_1 = \frac{1}{2\pi C_C (h_{ie} + R_C)} \quad (6.116)$$

From Eqs. (6.115) and (6.116), we have

$$|A_{vl}| = \frac{|A_{vm}|}{\sqrt{1 + \left(\frac{f_1}{f}\right)^2}} = \frac{|A_{vm}|}{\sqrt{1 + (\tan \theta_l)^2}} \quad (6.117)$$

where $$\tan \theta_l = \frac{f_1}{f}$$

or $$\theta_l = \tan^{-1} \frac{f_1}{f} \tag{6.118}$$

So the total phase shift of the voltage gain A_{vl} is the low frequency range is given by

$$\theta_l = 180° + \theta_l = 180° + \tan^{-1} \frac{f_1}{f} \tag{6.119}$$

From above equation it is obvious that with the decrease in frequency, θ_l increases and at $f = f_1$

$$\theta_l = 180° + 45° = 225°$$

6.15.4 Upper Cut-off Frequency

The upper cut-off frequency f_2 may be defined as the frequency at which the magnitude of the voltage gain in the high frequency range falls off to $\frac{1}{\sqrt{2}}$ or 0.707 times the magnitude of the gain in the mid frequency range. Thus at $f = f_2$.

$$|A_{vh}| = \frac{1}{\sqrt{2}} |A_{vm}|$$

From Eq. (6.112), the magnitude of A_{vh},

$$|A_{vh}| = \frac{h_{fe} R_C}{(h_{ie} + R_C) \sqrt{1 + \left\{ \frac{2\pi f C_d h_{ie} R_C}{h_{ie} + R_C} \right\}^2}} \tag{6.120}$$

From Eqs. (6.114) and (6.120), we have

$$\frac{A_{vh}}{A_{vm}} = \frac{1}{\sqrt{1 + \left\{ \frac{2\pi f C_d h_{ie} R_C}{h_{ie} + R_C} \right\}^2}} \tag{6.121}$$

If f_2 is the upper cut-off frequency then

$$\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{1 + \left\{ \frac{2\pi f C_d h_{ie} R_C}{h_{ie} + R_C} \right\}^2}}$$

or $$\frac{2\pi f C_d h_{ie} R_C}{h_{ie} + R_C} = 1$$

or upper cut-off frequency,

$$f_2 = \frac{h_{ie} + R_C}{2\pi C_d\, h_{ie}\, R_C} = \frac{1}{2\pi C_d}\left\{\frac{1}{h_{ie}} + \frac{1}{R_C}\right\} \tag{6.122}$$

From Eq. (6.121)

$$\frac{|A_{vh}|}{|A_{vm}|} = \frac{1}{\sqrt{1 + \left(\frac{f}{f_2}\right)^2}} = \frac{1}{\sqrt{1 + (\tan\theta_h)^2}} \tag{6.123}$$

where $$\theta_h = \tan^{-1}\frac{f}{f_2} \tag{6.124}$$

So $$\phi_h = 180° - \theta_h = 180° - \tan^{-1}\frac{f}{f_2} \tag{6.125}$$

At $f = f_2$, $\phi_h = 180° - 45° = 135°$

ϕ_h decreases with the increase in input frequency f. The variation of total phase shift between input and output voltages with frequency is shown in Fig. 6.35.

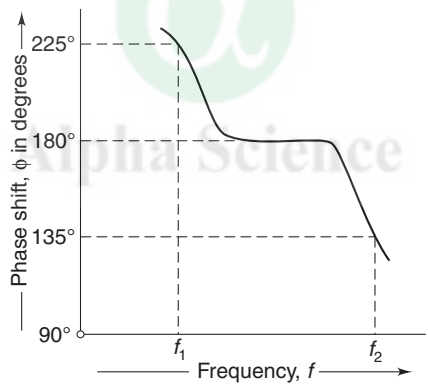


Fig. 6.35 *Phase relationship in an R-C coupled amplifier*

6.15.5 Frequency Response Curve

The frequency purpose curve of a typical *R-C* coupled transistor amplifier is shown in Fig. 6.36.

In *mid frequency range* (50 Hz – 20 kHz), the voltage gain of the amplifier is constant, as is obvious from the analysis. With the increase in the frequency in this range, the reactance of the coupling capacitor C_C reduces thereby increasing the gain but at the same time lower capacitive reactance causes higher loading resulting in lower voltage gain. Thus, the two effects cancel each other and uniform gain is obtained in mid frequency range.

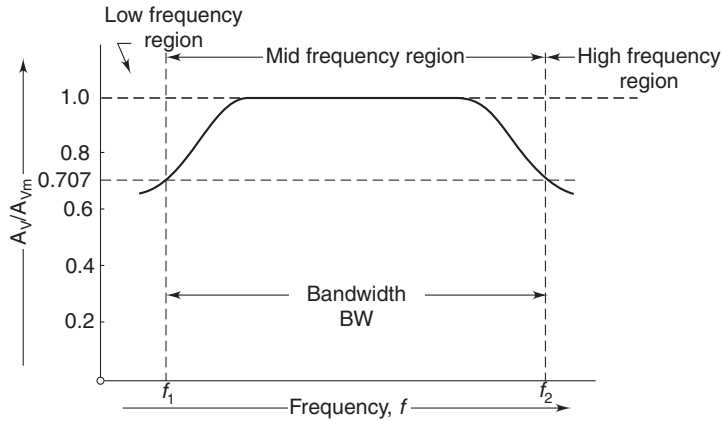


Fig. 6.36 *Frequency response curve of an R-C coupled amplifier*

At *low frequencies* (below 50 Hz) higher capacitive reactance of coupling capacitor C_C allows very small part of signal to pass from one stage to the next and also because of higher reactance of emitter bypass capacitor C_E, the emitter resistor R_E is not effectively shunted. Thus, the voltage gain falls off at low frequencies.

At *high frequencies* (exceeding 20 kHz), the gain of the amplifier decreases with the increase in frequency. Several factors are responsible for this reduction in gain. At high frequencies, the reactance of coupling capacitor C_C becomes very small and C_C behaves as a short circuit. This increases the loading of next stage and reduces the voltage gain. At high frequencies, capacitive reactance of base-emitter junction is low and so the base current is increased and current gain factor β is reduced. At high frequencies, the interelectrode capacitance C_C connects the output circuit to the input circuit. Thus, negative feedback takes place and the gain is reduced. Bedsides interelectrode capacitances, there are wiring capacitances C_{W_1}, and C_{W_2} and input capacitance C_{in} of the next stage. The effect of the capacitances C_{W1}, C_{W_2} and input capacitance C_{in} of the next stage can be represented by a single shunt capacitance given as

$$C_s = C_{W_1} + C_{W_2} + C_{\text{in}} \qquad (6.126)$$

The output section of the amplifier is shown in Fig. 6.37 from the a.c. point of view, for high frequency considerations. The capacitance C_s is the equivalent shunt capacitance as given by Eq. (6.126). In figure coupling and bypass capacitors C_C and C_E are not shown as they effectively represent short-circuits at high frequencies.

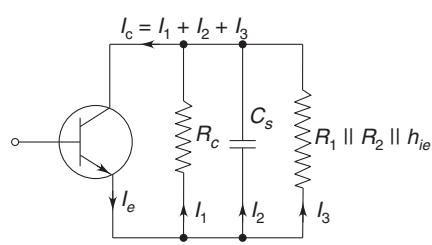


Fig. 6.37 *Output section of an R-C coupled amplifier at high frequencies*

From circuit shown in Fig. 6.37, it is obvious that collector current I_c is made up of three currents I_1, I_2 and I_3. With the increase in frequency of the input signal, the reactive impedance of the shunt capacitance C_s decreases, being inversely proportional to frequency, and as a result the current I_2 through this capacitance C_s increases. This reduces both currents I_1 and I_3 since total current $I_c = (I_1 + I_2 + I_3)$ is almost constant. The output voltage V_{out} being the product of I_3 and R (i.e., I_3R) decreases. Thus higher the frequency, the lower is the impedance offered by C_s and lower will be the output voltage V_{out}.

6.15.6 Bandwidth

Frequency response curve of an *R-C* coupled amplifier is shown in Fig. 6.36. The cut-off frequencies f_1 and f_2 are marked in the figure. The difference of the two frequencies (upper cut-off frequency f_2 and lower cutoff frequency f_1) is called the *bandwidth* (BW)

$$\therefore \qquad \text{BW} = f_2 - f_1 \tag{6.127}$$

when the bandwidth (BW) is multiplied by the gain at mid frequencies, gain-bandwidth (GBW) is obtained.

6.15.7 Advantages

Excellent frequency response (constant gain over the audio-frequency range, the most important region for speech, music etc.), cheaper in cost (resistors and capacitors used are very cheap) and very compact circuit (resistors and capacitors used being small and extremely light).

6.16 COMPARISON OF DIFFERENT TYPES OF COUPLINGS

S. No.	Type of Coupling / Particulars	R-C coupling	Transformer coupling	Direct coupling
1.	Size and weight	Small	Large and bulky	Very small
2.	Cost	Small	Costlier	Very small
3.	Frequency response	Excellent in the audio-frequency range	Poor	Best
4.	Impedance matching	Not good	Excellent	Good
5.	Uses	For voltage amplification	For power amplification	For amplification of extremely low frequency signals

6.17 VARIATIONS OF HYBRID PARAMETERS OF A TRANSISTOR

The variation of *h*-parameters depends upon junction temperature, collector current and collector-to-emitter voltage V_{CE}. Among these factors, the variations due to junction temperature and collector current are significant and thus discussed here.

6.17.1 Variation of *h* Parameters due to Temperature Variation

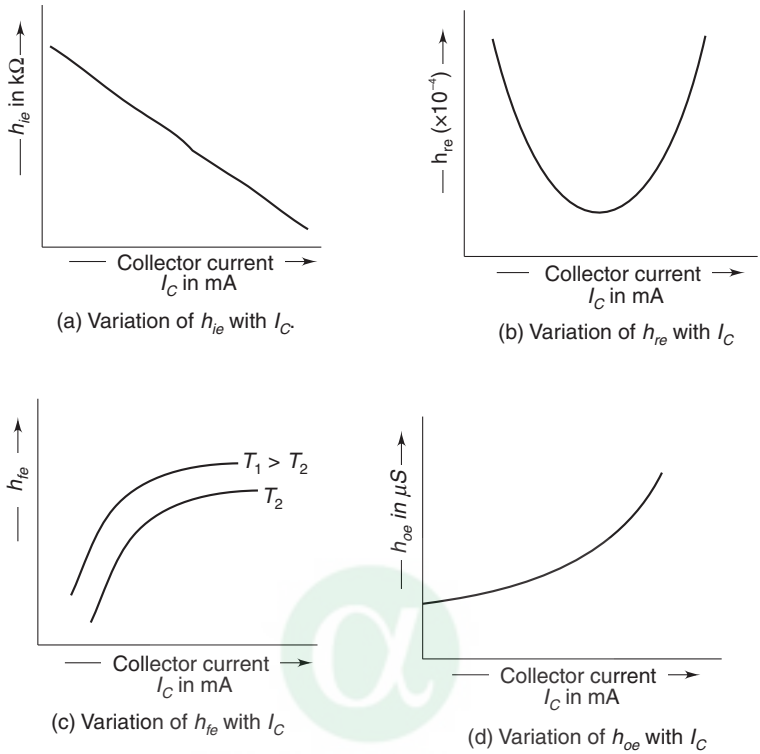
From the above discussion it is obvious that all the four *h* parameters of any transistor configuration (*viz.*, CB, CE and CC) can be determined from the slopes and spacing between curves at the quiescent point *Q* if it is specified. Since the characteristic curves, in general, are neither straight lines nor equally spaced for equal variations in collector-emitter voltage or base current, the values of *h*-parameters depend upon the position of the *Q*-point on the curves. It is also known that the shape and actual numerical values of the characteristic curves depend on the junction temperature. Most transistor specification sheets include curves of the variation of the *h* parameters with the quiescent point and temperature. Usually I_C = 1 mA is taken as the reference collector current and collector junction temperature of 25°C is taken as the reference temperature.

6.17.2 Variations of *h*-parameters due to Variation in Collector Current

(a) The parameter h_{ie} varies with operating value of collector current I_C as shown in Fig. 6.38(a), i.e., h_{ie} decreases with the increase in the operating value of I_C.

(b) The parameter h_{re} varies with operating value of I_C, as depicted in Fig. 6.38(b). From the curve shown in Fig. 6.38(b), it is obvious that h_{re} first decreases with the increase in the operating value of I_C, attains a minimum value at a certain value of I_C and thereafter increases with the increase in I_C.

(c) The parameter h_{fe} varies with I_C as shown in Fig. 6.38(c). The graph shown in Fig. 6.38(c) indicates that h_{fe} first increases with I_C, attains a maximum value at a particular value of I_C and then decreases slightly with I_C. The value of h_{fe} is temperature dependent and increases with the rise in temperature.



(a) Variation of h_{ie} with I_C.

(b) Variation of h_{re} with I_C

(c) Variation of h_{fe} with I_C

(d) Variation of h_{oe} with I_C

Fig. 6.38

Questions

6.1 Define *h*-parameters for a BJT transistor.

6.2 Draw the hybrid model of a BJT. Define

(i) h_{fe}

(ii) h_{ie}

(iii) h_{re}

(iv) h_{oe}

6.3 Give a mathematical analysis of Common-emitter amplifier using hybrid parameters and derive expressions for (i) current gain (ii) input resistance (iii) voltage gain (iv) power gain and (v) output resistance.

6.4 Explain graphical method for determination of *h*-parameters from the transistor characteristics.

6.5 Describe experimental method for determination of *h*-parameters.

6.6 What are multistage amplifiers? Derive the equations for overall voltage gain of a multistage amplifier in terms of individual voltage gains.

6.7 Classify the amplifiers based on different using conditions.

6.8 Explain the working of *R-C* coupled amplifier with the help of neat diagram. Explain the effect of coupling capacitor or low frequency response of *R-C* coupled amplifier.

6.9 Discuss the frequency response characteristic of *R-C* coupled amplifiers. Derive the general expression for gain at low and high frequencies.

6.10 Write short notes on:

(i) Two stage *R-C* coupled amplifiers

(ii) Low frequency response of *R-C* coupled amplifiers

(iii) High frequency response of *R-C* coupled amplifiers

6.11 Discuss the variations of hybrid parameters of a transistor.

7 Feedback in Amplifiers

7.1 INTRODUCTION

A feedback circuit is one where part of the output signal is added to the input either in phase or in antiphase. A general block diagram of a feedback circuit is shown in Fig. 7.1.

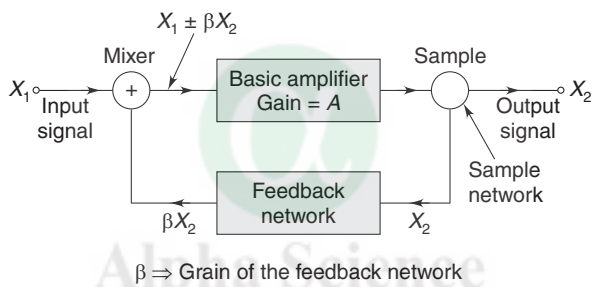


Fig. 7.1 *A basic feedback system*

The feedback system as shown in Fig. 7.1 has the following elements or parts:

1. The input X_1 is either a signal voltage v_s in series with a resistor R_S (a Thevenin's representation) or a signal current i_S in parallel with a resistor R_S (a Norton's representation).
2. The feedback network β is usually a two port network which may contain resistors, capacitors, and inductors. Most often it is simply a resistive configuration.
3. The output X_2 is either a voltage or a current. The sampling network, therefore, samples either voltage or current at the output of the basic amplifier. If X_2 is a voltage then it is sampled by connecting the feedback network in shunt across the output. This is known as voltage or node sampling. However, if X_2 is a current then the feedback network is placed in series with the output. This is known as the current or loop sampling. Some other sampling networks are also possible.

4. The mixer combines the output of the feedback network with the input signal either in series Z (loop) or in shunt (node).

Note that in Fig. 7.1, the output after being sampled is added to the input either to reduce the effective input signal or to increase the effective input signal to the amplifier. The former case is referred to as the inverse, degenerative, or negative feedback while the latter is called the regenerative or positive feedback.

7.2 FEEDBACK EQUATION

The input signal X_1 and the feedback signal X_f (= βX_2) are mixed or combined either in antiphase or in phase to form the single signal ($X_1 \mp \beta X_2$) which is then amplified by the amplifier section of the circuit. The amplifier output then goes into a sample circuit which feeds part of the output signal to the load and part of the signal to the feedback network. Note that A and β can be simply treated as 'gains'.

The symbol A in Fig. 7.1 represents the ratio of the output signal to the input signal of the basic amplifier. The ratio A is usually referred to as the forward or open-loop transfer function of the amplifier. The basic amplifier as depicted in Fig. 7.1 could be either a voltage amplifier or a current amplifier, or a transconductance amplifier, or a transcresistance amplifier. The transfer functions for these four different types of amplifiers are defined as follows:

For voltage amplifier $\Rightarrow A_V = \dfrac{v_o}{v_s}$

For current amplifier $\Rightarrow A_I = \dfrac{i_o}{i_i}$

For transconductance amplifier $\Rightarrow G_M = \dfrac{i_o}{v_s}$

For transresistance amplifier $\Rightarrow R_M = \dfrac{v_o}{i_i}$

Although G_M and R_M are defined as the ratio of two signals, one of these is a current waveform and the other is a voltage waveform. Here the symbol G_M or R_M does not represent an amplification in the usual sense of the word. Nevertheless, it is convenient to refer to each of the four quantities A_V, A_I, G_M and R_M as a transfer gain of the basic amplifier without feedback.

From Fig. 7.1 it is clear that

$$X_2 = A(X_1 \mp \beta X_2)$$

or $$X_2 (1 \pm A\beta) = AX_1$$

$$A_f = \frac{X_2}{X_1} = \frac{A}{1 \pm A\beta} \tag{7.1}$$

where

$A \Rightarrow$ Forward gain of the basic amplifier without feedback

$A_f \Rightarrow$ closed-loop gain of the amplifier with feedback

$A\beta \Rightarrow$ Loop-gain or return ratio loop transmission

$1 \mp A\beta \Rightarrow$ return difference

The terminology 'loop-gain' refers to the gain A of a signal path forward through the amplifier and then through the feedback path gain β, resulting in a closed loop back to its starting point. The gain around this closed loop is hence $A\beta$. The size and sign of $A\beta$, the loop-gain, very much controls the usefulness of the feedback.

Often the amount of feedback N introduced into an amplifier is expressed in decibels as

$$N(\text{dB}) = 20 \log \left(\frac{|A_f|}{|A|} \right) \tag{7.2}$$

If the output of the β-network X_f (= βX_2) combines with the input signal X_1 in antiphase (180° phase shift), then Eq. (7.1) assumes the form

$$A_f = \frac{X_2}{X_1} = \frac{A}{1 + A\beta} \tag{7.3}$$

which is known as the *negative feedback equation* for an amplifier. If the output of the β-network X_f, on the other hand, combines with the input signal X_1 in phase (0° phase shift), then Eq. (7.1) takes the form

$$A_f = \frac{X_2}{X_1} = \frac{A}{1 - A\beta} \tag{7.4}$$

which is known as the *positive feedback equation* for an amplifier.

Let us now consider some basic properties of Eqs. (7.3) and (7.4).

- If there is no feedback, that is, β = 0 Eq. (7.1) then becomes

$$A_f = \frac{A}{1 \pm 0} = A \tag{7.5}$$

- For negative feedback, the loop-gain $A\beta$ must be positive and as such $(1 + A\beta) > 1$ and from Eq. (7.3) the gain A_f with feedback is always less than unity. However, if $A\beta >> 1$, Eq. (7.3) becomes

$$A_f = \frac{A}{1 + A\beta} \Rightarrow \frac{A}{A\beta} \Rightarrow \frac{1}{\beta} \tag{7.6}$$

Equation (7.6) describes a case where the overall gain of the amplifier with feedback is no longer defined by the actual amplifier gain, i.e., the parameters of the basic amplifier in the forward path, but is defined only by the parameters of the feedback path. If the feedback components are

chosen to be of very stable type, then the overall gain of the amplifier A_f can also be very stable. This is an extremely useful property of the negative feedback amplifier.

- A positive feedback system being described by Eq. (7.4) gives infinite overall gain if $A\beta = 1$, i.e.,

$$A_f = \frac{A}{1 - A\beta} \Rightarrow \frac{A}{1 - 1} \Rightarrow \infty \Rightarrow \frac{X_2}{X_1} \tag{7.7}$$

which means that for $X_1 = 0$, $X_2 \neq 0$. In other words, a positive feedback amplifier with $A\beta = 1$ generates an output in the absence of an input and so it is termed an *oscillator*. Such an amplifier with positive feedback will be described in detail in Chapter 8.

7.3 MEANING OF THE LOOP-GAIN, *A*β

Consider the feedback system shown in Fig. 7.2. Note that the feedback loop is broken at X. From Fig. 7.2, we can write

$$\varepsilon_o = A\varepsilon_i \tag{7.8}$$

$$\varepsilon_f = \beta\varepsilon_o \tag{7.9}$$

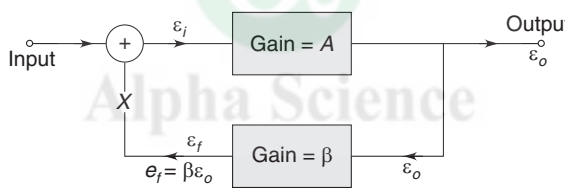


Fig. 7.2 *Meaning of feedback loop-gain*

Thus, using Eqs. (7.8) and (7.9), we may write

$$A\beta = \left(\frac{\varepsilon_o}{\varepsilon_i}\right)\left(\frac{\varepsilon_f}{\varepsilon_o}\right) = \left(\frac{\varepsilon_f}{\varepsilon_i}\right) \tag{7.10}$$

Thus, Eq. (7.10) represents the gain of the block consisting of amplifier A and the feedback network β.

7.4 SOME FUNDAMENTAL ASSUMPTIONS

For the feedback Eq. (7.1) to be valid:

- The basic amplifier must be unilateral from input to output and the reverse transmission should be zero.
- The feedback network, β-block, must be unilateral, so that it does not transmit a signal from input to output. This condition is often not

satisfied exactly, because β is a passive bilateral network. It is however, approximately valid for practical feedback connections.

- The reverse transmission factor β of the feedback network must be independent of the load and the source resistance R_L and R_S, respectively

7.5 SALIENT FEATURES OF NEGATIVE FEEDBACK AMPLIFIER

The salient features of negative feedback amplifiers are generally discussed under the topic of advantages and disadvantages. The advantages include:

- Improvement in gain stability
- Reduction in distortion and noise
- Improvement in bandwidth
- Control over impedances

The disadvantages are:

- Reduced gain
- Possibility of instability. Here, system instability refers to the occurrence of oscillations.

In the following sub-sections, we will discuss the salient features of the negative feedback amplifiers as mentional in this section.

7.5.1 Improvement in Gain Stability

The gain A of the basic amplifier changes with the change in the paramenters of the active devices constituting the amplifier. There are two factors responsible for changes in the gain of the amplifier.

- The device parameters are highly temperature dependent. This causes the gain of the amplifier to change with temperature.
- There is a wide spread in the values of paramenters of the devices of the same type. This causes the gain of the amplifier to change with the replacement of the concerned device in the amplifier.

The relative change in gain A being defined as (dA/A) is called the gain sensitivity. Referring to Fig. 7.1, the gain sensitivity of the negative feedback amplifier is defined as (dA_f/A_f). Thus using Eq. (7.3), we may write

$$\log A_f = \log A - \log (1 + A\beta)$$

which may be differentiated to yield

$$\frac{dA_f}{A_f} = \frac{1}{1 + A\beta} \frac{dA}{A} \tag{7.11}$$

For a negative feedback amplifier, $(1 + A\beta) >> 1$, and so from Eq. (7.11), we may write

$$\frac{dA_f}{A_f} < \frac{dA}{A} \tag{7.12}$$

Equation (7.12) indicates that the amplifier with negative feedback shows improved gain stability variations in the values of the gain A without feedback.

Sensitivity S is defined as the fractional change in amplification with negative feedback divided by the fractional change in amplification without feedback.

$$\text{Sensitivity} \Rightarrow \qquad S \Rightarrow \frac{dA_f/A_f}{dA/A} = \frac{1}{1 + A\beta} \tag{7.13}$$

and the desensitivity D is defined as

$$\text{Desensitivity} \Rightarrow \qquad D = |1 + A\beta| \tag{7.14}$$

7.5.2 Reduction in Distortion/Noise

There may be unwanted signals in an amplifier system in the form of a random noise, harmonics generated because of nonlinearity, man-made noise such as hum, or any other signal. Let us consider that the input signal is sinusoidal and that the distortion consists simply of a second-harmonic signal generated within the active devices. We assume that the second-harmonic component in the absence of feedback is equal to N_2. This situation is illustrated in Fig. 7.3.

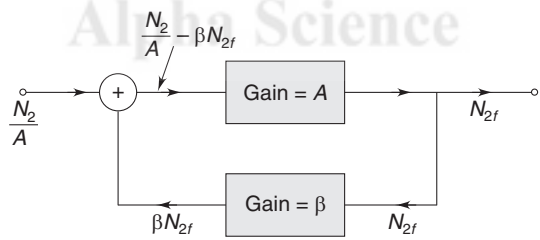


Fig. 7.3 *Illustrating reduction in noise/distortion*

Referring to Fig. 7.3, we can write

$$N_{2f} = A\left(\frac{N_2}{A} - \beta N_{2f}\right) = N_2 - A\beta N_{2f}$$

$$N_{2f} = \frac{N_2}{1 + A\beta} \tag{7.15}$$

where N_2 is the noise at the output without feedback and N_{2f} is the noise at the output with feedback.

Obviously, from Eq. (7.15)

$$N_{2f} < N_2$$

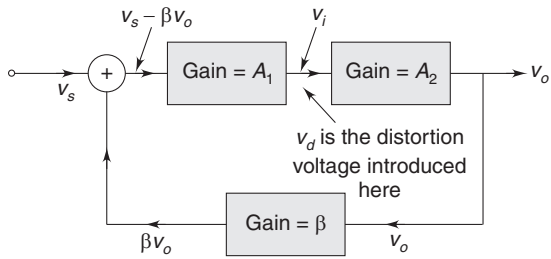


Fig. 7.4 *Illustration of reduction in distortion/noise*

For further illustration consider the feedback system shown in Fig. 7.4. Referring to this figure, we may write

$$v_i = A_1 (v_S - \beta v_o) + v_d \tag{7.16}$$

$$v_o = A_2 v_i \tag{7.17}$$

Using Eqs. (7.16) and (7.17),

$$v_o = A_2 A_1 (v_S - \beta v_o) + A_2 v_d$$

Simplification gives

$$v_o = \left(\frac{A_1 A_2}{1 + A_1 A_2 \beta}\right) v_S + \left(\frac{A_2}{1 + A_1 A_2 \beta}\right) v_d \tag{7.18}$$

For negative feedback $A_1 A_2 >> 1$, and hence using Eq. (7.18)

$$v_o = \frac{v_S}{\beta} + \frac{v_d}{A_1 \beta} \tag{7.19}$$

Here note that v_S is the signal voltage, and v_d is the distortion voltage introduced at the input of the second amplifier. Equation (7.19) shows that the disturbance is reduced by the gain of the amplifier between the input and the point of introduction of the disturbance.

7.5.3 Improvement in Bandwidth

A typical frequency response of an amplifier is depicted in Fig. 7.5. The lower and upper cut-off frequencies are labeled as f_L and f_H, respectively. The amplifier's bandwidth can be increased by using negative feedback. This illustrated in the following discussion.

Analysis in the Low-Frequency Region

The gain A of a single-pole amplifier in the low-frequency region may be written as

$$A_L = \frac{A_0}{1 - j(f_L/f)} \tag{7.20}$$

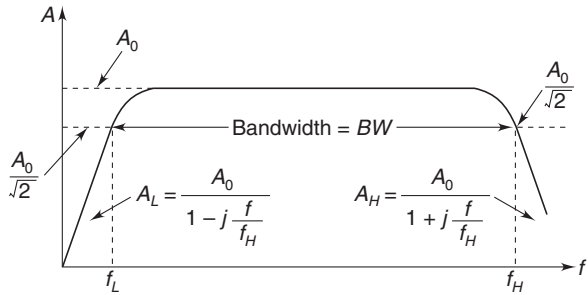


Fig. 7.5 *Frequency response of a typical amplifier*

where

A_L is the gain of the amplifier in the low-frequency region without feedback

A_0 is the gain of the amplifier in mid-band without feedback

f_L is the lower cut-off frequency

With negative feedback, using Eqs. (7.3) and (7.20), we can write an expression for the overall gain A_{Lf} in the low-frequency region as

$$A_{Lf} = \frac{A_L}{1 + A_L\beta} = \frac{\dfrac{A_0}{1 - jf_L/f}}{1 + \beta\, A_0/1 - jf_L/f} \tag{7.21}$$

Simplification of Eq. (7.21) gives

$$A_{Lf} = \frac{A_{0f}}{1 - \dfrac{jf_L}{f}} \tag{7.22}$$

where

$$A_{0f} = \frac{A_0}{1 + A_0\beta} \tag{7.23}$$

and

$$f_{Lf} = \frac{f_L}{1 + A_0\beta} \tag{7.24}$$

Note that A_{0f} is the mid-band gain with negative feedback and f_{Lf} is the lower cut-off frequency with negative feedback. It is obvious that

$$A_{0f} < A_0 \quad \text{and} \quad f_{Lf} < f_L$$

The results obtained are further illustrated in Fig. 7.6. Thus the lower cut-off frequency f_{Lf} with feedback is decreased by the factor $(1 + A_0\beta)$ and so also is the mid-band gain which is decreased with feedback by the same factor $(1 + A_0\beta)$.

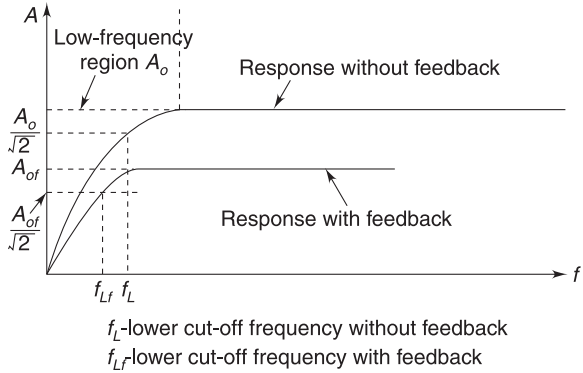


Fig. 7.6 *Effect of negative feedback on low-frequency response*

Analysis in the High-Frequency Region

The gain A of a single-pole amplifier in the high-frequency region is given by

$$A_H = \frac{A_0}{1 + j(f/f_H)} \tag{7.25}$$

where

A_0 is the mid-band gain without feedback

f_H is the upper cut-off frequency

A_H is the high-frequency gain without feedback

With negative feedback, using Eqs. (7.3) and (7.25), we may write

$$A_{Hf} = \frac{A_H}{1 + \beta A_H} = \frac{\dfrac{A_0}{1 + j(f/f_H)}}{1 + \beta \dfrac{A_0}{1 + j(f/f_H)}} \tag{7.26}$$

$$A_{Hf} = \frac{A_{0f}}{1 + j(f/f_H)} \tag{7.27}$$

where

$$A_{0f} = \frac{A_0}{1 + \beta A_0} \tag{7.28}$$

and

$$f_{Hf} = f_H\,(1 + \beta A_0) \tag{7.29}$$

Obviously, $A_{0f} < A_0$ and $f_{Hf} > f_H$. Thus we see that the upper cut-off frequency of the amplifier with negative feedback is increased by a factor of $(1 + A_0\beta)$.

Gain Bandwidth Product

For an amplifier without feedback the bandwidth BW is defined as

$$\text{BW} = f_H - f_L \tag{7.30}$$

Similarly, for an amplifier with feedback the bandwidth BW_f is defined as

$$\text{BW}_f = f_{Hf} - f_{Lf} \tag{7.31}$$

Thus using Eqs. (7.24) and (7.29),

$$\text{BW}_f = f_H (1 + \beta A_0) - \frac{f_L}{1 + \beta A_0} \tag{7.32}$$

This shows that the effective bandwidth with feedback BW_f is increased, i.e., $\text{BW}_f > \text{BW}$ for $f_H >> f_L$.

Gain in mid-band with feedback × Bandwidth with feedback

$$= A_{0f} \times \text{BW}_f \tag{7.33}$$

Thus,

$$\frac{A_0}{1 + \beta A_0} [f_H (1 + \beta A_0)] = A_0 f_H \tag{7.34}$$

which is equal to (Gain × Bandwidth) product without feedback. *Thus the gain bandwidth product of a negative feedback amplifier approximately remains constant*.

7.5.4 Control over Input and Output Impedances

The input and output resistances of an amplifier can be increased or decreased depending on whether current or voltage feedback is employed at the port. This property will be discussed while dwelling on the different types of feedback connections.

Feedback can be applied in several ways. It can sample output current or output voltage, and feedback a voltage or a current in parallel or in series with the input. Generally, the effect of feedback will be to improve the features of amplifier by the factor $(1 + A\beta)$.

7.6 TYPES OF FEEDBACK CONNECTIONS

There are four ways of connecting the feedback signal. Both voltage and current can be feedback to the input either in series or parallel in four different manners. These connections are:

1. Voltage-series feedback: called the voltage amplifier
2. Current-series feedback: called the transconductor amplifier
3. Current-shunt feedback: called the current amplifier
4. Voltage-shunt feedback: called the transresistance amplifier

7.6.1 Voltage-Series Negative Feedback Amplifier: Voltage Amplifier

The first word (voltage) in the name speaks about the nature of sampling and the second word (series) gives an idea about the nature of connection at the input.

The configuration in Fig. 7.7 shows the voltage-series feedback system. The equivalent circuit is depicted in Fig. 7.8.

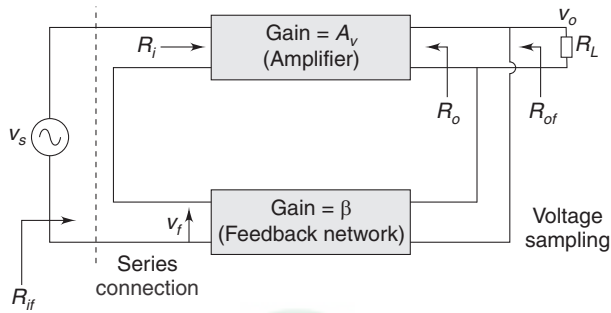


Fig. 7.7 *Voltage-series feedback connection*

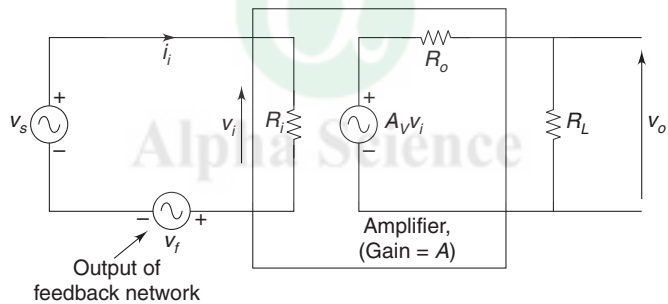


Fig. 7.8 *Equivalent circuit of Fig. 7.7*

Referring to Fig. 7.8, for the input loop, we write

$$v_S = i_i R_i + v_f \tag{7.35}$$

and

$$v_f = \beta v_o \tag{7.36}$$

For the output loop in Fig. 7.8, we can write

$$v_o = \frac{A_V v_i R_L}{R_o + R_L} = A_{VL} v_i \tag{7.37}$$

where A_{VL} stands for the voltage gain of the amplifier with the load (called the *loaded voltage gain*)and is given by

$$A_{VL} = \frac{A_V R_L}{R_o + R_L} \tag{7.38}$$

Now, $$v_o = A_{VL}(v_S - v_f) = A_{VL} v_S - A_{VL} \beta v_o \tag{7.39}$$

or $$v_o (1 + \beta A_{VL}) = A_{VL} v_S \tag{7.40}$$

or $$\frac{v_o}{v_S} = \frac{A_{VL}}{1 + \beta A_{VL}} = A_{vf} \tag{7.41}$$

Note that

$$\lim_{R_L \to \infty} A_{VL} = \lim_{R_L \to \infty} \left(\frac{A_V}{1 + (R_o/R_L)} \right) \Rightarrow A_V \tag{7.42}$$

which shows that the loaded gain A_{VL} is less than the unloaded gain A_V.

Relation for Input Resistance

In the input loop, we can write

$$v_S = i_i R_i + v_f = i_i R_i + \beta v_o \tag{7.43}$$

or $$v_S = i_i R_i + \beta \left(\frac{A_{VL}}{1 + \beta A_{VL}} \right) v_S \tag{7.44}$$

Simplification gives

$$\frac{v_S}{i_i} = R_i (1 + \beta A_{VL}) \Rightarrow R_{if} \tag{7.45}$$

Thus,

$$R_{if} = R_i (1 + \beta A_{VL}) \tag{7.46}$$

From Eq. (7.46), it may be seen that the input resistance is increased by a factor of the return difference.

Relation for Output Resistance

The output resistance is defined as

$$R_{of} = \frac{v}{i}\bigg|_{v_S = 0} \tag{7.47}$$

where v is the voltage source which we apply at the output terminals in place of R_L and i is the current supplied by the source with the signal source replaced by its internal resistance. Under the condition that $v_S = 0$, the equivalent circuit assumes the form as shown in Fig. 7.9. At the output loop in Fig. 7.9, we can write

$$v - A_V v_i = iR_o \tag{7.48}$$

or $$v = iR_o + A_V v_i \tag{7.49}$$

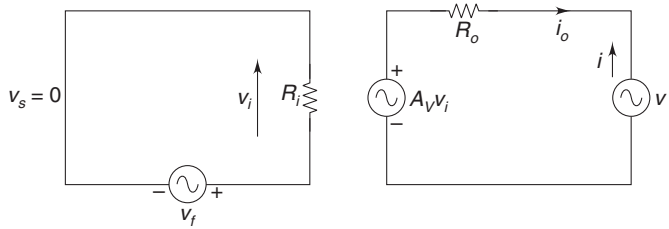


Fig. 7.9 *Determination of output resistance*

Since $v_i = -v_f$ and $v_f = \beta v$, simplification gives

$$\frac{v}{i}\bigg|_{v_S = 0} = \frac{R_o}{1 + \beta A_V} \Rightarrow R_{of} \tag{7.50}$$

and referring to Fig. 7.8, we have

$$R_{of}^{'} = R_{of} \parallel R_L.$$

From Eq. (7.50), it may be seen that the output resistance with feedback is decreased by a factor of the return difference.

Thus for a voltage amplifier, we have seen that the input resistance is increased and the output resistance is decreased by a factor of the return difference. The negative feedback thus helps in improving the features of a voltage amplifier.

7.6.2 Current-Series Negative Feedback Amplifier: Transconductance Amplifier

A current-series feedback system is shown in Fig. 7.10. Note that here the input is voltage and the output is current and hence the transfer function has the unit of conductance. The equivalent circuit is shown in Fig. 7.11.

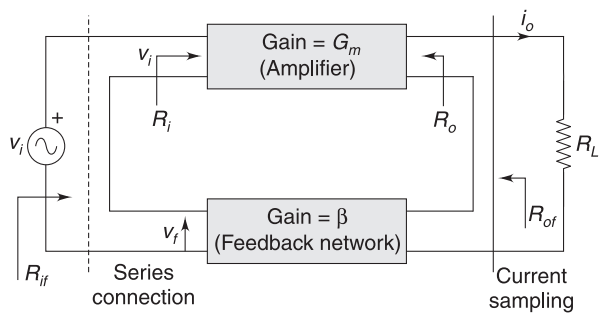


Fig. 7.10 *Current-series feedback connection*

Gain Relation

Referring to the Fig. 7.11, in the input loop, we can write

$$v_S = i_i R_i + v_f = i_i R_i + \beta i_o \tag{7.51}$$

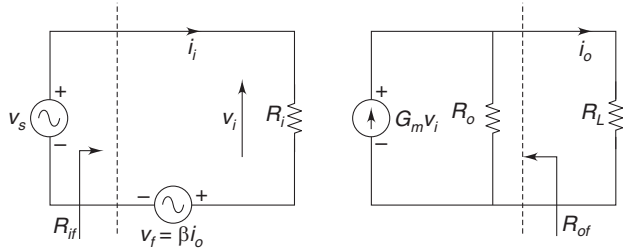


Fig. 7.11 *Equivalent circuit of Fig. 7.10*

For the output loop, we have

$$i_o = \left(\frac{R_o}{R_o + R_L}\right) G_m \, v_i$$

Simplification gives

$$G_{Mf} = \frac{i_o}{v_S} = \frac{G_M}{1 + \beta G_M} \tag{7.52}$$

Note that

$$G_M = R_o \left(\frac{G_M}{R_o + R_L}\right) \tag{7.53}$$

and

$$\lim_{R_L \to 0} \left(\frac{G_m \, R_o}{R_o + R_L}\right) \Rightarrow G_m \tag{7.54}$$

The gain function being gives by Eq. (7.54) is thus decreased in the current-series feedback by a factor of the return difference.

Relation for Input Resistance

For the equivalent circuit shown in Fig. 7.11, we can write

$$v_S = i_i \, R_i + v_f = i_i \, R_i + \beta i_o$$

$$= i_i \, R_i + \beta v_S \left(\frac{G_M}{1 + \beta G_M}\right)$$

or

$$v_S \left(1 - \frac{\beta G_M}{1 + \beta G_M}\right) = i_i \, R_i$$

or

$$\frac{v_S}{i_i} = R_i \, (1 + \beta G_M) \Rightarrow R_{if}$$

or

$$R_{if} = R_f\,(1 + \beta G_M) \tag{7.55}$$

The input resistance is thus increased by a factor of the return difference.

Relation for Output Resistance

The equivalent circuit to calculate the output resistance under the condition that the signal voltage v_S is zero is shown in Fig. 7.12.

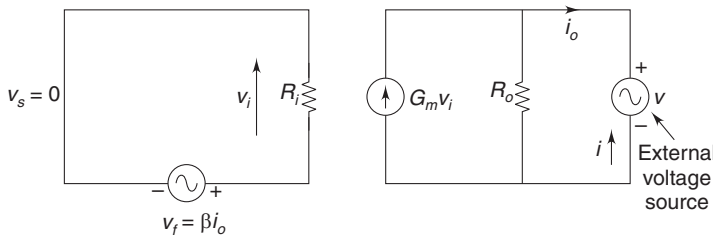


Fig. 7.12 *Determination of output resistance*

Referring to Fig. 7.12, in the output loop, we may write

$$i = \frac{v - G_m\, v_i\, R_o}{R_o} = \frac{v}{R_o} - G_m\, v_i$$

Since R_{of} is to be calculated for $v_S = 0$, so

$$0 = v_i + \beta i_o$$

or referring to Fig. 7.12,

$$v_i = -\beta i_o = \beta i \quad (\text{since } i = -i_o)$$

Therefore,

$$i = \frac{v}{R_o} - G_m\, \beta i$$

or

$$i\,(1 + \beta G_m) = \frac{v}{R_o}$$

or

$$i = \frac{v}{R_o\,(1 + \beta G_m)}$$

Therefore,

$$\left.\frac{v}{i}\right|_{v_S = 0} = R_o\,(1 + \beta G_m) \Rightarrow R_{of}$$

Thus, we get

$$R_{of} = R_o\,(1 + \beta G_m) \tag{7.56}$$

and referring to Fig. 7.11, we may write

$$R'_{of} = R_{of} \parallel R_L = \frac{R_{of}\, R_L}{R_{of} + R_L} \tag{7.57}$$

Equation (7.56) shows that the output resistance of a current-series or transconductance feedback amplifier is increased by a factor of the return difference.

The negative feedback in this current-series amplifier is shown to have improved the features of the amplifier.

7.6.3 Current-Shunt Negative Feedback Amplifier: Current Amplifier

A current-shunt negative feedback system is shown in Fig. 7.13. Here note that the input is current and the output is also current, so it is a current amplifier. The equivalent circuit is shown in Fig. 7.14. Note that the feedback current and the signal current are in phase opposition because of negative feedback.

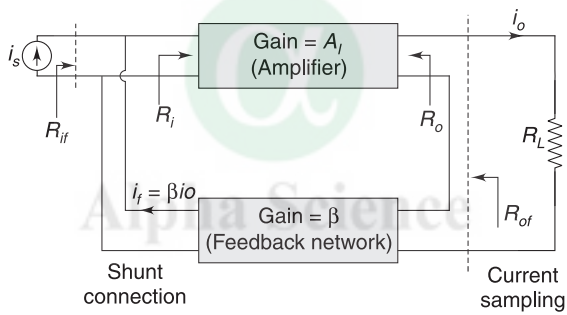


Fig. 7.13 *Current-shunt negative feedback*

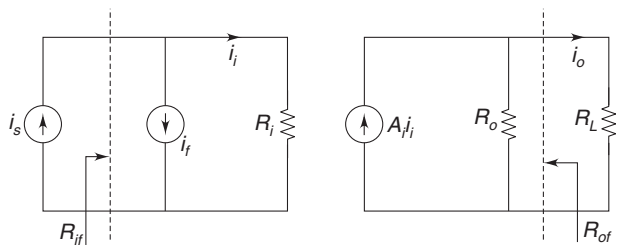


Fig. 7.14 *Equivalent circuit of Fig. 7.13*

Gain Relation

Referring to Fig. 7.14 for the output loop, we can write

$$i_o = \frac{A_I\, i_i\, R_o}{R_o + R_L} = A_{IL}\, i_i \tag{7.58}$$

where

$$A_{IL} = \frac{A_I R_o}{R_o + R_L} \tag{7.59}$$

Further note that

$$\lim_{R_L \to 0} \left(\frac{A_I R_o}{R_o + R_L} \right) \Rightarrow A_I \tag{7.60}$$

Here A_{IL} is called the current gain with load, i.e., the *loaded current gain*. Also for the input loop in Fig. 7.14, we may write

$$i_S = \beta i_o + i_i \tag{7.61}$$

Thus,

$$i_o = A_{IL}\, i_i = A_{IL}\, (i_S - \beta i_o)$$

or

$$i_o\, (1 + \beta A_{IL}) = i_S\, A_{IL}$$

or

$$\frac{i_o}{i_S} = \frac{A_{IL}}{1 + \beta A_{IL}} \Rightarrow A_{if}$$

Thus,

$$A_{if} = \frac{A_{IL}}{1 + \beta A_{IL}} \tag{7.62}$$

Relation for Input Resistance

Referring to Fig. 7.14 for the input loop, we can write

$$\begin{aligned} v_i &= i_i\, R_i = R_i\, (i_S - \beta i_o) \\ &= R_i\, i_S - \beta R_i\, \frac{A_{IL}\, i_S}{1 + \beta A_{IL}} \end{aligned}$$

or

$$\frac{v_i}{i_S} = \frac{R_i}{1 + \beta A_{IL}} \Rightarrow R_{if}$$

Thus

$$R_{if} = \frac{R_i}{1 + \beta A_{IL}} \tag{7.63}$$

Relation for Output Resistance

In order to calculate the output resistance we assume a voltage source v connected in place of R_L, with input signal replaced by its internal resistance, which sends a current i in the opposite direction as shown in Fig. 7.15. We define the output resistance R_{of} as

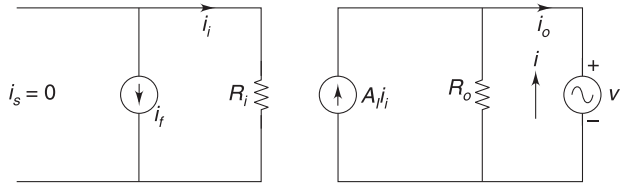


Fig. 7.15 *Determination of output resistance*

$$R_{of} = \left.\frac{v}{i}\right|_{i_S = 0}$$

The current signal source here is assumed to be open-circuited. In the output loop, we can write

$$i = \frac{v - A_I\, i_i\, R_o}{R_o} = \frac{v}{R_o} - A_I\, i_i$$

As $i_S = 0$, $i_i = -\beta i_o = \beta i$ as $i = -i_o$. Therefore,

$$i = \frac{v}{R_o} - A_I\, \beta i$$

or

$$\left.\frac{v}{i}\right|_{i_S = 0} = R_o\,(1 + \beta A_I) \Rightarrow R_{of}$$

or

$$R_{of} = R_o\,(1 + \beta A_I) \tag{7.64}$$

and further referring to Fig. 7.14,

$$R'_{of} = R_{of} \parallel R_L \tag{7.65}$$

The negative feedback in the current amplifier thus decreases the input resistance and increases the output resistance which enhances the features of the current amplifier.

7.6.4 Voltage-Shunt Negative Feedback: Transresistance Amplifier

The voltage-shunt feedback system is shown in Fig. 7.16. Here it should be noted that the input is the current and the output is the voltage and hence the transfer function has the unit of resistance. Because of this, the system is known as the transresistance amplifier. The equivalent circuit of Fig. 7.16 is shown in Fig. 7.17.

Gain Relation

Referring to Fig. 7.17 for the input loop, we may write

$$i_S = \beta v_o + i_i \tag{7.66}$$

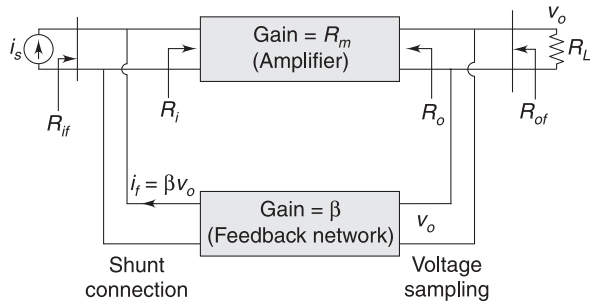


Fig. 7.16 *Voltage-shunt negative feedback*

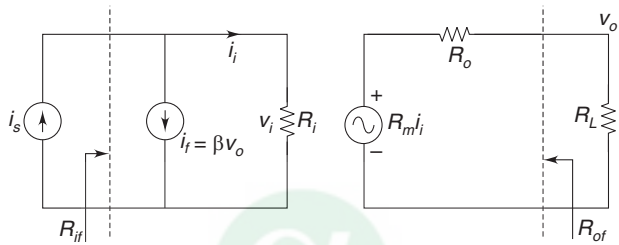


Fig. 7.17 *Equivalent circuit of Fig. 7.16*

and for the output loop

$$v_o = \frac{R_m\, i_i\, R_L}{R_o + R_L} = R_M\, i_i \tag{7.67}$$

where

$$R_M = \frac{R_m\, R_L}{R_o + R_L} \tag{7.68}$$

Further note that

$$\lim_{R_L \to \infty} \left(\frac{R_m\, R_L}{R_o + R_L} \right) \Rightarrow R_m \tag{7.69}$$

Using Eq. (7.66) in Eq. (7.67), we get

$$v_o = R_M\, i_i = R_M\, (i_S - \beta v_o)$$

or

$$\frac{v_o}{i_S} = \frac{R_M}{1 + \beta R_M} \Rightarrow R_{Mf}$$

Thus,

$$R_{Mf} = \frac{R_M}{1 + \beta R_M} \tag{7.70}$$

Relation for Input Resistance

Referring to Fig. 7.17, an expression for the input resistance may be written as

$$R_{if} = \frac{v_i}{i_S} = \frac{i_i R_i}{i_S} = \frac{R_i i_i}{i_S}$$

Using Eq. (7.67) in Eq. (7.66), we get

$$\frac{i_i}{i_S} = \frac{1}{1 + \beta R_M}$$

Thus,

$$R_{if} = R_i \left(\frac{1}{1 + \beta R_M} \right) \tag{7.71}$$

Relation for Output Resistance

In order to calculate the output resistance we assume a voltage source v connected in place of R_L, with input signal replaced by its internal resistance. As shown the voltage source v sends a current i in the opposite direction as depicted in Fig. 7.18. From the input loop, we can write $i_i = -\beta v_o = -\beta v$ because $v = v_o$ and $i_S = 0$.

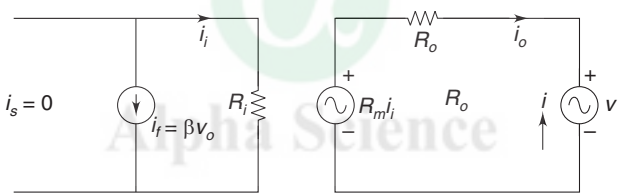


Fig. 7.18 *Determination of output resistance*

For the output loop in Fig. 7.18, we can write

$$i = \frac{v - R_m i_i}{R_o} = \frac{v + \beta R_m v}{R_o}$$

$$= v \left(\frac{1 + \beta R_m}{R_o} \right)$$

Thus,

$$\left. \frac{v}{i} \right|_{i_S = 0} = R_{of} = \frac{R_o}{1 + \beta R_m} \tag{7.72}$$

and referring to Fig. 7.17,

$$R'_{of} = R_{of} \parallel R_L \tag{7.73}$$

The negative feedback in the transresistance amplifier decreases the input and output resistance, thereby improving the features of the amplifier.

7.7 ADVANTAGES OF NEGATIVE FEEDBACK

There are numerous advantages of negative feedback which outweigh its only drawback of reduction in gain. These are as follows:

1. **Gain Stability** The voltage gain of an amplifier with negative feedback is given as

$$A_f = \frac{A}{1 + A\beta}.$$

If $A\beta >> 1$, then the expression becomes

$$A_f = \frac{1}{\beta}$$

We overall gain A_f is independent of internal gain and depends only on feedback ratio β and β in turn depends on the passive elements such as resistors. Resistors remain fairly constant and so the gain is stabilized.

2. **Reduced Non-linear Distortion** A large signal stage has non-linear distortion which is reduced by a factor $(1 + A\beta)$ when negative feedback is used.
3. **Reduced Noise** There is always a noise voltage in the amplifier which is reduced by a factor $(1 + A\beta)$ when negative feedback is used.
4. **Increased Bandwidth (or Improved Frequency Response)** The bandwidth (BW) of an amplifier without feedback is equal to the separation between 3 dB frequencies f_1 and f_2. If A is the gain then gain-bandwidth product is $A \times$ BW. With the negative feedback, the amplifier gain is reduced and since the gain-bandwidth product has to remain constant, so the BW will increase to compensate for the reduction in gain.
5. **Increased Input Impedance** The input impedance of the amplifier with negative feedback is increased by a factor $(1 + A\beta)$.
6. **Reduced Output Impedance** The output impedance of the amplifier with negative feedback is reduced by a factor $(1 + A\beta)$.

7.8 DISADVANTAGES OF NEGATIVE FEEDBACK SYSTEM

7.8.1 Reduction in Gain

In all the four types of negative feedback connections it has been shown that the gain is reduced by a factor $(1 + A\beta)$ where A is the forward gain and β is the gain of the feedback network. The gain A takes the form of either voltage gain or current gain or transconductance gain or transresistance gain depending upon the use of the basic amplifier in the feedback system.

7.8.2 Possibility of Instability (Nyquist-Plot)

So far we assumed ideal feedback conditions, i.e., real amplifier gain A and the real feedback gain β. In the practice we may encounter complex gain A, especially at the extreme edges of the passband of the amplifier, as well as complex gain β. If negative feedback is applied, then the gain function is explicity given by the expression

$$A_f = \frac{A}{1 + A\beta} \tag{7.74}$$

Hence as $A\beta$ approaches the value $-1 + j0$, the denominator in Eq. (7.74) becomes very small, i.e., there occurs an increase in gain, And at the frequency where $A\beta$ equals $-1 + j0$, the amplifier has infinite gain and does not need any input to give an output. This condition indicates positive feedback when oscillations begin to occur. In other words, the amplifier becomes unstable. If the amplifier oscillates at some low or high frequency, it is no longer useful as an amplifier. A proper negative feedback amplifier design requires that the circuit be stable at all frequencies, not merely those in the range of interest. Otherwise, a transient disturbance could cause a seemingly stable amplifier to suddenly start oscillating.

Thus, the stability of any circuit or system with negative feedback can be studied by looking at its $A\beta$ plot as a function of frequency. Nyquist showed that such a plot of $|A\beta| \angle\theta$ in the polar coordinates is a closed curve on the complex plane, when all frequencies are taken into account. The nature and location of this curve, known as the Nyquist plot, indicates either stability or instability.

For an amplifier in general, we can say that the loop-gain $A\beta$ is a function of frequency and thus we can write the loop gain $A\beta$ as

$$\text{Loop-gain} = A\beta(j\omega) \tag{7.75}$$

The Nyquist plot is obtained when we plot the imaginary part of $A\beta(j\omega)$ versus the real part of $A\beta(j\omega)$ with frequency as a parameter ($0 \le \omega \le \infty$). The Nyquist plot, in effect, combines in a single plot the two Bode plots of the gain versus frequency and the phase shift versus frequency.

A typical plot of the loop-gain $A\beta(j\omega)$ in the complex plane is shown in Fig. 7.19. A few points at frequencies ω_1, ω_2, ω_3, ω_4, are shown in Fig. 7.19. At the frequency ω_3, the magnitude of the loop gain is the length of phasor OX and its phase shift is the angle between OX and the conventional positive real axis which represents zero phase shift. Instead of drawing a whole family of phasors similar

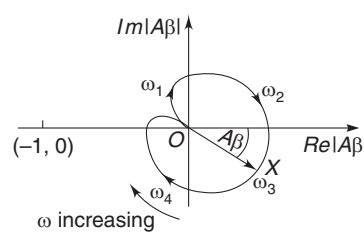


Fig. 7.19 *Nyquist plot*

to *OX*, one each corresponding to other frequencies, only the locus of *X*, the end of the phasor, is plotted with some frequencies as shown in Fig. 7.19.

The Nyquist criteria for stability can thus be stated as follows:

The amplifier is unstable if the Nyquist curve as plotted encloses (encircles) the – 1 point, and it is stable otherwise.

An example of the Nyquist criteria is demonstrated by the curve in Fig. 7.20. The Nyquist plot in Fig. 7.20(a) is stable since it does not encircle the – 1 point, whereas that shown in Fig. 7.20(b) is unstable since the curve does encircle the – 1 point. Note that encircling the – 1 point means that at a phase shift of 180° the loop-gain $A\beta$ is greater than unity, therefore, the feedback signal is in phase with the input and large enough to result in a larger input signal than that applied, with the result that oscillation occurs.

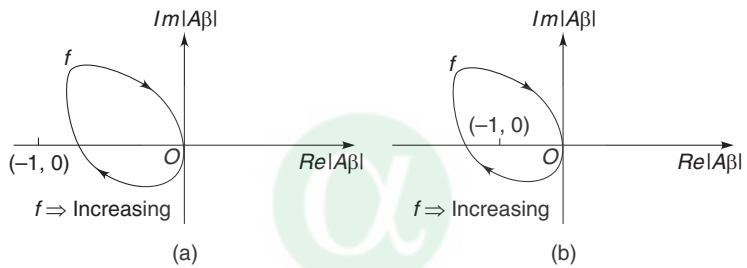


Fig. 7.20 *Illustration of Nyquist criteria*

Questions

7.1 What are different types of feedback?

7.2 What do you understand by feedback in amplifiers? Explain the terms feedback factor and open-loop gain.

7.3 What do you understand by feedback in amplifiers? List the four basic negative feedback configurations and describe the effect of feedback on R_i and R_o in each of these configurations.

7.4 List five characteristics of an amplifier which are modified by negative feedback. Explain them in brief.

7.5 Discuss the effects of negative feedback on gain, stability of gain and noise.

7.6 Prove in a negative feedback amplifier

$$\left|\frac{dA_f}{A_f}\right| = \frac{1}{1 + \beta A}\left|\frac{dA}{A}\right|$$

where A_f = gain with feedback

A = transfer gain

β = feedback factor

7.7 Make a distinction between the "voltage" feedback and "current" feedback in amplifier circuits. Discuss the merits in each case and derive the expressions for the net output impedance in each case.

7.8 Write short notes on any two:

(i) Advantages of negative feedback

(ii) Stability criterion

(iii) Classification of feedback amplifiers

(iv) Nyquist plot

8 Sinusoidal Oscillators

8.1 INTRODUCTION

An a.c. generator (called an alternator) generates sinusoidal a.c. power of 50 Hz, it cannot be called an oscillator. An alternator is a mechanical device that has rotating parts (e.g., a loop carrying a coil), converts mechanical energy into ac energy but cannot produce ac energy of high frequency (exceeding 1000 Hz) because for generating higher frequencies, either the speed of rotation of the armature has to be made extremely high or the number of poles has to be made large and both options are impracticable. On the other hand, an oscillator is a non-rotating electronic device that converts d.c. energy into a.c. energy of frequency ranging from a few Hz to many GHz.

Though alternator generates large amount of a.c. power but for several applications such as radio transmitters and receivers, radars, etc., an oscillator is preferred owing to the under-mentioned advantages over alternators:

(i) Portable and cheap in cost.

(ii) An oscillator has no rotating parts, so it is silent and there is no wear and tear.

(iii) Frequency of oscillator may be conveniently varied.

(iv) Frequency once set, remains constant for a considerable time period.

(v) Voltages or currents of any frequency (20 Hz to extremely high frequencies) adjustable over a wide range can be generated.

(vi) Voltages free from harmonics as well as rich with harmonics can be generated by sinusoidal oscillators and relaxation oscillators respectively.

(vii) High operation efficiency – due to absence of moving parts.

For having a sinusoidal oscillator, we require an amplifier with positive feedback. The idea is to use the feedback signal in place of an input signal. If the (feedback) loop-gain and phase are correct, there will be an output signal even though there is no external input signal (see Fig. 8.1).

Thus, an oscillator is an amplifier that is modified by positive feedback to supply its own input signal.

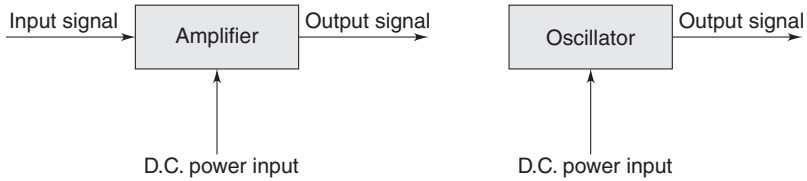


Fig. 8.1

8.2 KINDS OF OSCILLATORS

When employed as an instrument, an oscillator may be called with different names such as *test oscillator* (or *signal generator*), standard signal generator (or function generator) depending on the function performed by the device, type of signal produced by it, order of sophistication etc.

Oscillators may be broadly divided into two categories *viz*., *harmonic* oscillators and *relaxation* oscillators. Both types can include active devices such as BJTs, FETs and op-amps, and passive components such as resistors, inductors and capacitors. In harmonic oscillators, the energy always flows in one direction—from the active to passive components. However, in relaxation oscillators, the energy is exchanged between the active and passive components. In harmonic oscillators, the frequency of oscillations is determined by the feedback path. However, in relaxation oscillators, the frequency is determined by time constants – specifically, the charge and discharge time constants during the exchange of energy. Harmonic oscillators can develop low – distortion sinusoidal output waveforms, but relaxation oscillators can only generate nonsinusoidal waveforms such as sawtooth, square or triangular. Harmonic oscillators will be discussed here.

Oscillators can also be categorized on the basis of design principle used, and the frequency range over which they are employed.

The oscillators, according to the design principle used are feedback oscillators and negative resistance oscillators. *Feedback oscillators* are widely employed.

The oscillators, according to operating frequency, may be classified as follows:

Type of Oscillators	*Approximate range*
1. Audio frequency (AF) oscillators	20 Hz-20 kHz
2. Radio frequency (RF) oscillators	20 Hz-30 MHz
3. Very low frequency (VLF) oscillators	15-100 kHz
4. Low frequency (LF) oscillators	100-500 kHz
5. Broadcast oscillators	500 kHz-1.5 MHz
6. Video frequency oscillators	0-5 MHz

7. High frequency (HF) oscillators — 1.5-30 MHz
8. Very high frequency (VHF) oscillators — 30-300 MHz
9. Ultra high frequency (UHF) oscillators — 300-3,000 MHz
10. Microwave oscillators — Beyond 3 GHz (3,000 MHz)

8.3 OPERATION OF OSCILLATOR

The use of positive feedback that results in a feedback amplifier having closed-loop gain $|A_f|$ exceeding unity and satisfies the phase conditions results in operation as an oscillator circuit. An oscillator circuit then provides a constantly varying output signal. If the output signal varies sinusoidally, the circuit is referred to as a *sinusoidal oscillator* and on the other hand if the output voltage rises quickly to one voltage level and later drops, quickly to another voltage level, the circuit is usually referred to as a pulse or *square-wave generator*.

To understand how an oscillator produces an output signal without an external input signal, let us consider the feedback circuit shown in Fig. 8.2(a), where V_{in} is the voltage of a.c. input driving the input terminals *bc* of an amplifier having voltage gain *A*. The amplified output voltage is

$$V_{out} = AV_{in} \tag{8.1}$$

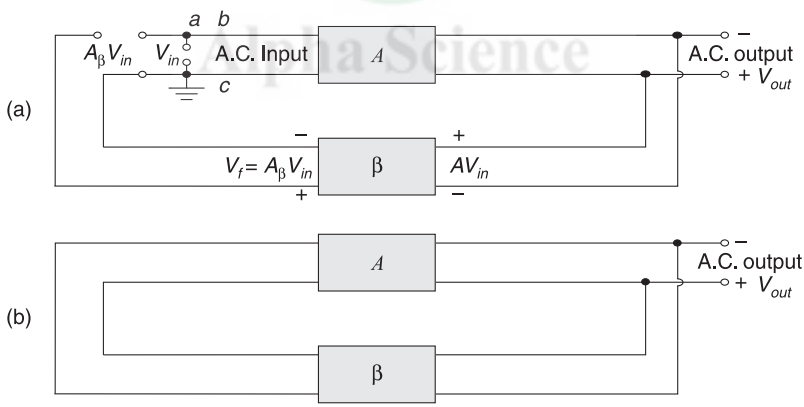


Fig. 8.2

This voltage drives a feedback circuit that is usually a *resonant circuit*, as we get maximum feedback at one frequency. The feedback voltage returning to point *a* is given by equation

$$V_f = A\beta V_{in} \tag{8.2}$$

where β is the gain of feedback network.

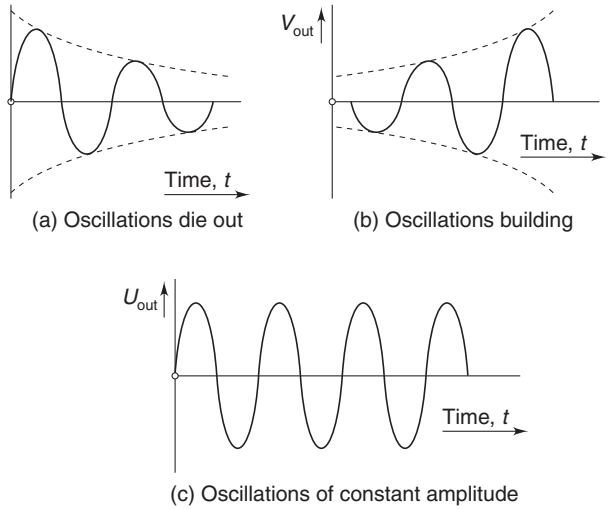


Fig. 8.3

If the phase shift through the amplifier and feedback circuit is zero, then $A\beta V_{in}$ is in phase with the input signal V_{in} that drives the input terminals of the amplifier.

Now we consider point '*a*' to point '*b*' and simultaneously remove voltage source V_{in}, then feedback voltage $A\beta V_{in}$ drives the input terminals *bc* of the amplifier, as shown in Fig. 8.2(b). In case $A\beta$ is less than unity, $A\beta V_{in}$ is less than V_{in} and the output signal will die out, as illustrated in Fig. 8.3(a). On the other hand if $A\beta$ is greater than unity, the output signal will build up, as illustrated in Fig. 8.3(b). If $A\beta$ is equal to unity, $A\beta V_{in}$ equals V_{in} and the output signal is a steady sinewave, as illustrated in Fig. 8.3(c). In this case, the circuit supplies its own input signal and produces a sinusoidal output.

Certain conditions are required to be fulfilled for sustained oscillations and these conditions are that (i) the loop gain of the circuit must be equal to (or slightly greater than) unity and (ii) phase shift around the circuit must be zero. These two conditions for sustained oscillators are called *Barkhausen criteria*.

For initiation of oscillations, supply of an input signal is not essential. Only the condition $\beta A = 1$ must be satisfied for self-sustained oscillations to result. In practices βA is made slightly greater than unity, the system starts oscillating by amplifying noise voltage which is always present. Saturation factors in the practical circuits *provide* an average value of βA of 1. The resulting waveforms are never exactly sinusoidal. However, the closer the value of βA is to exactly 1, the more nearly sinusoidal is the waveform. Figure 8.4 shows how the noise voltage results in a buildup of a steady-state Oscillation condition.

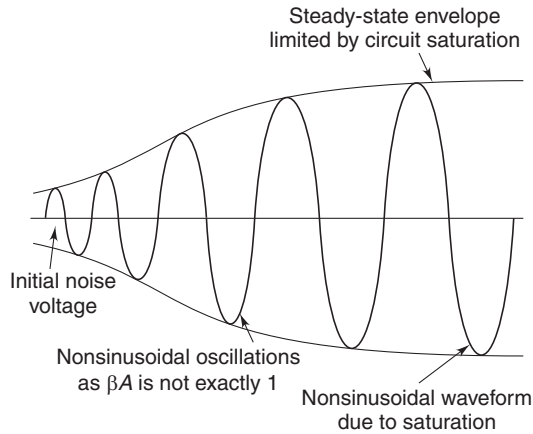


Fig. 8.4 *Buildup of steady-state oscillations*

Another way of seeing how the feedback circuit provides operation as an oscillator is obtained by noting the denominator in the basic feedback equation, $A_f = A/(1 + \beta A)$. When $\beta A = -1$ or magnitude 1 at a phase angle of 180°, the denominator becomes zero and the gain with feedback, A_f, becomes infinite. Thus, an infinitesimal signal (noise voltage) can provide a measurable output voltage, and the circuit acts as an oscillator even without an input signal.

By deliberate design the phase shift around the loop is made 0° at the resonant frequency. Above and below the resonant frequency, the phase shift is different from 0°. Thus, oscillations are obtained at only one frequency, the resonant frequency of the feedback circuit.

To understand and apply the Barkhausen criterion, we must consider both the gain and phase shift of $A\beta$ as a function of frequency. Reactive elements, capacitance in particular, contained in the amplifier and/or feedback causes the gain magnitude and phase shift to vary with the frequency. In general, there will be only one frequency at which the gain magnitude is unity and at which, simultaneously, the total phase shift is equivalent to 0° (in phase – a multiple of 360°). The *system will oscillate at the frequency that satisfies these conditions*. Designing an oscillator amounts to selection of reactive components and their incorporation into circuit in such a way that the conditions will be satisfied at a desired frequency.

To show the dependence of the loop gain $A\beta$ on frequency, we express $A\beta(j\omega)$, a complex phasor that can be expressed in both polar and rectangular form:

$$A\beta(j\omega) = |A\beta| \angle\theta = |A\beta| \cos\theta + j|A\beta| \sin\theta \qquad (8.3)$$

where $|A\beta|$ is the magnitude of the loop gain, a function of frequency, and θ is the phase shift, also a function of frequency. For satisfaction of Barkhausen criterion, we must have

$$|A\beta| = 1 \tag{8.4}$$

and

$$\theta = \pm 360^\circ \, n \tag{8.5}$$

where n is any integer, including 0. In polar and rectangular forms, the **Barkhausen criterion** is expressed as

$$A\beta(j\omega) = 1 \angle \pm 360^\circ \, n = 1 + j0 \tag{8.6}$$

8.4 ESSENTIALS OF AN OSCILLATOR CIRCUIT

From the above discussion it can be inferred that an oscillator circuit must have the following three elements:

1. Oscillatory circuit or frequency determining element.
2. Amplifier
3. Positive feedback network

The oscillatory circuit or a frequency-determining element can be an inductance-capacitance network (*L*-*C* tank), resistance-capacitance (*R*-*C* network), or a quartz crystal, depending on the frequency and waveshape desired.

The electronic amplifier receives d.c. power from the battery or d.c. power supply and converts it into a.c. power for supply to the tank circuit. The oscillations occurring in the tank circuit are applied to the input of the electronic amplifier. Because of the amplifying properties of the amplifier, we get increased output of these oscillations. This amplified output of oscillations is because of d.c. power supplied from the external source (a battery or power supply). The output of the amplifier can be supplied to the tank circuit to meet the losses.

The *feedback* network supplies a part of the output power to the frequency-determining element in correct phase to aid the oscillations. In other words, feedback circuit must be a *positive feedback network*. The block diagram of an oscillator is shown in Fig. 8.5.

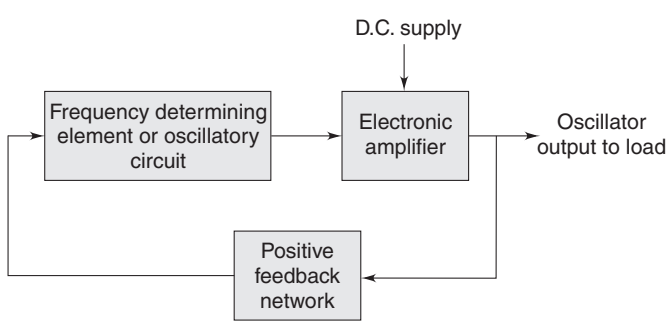


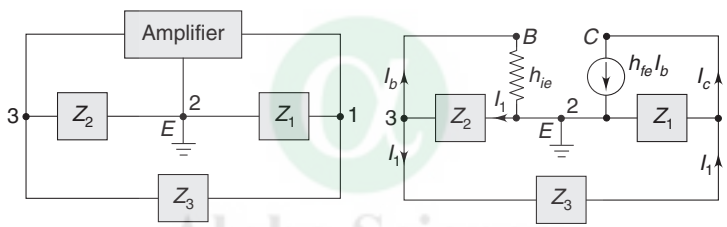
Fig. 8.5 *Block diagram of an oscillator*

8.4.1 General Form of an L-C Oscillator

In the general form of an oscillator depicted in Fig. 8.6(a), any of the active devices such as vacuum tube, BJT, FET or op-amp may be used in the amplifier section, Z_1, Z_2, and Z_3 are the reactive elements constituting the feedback tank circuit which determines the frequency of oscillation. Here, Z_1 and Z_2 serve as an ac voltage divider for the output voltage and feedback signal. Thus, the voltage across Z_2 is the feedback signal. The equivalent circuit is shown in Fig. 8.6(b) with the following two assumptions:

(i) h_{re} of transistor is negligibly small and, therefore, the feedback source h_{re} V_{out} is negligible.

(ii) h_{oe} of the transistor is very small i.e., the output resistance $\frac{1}{h_{oe}}$ is very large and, therefore, $\frac{1}{h_{oe}}$ is omitted from the equivalent circuit.



(a) General form of an oscillator (b) Equivalent circuit

Fig. 8.6

Let us determine the load impedance between output terminals 1 and 2. Here Z_2 and h_{ie} are in parallel and their resultant impedance is in series with impedance Z_3.

The equivalent impedance is in parallel with impedance Z_1.

Thus load impedance between output terminals is given as

$$Z_L = Z_1 \parallel [Z_3 + (Z_2 \parallel h_{ie})]$$

$$= Z_1 \parallel \left[Z_3 + \frac{Z_2 h_{ie}}{Z_2 + h_{ie}} \right]$$

$$= Z_1 \parallel \left[\frac{Z_3 (Z_2 + h_{ie}) + Z_2 h_{ie}}{Z_2 + h_{ie}} \right]$$

$$= Z_1 \parallel \left[\frac{h_{ie} (Z_2 + Z_3) + Z_2 Z_3}{Z_2 + h_{ie}} \right]$$

or
$$\frac{1}{Z_L} = \frac{1}{Z_1} + \frac{Z_2 + h_{ie}}{h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3}$$
$$= \frac{h_{ie}\,(Z_1 + Z_2 + Z_3) + Z_1\,Z_2 + Z_2\,Z_3}{Z_1\,[h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3]}$$

or
$$Z_L = \frac{Z_1\,[h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3]}{h_{ie}\,(Z_1 + Z_2 + Z_3) + Z_1\,Z_2 + Z_2\,Z_3} \tag{8.7}$$

The voltage gain of a CE amplifier without feedback is given as

$$A = \frac{-\,h_{fe}\,Z_L}{h_{ie}} \tag{8.8}$$

The output voltage between terminals 1 and 2 is given as

$$V_{\text{out}} = Z_3 + \left[\frac{Z_2\,h_{ie}}{Z_2 + h_{ie}}\right] I_1$$
$$= \left[\frac{h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3}{Z_2 + h_{ie}}\right] I_1$$

The voltage feedback to the input terminals 2 and 3 given as

$$V_f = \frac{Z_2\,h_{ie}}{Z_2 + h_{ie}}\,I_1$$

So feedback fraction,

$$\beta = \frac{V_f}{V_{\text{out}}} = \frac{Z_2\,h_{ie}}{h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3} \tag{8.9}$$

Applying the criterion of oscillation i.e., $A\beta = 1$, we have

$$\frac{-\,h_{fe}\,Z_L}{h_{ie}} \cdot \frac{Z_2\,h_{ie}}{h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3} = 1$$

or
$$\frac{h_{fe}\,Z_1\,[h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3]}{h_{ie}\,(Z_1 + Z_2 + Z_3) + Z_1\,Z_2 + Z_2\,Z_3} \cdot \left[\frac{Z_2}{h_{ie}\,(Z_2 + Z_3) + Z_2\,Z_3}\right] = -\,1$$

or
$$\frac{h_{fe}\,Z_1\,Z_2}{h_{ie}\,(Z_1 + Z_2 + Z_3) + Z_1\,Z_2 + Z_2\,Z_3} = -\,1$$

or
$$h_{ie}\,(Z_1 + Z_2 + Z_3) + Z_1\,Z_2 + Z_2\,Z_3 = -\,h_{fe}\,Z_1\,Z_2$$

or
$$h_{ie}\,(Z_1 + Z_2 + Z_3) + Z_1\,Z_2\,(1 + h_{fe}) + Z_2\,Z_3 = 0 \tag{8.10}$$

This is the general equation for the oscillator.

8.5 TYPES OF TRANSISTOR OSCILLATORS

A transistor can be operated as an oscillator for producing continuous undamped oscillations of any desired frequency, if tank (or oscillatory) and feedback circuits are properly connected to it. All oscillators under different names have similar function i.e., they generate continuous undamped output. However, they differ in methods of supplying energy to the tank or oscillatory circuit to meet the losses and the frequency ranges over which they are used.

The frequency spectrum over which oscillators are employed to produce sinusoidal signals is extremely wide (from less than 1 Hz to many GHz). However, no single oscillator design is practical for generating signals over this entire range. Instead, a variety of designs are employed, each of which generates sinusoidal outputs most advantageously over various portions of the frequency spectrum.

Oscillators, which use *inductance-capacitance* (*L-C*) circuits as their tank or oscillatory circuits, are very popular for generating high frequency (e.g., 10 kHz to 100 MHz) outputs. The most widely used *L-C* oscillators are the *Hartley* and *Colpitt's oscillators*. Although they slightly differ from one another in their electronic circuitry but they have virtually identical frequency ranges and frequency-stability characteristics. However, such oscillators are not suitable for generating low-frequency sinusoidal outputs. This is due to the fact that the components required in construction of low-frequency *L-C* resonant circuits are too bulky and heavy. So *resistor-capacitor* (*R-C*) *oscillators* are generally employed for generating low-frequency (from 1 Hz to about 1 MHz) sinusoidal signals. Two most common *R-C* oscillators are the *Wien-bridge* and *phase shift types*.

Other less frequently used oscillators are the *crystal oscillators* and the *negative resistance oscillators*. The operating frequency ranges of various types of most commonly used oscillators are given below:

Type of oscillator	*Approximate frequency ranges*
Wien bridge oscillator	1 Hz — 1 MHz
Phase shift oscillator	1 Hz — 10 MHz
Hartley oscillator	10 kHz — 100 MHz
Colpitt's oscillator	10 kHz — 100 MHz
Negative resistance oscillator	> 100 MHz
Crystal oscillator	Fixed frequency

The oscillators of various types will be discussed in the following sections.

8.6 COLPITT'S OSCILLATOR

The Colpitt's oscillator circuit is a superb circuit and is widely used in commercial signal generators up to 100 MHz. The basic circuit of a transistor Colpitt's oscillator is shown in Fig. 8.7. It basically consists of a single stage inverting amplifier and an L-C phase shift network, as obvious from the circuit diagram shown. The two series capacitors C_1 and C_2 form the potential divider used for providing the feedback voltage— the voltage developed across capacitor C_2 provides the regenerative feedback required for sustained oscillations. Parallel combination of R_E and C_E along with resistors R_1 and R_2 provides the stabilized self bias. The collector supply voltage V_{CC} is applied to the collector through a radio frequency choke (RFC) which permits an easy flow of direct current but at the same time it offers very high impedance to the high frequency currents. The presence of coupling capacitor C_{C_2} in the output circuit does not permit the dc currents to go to the tank circuit (the flow of d.c. current in a tank circuit reduces its Q). The radio frequency energy developed across RFC is capacitively coupled to the tank circuit through the capacitor C_{C_2}. The output of the phase shift L-C network is coupled from the junction of L and C_2 to the amplifier input at base through coupling capacitor C_{C_1}, which blocks dc but provides path to ac. Transistor itself produces a phase shift of 180° and another phase shift of 180° is provided by the capacitive feedback. Thus, a total phase shift of 360° is obtained which is an essential condition for developing oscillations. The output voltage is derived from a secondary winding L' coupled to the inductance L. The frequency is determined by the tank circuit and is varied by gang-tuning the two capacitors C_1 and C_2. It is to be noted that capacitors C_1 and C_2 are ganged. As the tuning is varied, values of both capacitors vary simultaneously, the ratio of the two capacitances remaining the same.

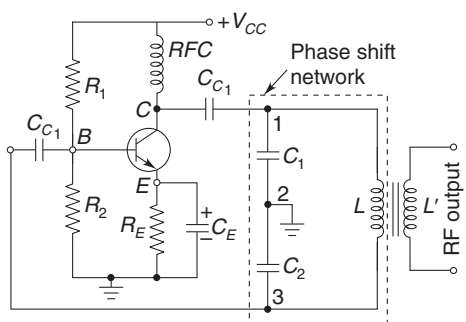


Fig. 8.7 *Basic circuit for transistor Colpitt's oscillator*

Working When the collector supply voltage V_{CC} is switched on, the capacitors C_1 and C_2 are charged. These capacitors C_1 and C_2 discharge through the coil L, setting up oscillations of frequency $f = \frac{1}{2\pi}\sqrt{\frac{1}{LC_1} + \frac{1}{LC_2}}$. The oscillations across capacitor C_2 are applied to the base-emitter junction and appear in the amplified form in the collector circuit. Of course, the amplified output in the collector circuit is of the same frequency as that of the oscillatory circuit. This amplified output in the collector circuit is supplied to the tank circuit in order to meet the losses. Thus the tank circuit is getting energy continuously from the collector circuit to make up for the losses occurring in it and, therefore, ensures undamped oscillations. The energy supplied to the tank circuit is of correct phase, as already explained, and if $A\beta$ exceeds unity, oscillations are sustained in the circuit.

Frequency of Oscillation

Here referring to Fig. 8.6,

$$Z_1 = \frac{1}{j\omega C_1} = \frac{-j}{\omega C_1};$$

$$Z_2 = \frac{1}{j\omega C_2} = \frac{-j}{\omega C_2};$$

and $$Z_3 = j\omega L$$

Substituting these values in Eq. (8.10), we have

$$h_{ie}\left[\frac{-j}{\omega C_1} - \frac{j}{\omega C_2} + j\omega L\right] + \left[\frac{-j}{\omega C_1} \cdot \frac{-j}{\omega C_2}\right](1 + h_{fe}) + \left[\frac{-j}{\omega C_2}\right] j\omega L = 0$$

or $$-jh_{ie}\left[\frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L\right] - \frac{1 + h_{fe}}{\omega^2 C_1 C_2} + \frac{L}{C_2} = 0 \quad (8.11)$$

Equating the imaginary component of the above Eq. (8.11) to zero, we have

$$h_{ie}\left[\frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L\right] = 0$$

or $$\frac{1}{\omega C_1} + \frac{1}{\omega C_2} = \omega L \qquad \because h_{ie} \neq 0$$

or $$\frac{C_1 + C_2}{\omega C_1 C_2} = \omega L$$

or $$\omega^2 = \frac{C_1 + C_2}{LC_1 C_2}$$

or $$\omega = \sqrt{\frac{C_1 + C_2}{LC_1 C_2}} = \sqrt{\frac{1}{LC_1} + \frac{1}{LC_2}}$$

or $$f = \frac{1}{2\pi}\sqrt{\frac{1}{LC_1} + \frac{1}{LC_2}} \tag{8.12}$$

The above Eq. (8.12) gives the frequency of oscillation. Equating the real component of Eq. (8.11) to zero, we have

$$\frac{1 + h_{fe}}{\omega^2 C_1 C_2} = \frac{L}{C_2}$$

or $$1 + h_{fe} = \omega^2 LC_1$$

or $$1 + h_{fe} = \frac{C_1 + C_2}{LC_1 C_2} \times LC_1 = \frac{C_1 + C_2}{C_2} = 1 + \frac{C_1}{C_2}$$

or $$h_{fe} = \frac{C_1}{C_2} \tag{8.13}$$

As for other oscillator circuits, the loop gain must be greater than unity to ensure that the circuit oscillates.

So $$A\beta \geq 1 \text{ or } A \geq \frac{1}{\beta} \geq \frac{C_2}{C_1} \tag{8.14}$$

8.7 HARTLEY OSCILLATOR

The transistor Hartley oscillator is as popular as Colpitt's oscillator and is widely used as a local oscillator in radio receivers. The circuit arrangement is shown in Fig. 8.8. Hartley oscillator circuit is similar to Colpitt's oscillator circuit, except that phase shift network consists of two inductors L_1 and L_2 and a capacitor C instead of two capacitors and one inductor. The output of the amplifier is applied across inductor L_1 and the voltage across inductor L_2 forms the feedback voltage. The coil L_1 is inductively coupled to coil L_2, the combination functions as an auto-transformer. However, because of direct connection, the junction of L_1 and L_2 cannot be directly grounded. Instead, another capacitor C_L is used. The operation of the circuit is similar to that of the Colpitt's oscillator circuit.

Considering the fact that there exists mutual inductance between coils L_1 and L_2 because the coils are wound on the same core, their net effective inductance is increased by mutual inductance M. So in this case effective inductance is given by the equation

$$L = L_1 + L_2 + 2M \tag{8.15}$$

and resonant or oscillation frequency is given by the equation

$$f = \frac{1}{2\pi\sqrt{[C(L_1 + L_2 + 2M)]}},$$

as derived below:

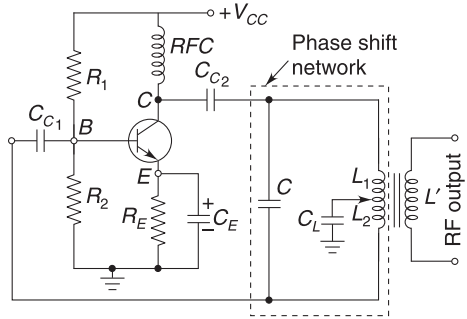


Fig. 8.8 *Basic circuit for Hartley oscillator*

Frequency of Oscillation The general Eq. (8.10) derived in Section 8.4.1 is reproduced here

$$h_{ie}\,(Z_1 + Z_2 + Z_3) + Z_1\,Z_2\,(1 + h_{fe}) + Z_2\,Z_3 = 0 \tag{8.16}$$

Here $Z_1 = j\omega L_1 + j\omega M;$

$Z_2 = j\omega L_2 + j\omega M$

and $Z_3 = \dfrac{1}{j\omega C} = -\,j/\omega C$

Substituting these values in general Eq. (8.10) of Section 8.4.1

$$h_{ie}\left[(j\omega L_1 + j\omega M) + (j\omega L_2 + j\omega M) - \frac{j}{\omega C}\right] + (j\omega L_1 + j\omega M) \times$$

$$(j\omega L_2 + j\omega M)\,(1 + h_{fe}) + (j\omega L_2 + j\omega M)\left(\frac{-j}{\omega C}\right) = 0$$

or $$j\omega h_{ie}\left[L_1 + L_2 + 2M - \frac{1}{\omega^2 C}\right]$$

$$-\,\omega^2\,(L_2 + M)\left[(L_1 + M)\,(1 + h_{fe}) - \frac{1}{\omega^2 C}\right] = 0 \tag{8.17}$$

Equating the imaginary part of above Eq. (8.17) to zero, we get

$$\omega h_{ie}\left[L_1 + L_2 + 2M - \frac{1}{\omega^2 C}\right] = 0$$

or $$L_1 + L_2 + 2M - \frac{1}{\omega^2 C} = 0 \qquad \because \omega h_{ie} \neq 0$$

or $$\omega^2 C = \frac{1}{L_1 + L_2 + 2M}$$

or $$f = \frac{\omega}{2\pi} = \frac{1}{2\pi\sqrt{C(L_1 + L_2 + 2M)}} \tag{8.18}$$

The above Eq. (8.18) gives the frequency of oscillation.

Equating the real component of Eq. (8.17) to zero, we get

$$(L_1 + M)(1 + h_{fe}) - \frac{1}{\omega^2 C} = 0$$

or $$(1 + h_{fe}) = \frac{1}{\omega^2 C(L_1 + M)} = \frac{L_1 + L_2 + 2M}{L_1 + M}$$

$$= 1 + \frac{L_2 + M}{L_1 + M}$$

or $$h_{fe} = \frac{L_2 + M}{L_1 + M} \tag{8.19}$$

As for other oscillator circuits, the loop gain must be greater than 1 to ensure that circuit oscillates.

So $$A\beta \geq 1$$

or $$A \geq \frac{1}{\beta} \geq \frac{L_2 + M}{L_1 + M} \tag{8.20}$$

8.8 PHASE SHIFT OSCILLATOR

The circuit arrangement of a phase shift oscillator using *NPN* transistor in CE configuration is shown in Fig. 8.9. As usual, the voltage divider $R_1 - R_2$ provides dc emitter base bias, R_E and C_E combination provides temperature stability and prevent ac signal degeneration and collector resistor R_C controls the collector voltage. The oscillator output voltage is capacitively coupled to the load by C_C.

In case of a transistor phase shift oscillator, the output of the feedback network is loaded appreciably by the relatively small input resistance (h_{ie}) of the transistor. Hence, instead of employing voltage-series feedback (as used in case of FET phase shift oscillator), voltage-shunt feedback is used for a transistor

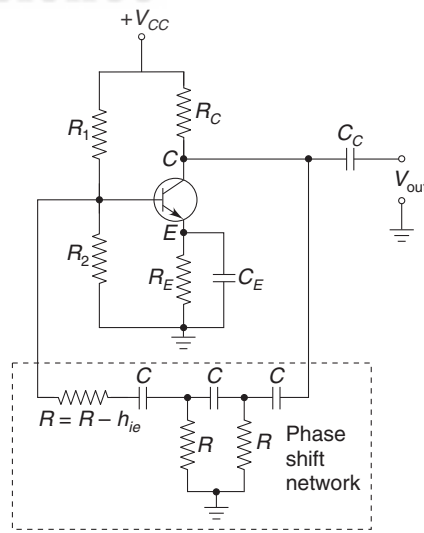


Fig. 8.9 *Basic circuit of a transistor phase shift oscillator*

phase shift oscillator, as shown in Fig. 8.9. In this circuit, the feedback signal is coupled through the feedback resistor R' in series with the amplifier stage input resistance h_{ie}. The value of R' should be such that when added with amplifier stage input resistance h_{ie}, it is equal to R i.e., $R' + h_{ie} = R$.

Operation The circuit is set into oscillations by any random or variation caused in the base current, that may be either due to noise inherent in the transistor or minor variation in voltage of d.c. power supply. This variation in base current is amplified in collector circuit. The output of the amplifier is supplied to an R-C feedback network. The *R-C* network produces a phase shift of 180° between output and input voltages. Since CE amplifier produces a phase reversal of the input signal, total phase shift becomes 360° or 0° which is essential for *regeneration* or for sustained oscillations. The output of this network is thus in the same phase as the originally assumed input to the amplifier and is applied to the base terminal of the transistor. Thus, sustained variation in collector current between saturation and cut-off values are obtained. *R-C* phase shift network is the frequency determining network.

Frequency of Oscillation The equivalent circuit for the analysis of a transistor phase shift oscillator (circuit shown in Fig. 8.9) is shown in Fig. 8.10. The equivalent circuit shown in Fig. 8.10 is simplified if the following assumptions are made.

(i) h_{re} of the transistor is usually negligibly small and therefore, $h_{re}\ V_{out}$ is omitted from the circuit.

(ii) h_{oe} of the transistor is very small i.e., $1/h_{oe}$ is much larger than R_C. Thus the effect of h_{oe} can be neglected.

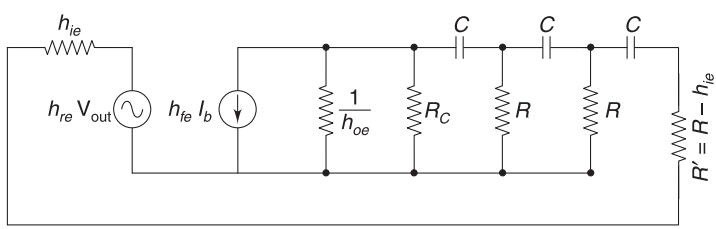


Fig. 8.10 *Equivalent circuit of transistor phase-shift oscillator circuit shown in Fig. 8.9*

Making above assumptions and replacing source by equivalent voltage source, the simplified equivalent circuit is shown in Fig. 8.11.

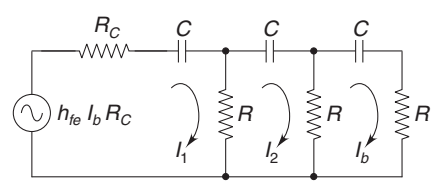


Fig. 8.11

Applying Kirchhoff's voltage law to the three loops shown in Fig. 8.11, we have

$$\left(R + R_C + \frac{1}{j\omega C}\right) I_1 - RI_2 + h_{fe}\, I_b\, R_C = 0 \tag{8.21}$$

$$-RI_1 + \left(2R + \frac{1}{j\omega C}\right) I_2 - RI_b = 0 \tag{8.22}$$

$$0 - RI_2 + \left(2R + \frac{1}{j\omega C}\right) I_b = 0 \tag{8.23}$$

As the currents I_1, I_2, and I_b are non-vanishing, the determinant of the coefficients of I_1, I_2, and I_b must be zero. Substituting $\frac{1}{\omega C} = X_C$, we have

$$\begin{vmatrix} (R + R_C - jX_C) & (-R) & (h_{fe}\, R_C) \\ (-R) & (2R - jX_C) & (-R) \\ 0 & (-R) & (2R - jX_C) \end{vmatrix} = 0 \tag{8.24}$$

or $(R + R_C - jX_C)\,[(2R - jX_C)^2 - R^2]$

$$+ \mathrm{R}\,[(-R)\,(2R - jX_C) - h_{fe}\, R_C\,(-R)] = 0$$

or $(R + R_C - jX_C)\,(3R^2 - X_C^2 - j4RX_C]$

$$- \mathrm{R}\,[2R^2 - jRX_C - h_{fe}\, R_C R) = 0$$

or $R^3 + R^2R_C\,(3 + h_{fe}) - 5RX_C^2$

$$- \mathrm{R}CX_C^2 - 6jR^2X_C - j4RR_CX_C + jX_C^3 = 0 \tag{8.25}$$

Equating the imaginary component of the above equation to zero, we have

$$6R^2X_C + 4RR_CX_C - X_C^3 = 0$$

or $$X_C = \sqrt{6R^2 + 4RR_C}$$

or $$2\pi fC = \frac{1}{\sqrt{6R^2 + 4RR_C}} \qquad \because X_C = \frac{1}{\omega C} = \frac{1}{2\pi fC}$$

or $$f = \frac{1}{2\pi RC\sqrt{6 + \dfrac{4R_C}{R}}} \tag{8.26}$$

If $R = R_C$, then

$$f = \frac{1}{2\pi RC\sqrt{10}} \tag{8.27}$$

The above equation gives frequency of oscillation.

Equating the real component of Eq. (8.25) to zero, we have

$$R^3 + R^2R_C\,(3 + h_{fe}) - X_C^2\,(5R + R_C) = 0$$

or $$R^3 + R^2R_C\,(3 + h_{fe}) - (6R^2 + 4RR_C)\,(5R + R_C) = 0$$

or $$-29R^3 - 23R^2R_C + h_{fe}\,R^2R_C - 4RR_C^2 = 0$$

or $$\frac{-29R}{R_C} - 23 + h_{fe} - 4\frac{R_C}{R} = 0$$

or $$h_{fe} = 23 + 29\frac{R}{R_C} + 4\frac{R_C}{R} \tag{8.28}$$

For the loop gain to be greater than unity, the requirement of the current gain of the transistor is found to be

$$h_{fe} > 23 + 29\frac{R}{R_C} + 4\frac{R_C}{R} \tag{8.29}$$

If $R = R_C$, then

$$h_{fe} > (23 + 29 + 4) \text{ i.e., } 56 \tag{8.30}$$

8.9 WIEN BRIDGE OSCILLATOR

It is one of the most popular type of oscillators used in audio and sub-audio frequency ranges (20 – 20 kHz). This type of oscillator is simple in design, compact in size, and remarkably stable in its frequency output. Furthermore, its output is relatively free from distortion and its frequency can be varied easily. However, the maximum frequency output of a typical Wien bridge oscillator is only about 1 MHz.

This is also, in fact, a phase shift oscillator. It employs two transistors, each producing a phase shift of 180°, and thus producing a total phase shift of 360° or 0°.

The circuit diagram of Wien bridge oscillator is shown in Fig. 8.12. It is essentially a two-stage amplifier with an *R-C* bridge circuit. *R-C* bridge circuit (Wien bridge) is a lead-lag network (R_1-C_1 and R_2-C_2). The phase shift across the network lags with increasing frequency and leads with decreasing frequency.

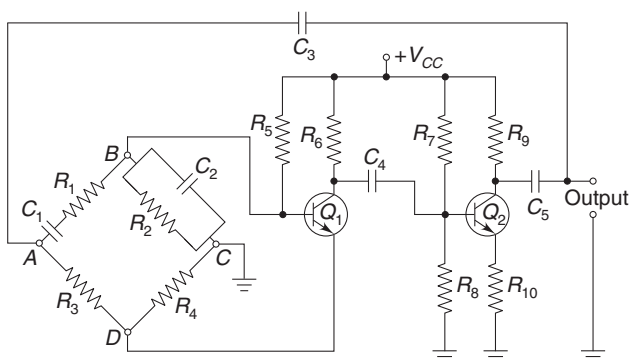


Fig. 8.12 *Wien bridge oscillator circuit*

By adding Wien bridge feedback network, the oscillator becomes sensitive to a signal of only one particular frequency. This particular frequency is that at which Wien bridge is balanced and for which the phase shift is 0°. If the Wien bridge feedback network is not employed and output of transistor Q_2 is fed back to transistor Q_1 for providing regeneration required for producing oscillations, the transistor Q_1 will amplify signals over a wide range of frequencies and thus direct coupling would result in poor frequency stability. Thus, by employing Wien bridge feedback network frequency stability is increased.

In the bridge circuit R_1 in series with C_1, R_3, R_4 and R_2 in parallel with C_2 form the four arms.

From the analysis of the bridge circuit it is obvious that the bridge will be balanced only when

$$R_3 \left[\frac{R_2}{1 + j\omega C_2 R_2}\right] = R_4 \left(R_1 - \frac{j}{\omega C_1}\right) \tag{8.31}$$

or $$R_2 R_3 = R_4 (1 + j\omega C_2 R_2) (R_1 - j/\omega C_1)$$

or $$R_2 R_3 - R_4 R_1 - \frac{C_2}{C_1} R_2 R_4 + \frac{jR_4}{\omega C_1} - j\omega C_2 R_2 R_1 R_4 = 0$$

Separating real and imaginary terms, we have

$$R_2 R_3 - R_4 R_1 - \frac{C_2}{C_1} R_2 R_4 = 0 \tag{8.32}$$

or $$\frac{C_2}{C_1} = \frac{R_3}{R_4} - \frac{R_1}{R_2} \tag{8.33}$$

and $$\frac{R_4}{\omega C_1} - \omega C_2 R_2 R_1 R_4 = 0 \tag{8.34}$$

or $$\omega^2 = \frac{1}{C_1 C_2 R_1 R_2} \tag{8.35}$$

or $$\omega = \frac{1}{\sqrt{C_1 C_2 R_1 R_2}}$$

and $$f = \frac{1}{2\pi \sqrt{R_1 C_1 R_2 C_2}} \tag{8.36}$$

if $C_1 = C_2 = C$ and $R_1 = R_2 = R$, then

$$f = \frac{1}{2\pi CR} \tag{8.37}$$

and $$R_3 = 2R_4 \tag{8.38}$$

Thus, we see that in a bridge circuit the output will be in phase with the input only when the bridge is balanced, i.e., at resonant frequency given by Eq. (8.37). At all other frequencies the bridge is off-balance i.e., the voltage fed back and output voltage do not have the correct phase relationship for sustained oscillations.

So this bridge circuit can be used as feedback network for an oscillator, provided that the phase shift through the amplifier is zero. This requisite condition is achieved by using a two-stage amplifier, as illustrated in Fig. 8.12. In this arrangement the output of the second stage is supplied back to the feedback network and the voltage across the parallel combination C_2 R_2 is fed to the input of the first stage. Transistor Q_1 serves as an oscillator and amplifier whereas the transistor Q_2 as an inverter to cause a phase shift of 180°. The circuit uses positive and negative feedbacks. The positive feedback is through R_1, C_1, R_2, C_2 to transistor Q_1 and negative feedback is through the voltage divider R_3 – R_4 to the emitter of transistor Q_1. Resistors R_3 and R_4 are used to stabilize the amplitude of the output.

The two transistors Q_1 and Q_2 thus cause a total phase shift of 360° and ensure proper positive feedback. The negative feedback is provided in the circuit to ensure constant output over a range of frequencies. This is achieved by taking resistor R_4 in the form of a temperature sensitive lamp, whose resistance increases with the increase in current. In case the amplitude of the output tends to increase, more current would provide more negative feedback. Thus the output would regain its original value. A reverse action would take place in case the output tends to fall.

The amplifier voltage gain,

$$A = \frac{R_3 + R_4}{R_4} = \frac{R_3}{R_4} + 1 = 3 \tag{8.39}$$

$\because$ from Eq. (8.38) $R_3 = 2R_4$

The above corresponds with the feedback network attenuation of 1/3. Thus, in this case, voltage gain A must be equal to or greater than 3, to sustain oscillations.

To have a voltage gain of 3 is not, difficult. On the other hand, to have a gain as low as 3 may be difficult. For this reason also negative feedback is essential.

Operation The circuit is set in oscillation by any random change in base current of transistor Q_1, that may be due to noise inherent in the transistor or variation in voltage of dc supply. This variation in base current is amplified in collector circuit of transistor Q_1 but with a phase shift of 180°. The output of transistor Q_1 is fed to the base of second transistor Q_2 through capacitor C_4. Now a still further amplified and twice phase-reversed signal appears at the collector of the transistor Q_2. Having been inverted twice, the output signal will be in phase with the signal input to the base of transistor Q_1. A part of the output signal at transistor Q_2 is fed back to the input points of the bridge circuit (points *A*-*C*). A part of this feedback signal is applied to emitter resistor R_4 where it produces degenerative effect (or negative feedback). Similarly, a part of the feedback signal

is applied across the base-bias resistor R_2 where it produces regenerative effect (or positive feedback). At the rated frequency, effect of regeneration is made slightly more than that of degeneration so as to obtain sustained oscillations.

The continuous frequency variation in this oscillator can be had by varying the two capacitors C_1 and C_2 simultaneously. These capacitors are variable air-gang capacitors. We can change the frequency range of the oscillator by switching into the circuit different values of resistors R_1 and R_2.

The advantage and disadvantages of Wien bridge oscillators are given below:

Advantages

1. It provides a stable low distortion sinusoidal output over a wide range of frequency.
2. The frequency range can be selected simply by using decade resistance boxes.
3. The frequency of the oscillator can be easily varied by varying capacitances C_1 and C_2 simultaneously.
4. The overall gain is high because of two transistors.

Disadvantages

1. The circuit needs two transistors and a large number of other components.
2. The maximum frequency output is limited because of amplitude and the phase shift characteristics of the amplifier.

Questions

8.1 What is an amplifier? How does it differ from an amplifier?

8.2 Give the two Barkhausen conditions required for sinusoidal oscillations to be sustained with the help of neat diagrams.

8.3 State and explain the condition under which feedback amplifier works as an oscillator.

8.4 Draw the circuit diagram of a Colpitt's oscillator, using transistor. Derive an expression for its frequency of oscillations. Deduce the starting condition for this oscillator.

8.5 Draw a neat circuit diagram of a Colpitt oscillator using *NPN* transistor. Give its equivalent circuit. Derive expressions for the following:

(i) the frequency of oscillations

(ii) The maximum gain for sustained oscillations

8.6 State Barkhausen criteria and explain the working of Hartley oscillator.

8.7 Explain briefly, how the oscillations are maintained in Hartley oscillator. Also derive the expression of frequency of oscillation.

8.8 What is the Barkhausen criteria for the feedback oscillator? Draw a neat diagram of a phase shift oscillator using BJT. Derive an expression for its frequency of oscillation.

8.9 Sketch and explain the circuit of a phase-shift oscillator using a bipolar junction transistor.

8.10 What is Barkhausen criterion? What do you mean by frequency stability and stability criterion in an oscillator? Explain the Wien bridge oscillator in detail.

8.11 Draw the circuit of a Wien bridge oscillator and explain how oscillations are generated?

8.12 Write short notes on:

(i) Stability criteria for an oscillator

(ii) Hartley oscillator

(iii) Colpitt oscillator

(iv) *R-C* phase shift oscillator

(v) Wien bridge oscillator

9 Multivibrators

9.1 INTRODUCTION

In certain circuits, we often need making (switching on) and breaking (switching off) of the circuit. It is also desirable that this making and breaking operation be very fast and without sparking. A circuit that can turn on or off current in it is referred to as a *switching circuit*.

9.2 TRANSISTOR AS A SWITCH

A transistor can be employed as an electronic switch. Operating a transistor as a switch means operating it at either saturation or cut-off but nowhere else along the load line. When a transistor is saturated, it is like a *closed switch* from the collector to emitter. When the transistor is cut-off, it is like an *open switch*. This operation of a transistor as a switch is illustrated in Fig. 9.1.

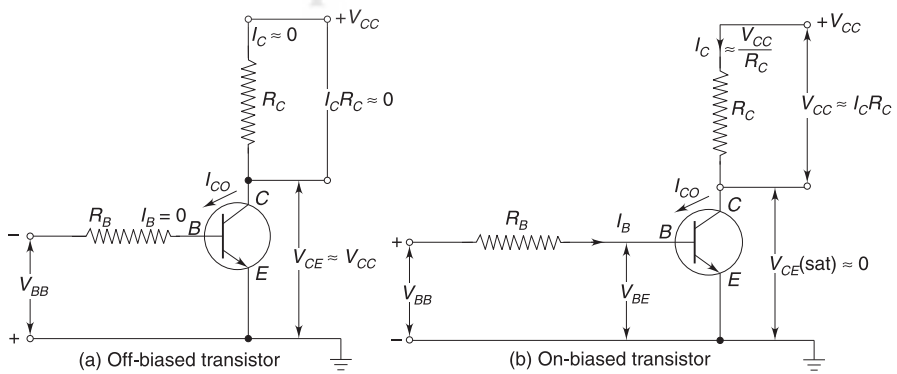


Fig. 9.1 *(a) Off-biased transistor (open switch) and (b) On-biased transistor (closed switch)*

Refer to Fig. 9.2, when a transistor is working as a switch, the cut-off (off) and saturation (on) regions are the stable regions of its operation while active region is the unstable (or transient) region of its operation. The operation of transistor passes through the active region when its condition is changed from one state to another (i.e. from off state to on-state or vice versa).

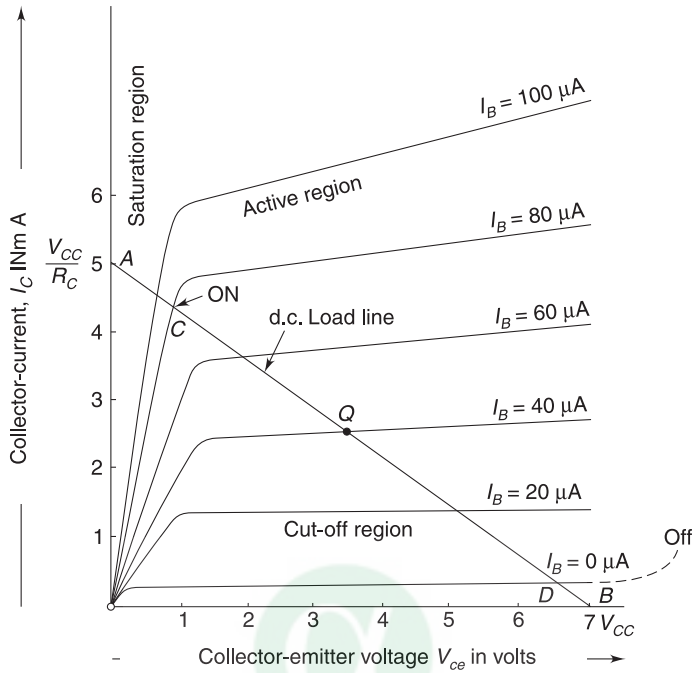


Fig. 9.2

In Fig. 9.2, the line-segment d.c. corresponds to the active region. The collector current I_C increases from I_{CO} to $I_{C\,(sat)}$ along the path d.c. as the transistor is switched-on. When the transistor is switched-off, I_C decreases from $I_{C\,(sat)}$ to I_{CO} along path CD. Due to I_{CO}, the transistor does not behave as an ideal switch (i.e., due to minority carrier leakage current), but for high speed switching transistor is superior to switches of other types.

Proper design for switching operation requires that the operating point switches from cut-off to saturation along the load line, as shown in Fig. 9.3. For our purposes, we will assume the $I_C = I_{CEO} = 0$ mA when $I_B = 0$ μA as shown in Fig. 9.3 (an excellent approximation in the light of improving fabrication techniques). In addition we will assume that $V_{CE} = V_{CE\,(sat)} = 0$ V rather than the typical 0.1 to 0.3 V level.

Transistor switching circuit is shown in Fig. 9.3(a). Summing up voltages around the input loop gives

$$I_B R_B + V_{BE} - V_{BB} = 0$$

or base current,

$$I_B = \frac{V_{BB} - V_{BE}}{R_B} \tag{9.1}$$

For a silicon transistor V_{BE} is typically 0.7 V.

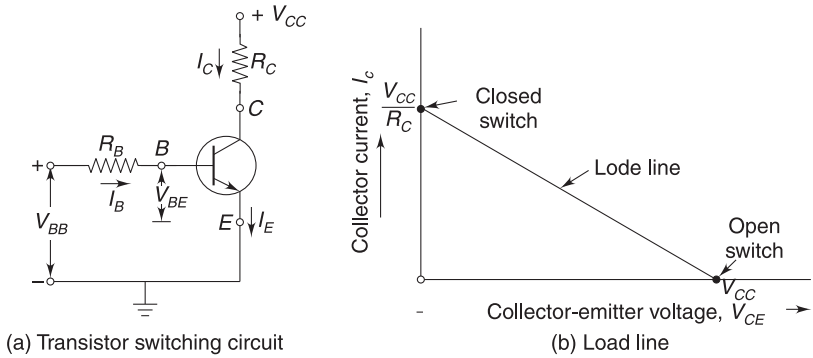


Fig. 9.3

If the base current is larger than or equal to $I_{B\,(sat)}$, the operating point Q is at the upper end of the load line (Fig. 9.3(b)). In this case, the transistor appears like a closed switch. On the other hand, if the base current is zero, the transistor operates at the lower end of the load line and the transistor appears like an open switch.

In Fig. 9.4, we see a transistor switching circuit that is being driven by a voltage step. This is the kind of waveform we get in a digital computer. When the input voltage is zero, the transistor is cut-off, it appears like an open switch. With no current through the collector resistor, the output voltage equals V_{CC} (+ 10 V).

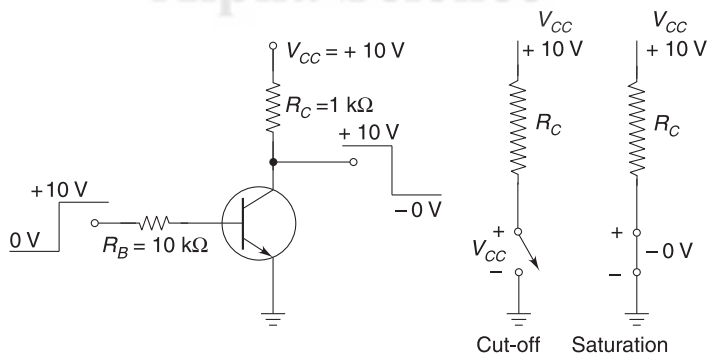


Fig. 9.4

When the input voltage is 10 V, the base current is

$$I_B = \frac{V_{\text{in}} - V_{BE}}{R_B} = \frac{10 - 0.7}{10\ \text{k}\Omega} = 0.93\ \text{mA}$$

Visualizing the transistor shorted between the collector and emitter, the output voltage ideally drops to zero, and the saturation current is

$$I_{C\,(\text{sat})} = \frac{V_{CC}}{R_C} = \frac{10}{1\ \text{k}\Omega} = 10\ \text{mA}$$

If the h_{fe} of the transistor is 100 and $I_B = 0.93$ mA, the collector current will attempt to be $I_C = h_{fe}\ I_B = 100 \times 0.93\ \text{mA} = 93\ \text{mA}$.

Obviously, the collector current can never attain this value as it is limited by the collector resistance R_C of 1 kΩ. The conclusion to be drawn is that when input voltage is 10 V, the collector current will be the maximum and equal to saturation current $I_{C\,(\text{sat})}$ and V_{CE} will be approximately zero.

Thus, we see that when input voltage is zero, output voltage V_{CE} is maximum and equals V_{CC} and when input voltage is maximum, the output voltage V_{CE} is minimum and is almost zero. It means output is inverse of the input and the transistor switch acts like an *inverter*.

The power loss in the on-state is quite low because $V_{CE\,(\text{sat})}$ is very small and power loss is given as

$$\text{Power loss} = \text{Output voltage} \times \text{output current}$$

$$= V_{CE\,(\text{sat})} \times I_{C\,(\text{sat})}$$

Thus, the transistor efficiency is quite high in *on-state*.

The power loss in the *off-state* equals collector supply voltage multiplied by reverse leakage current I_{CEO}. Since I_{CEO} is negligibly small as compared to full-load current that flows in the *on-state*, power loss in the transistor is quite small in the *off-state*. Thus the efficiency of the transistor is high in *off-state* too.

9.3 FIXED CURRENT BIAS TRANSISTOR SWITCHING CIRCUIT

The fixed current bias circuit for transistor switching is shown in Fig. 9.5. In this case, the transistor is biased into saturation. Unlike amplifier circuits, the bias conditions of properly designed switching circuits do not normally change with different transistor h_{fe} values. The switching transistor is either *on* or *off*. Although too unpredictable for biasing amplifier circuits, fixed current bias is quite satisfactory for switching circuits.

For the circuit shown in Fig. 9.5, base current I_B is determined using the specified minimum value of h_{fe} and collector current I_C level required for saturation. Base resistance R_B is then determined from the collector supply voltage V_{CC} and base current I_B levels. This means that all h_{fe} values higher than $h_{fe\,(\text{min})}$ will give collector currents larger than that required for saturation. Thus, the transistor is always saturated regardless of the value of h_{fe}.

The negative-going input pulse, applied to the transistor base via capacitor C_C, drives the base below the emitter voltage level to switch the device *off*. When

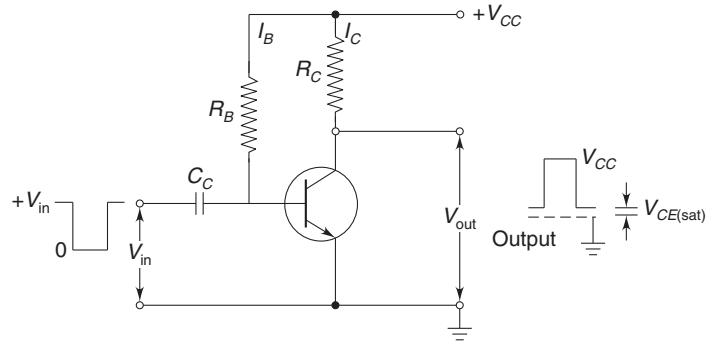


Fig. 9.5 *Fixed current bias transistor switching circuit*

the transistor goes *off*, collector current I_C ceases to flow and output voltage V_{out} goes from $V_{CE\ (sat)}$ to approximately V_{CC}.

9.4 MULTIVIBRATORS

A multivibrator is basically a two stage *R-C* coupled amplifier with positive feedback from the output of one amplifier to the input of the other, as illustrated in Fig. 9.6. Multivibrator is a switching circuit and may be defined as an electronic circuit that generates nonsinusoidal waves such as rectangular waves, sawtooth waves, square waves etc. Multivibrators are capable of storing binary numbers, counting pulses, synchronizing arithmetic operations and performing other essential functions used in digital systems.

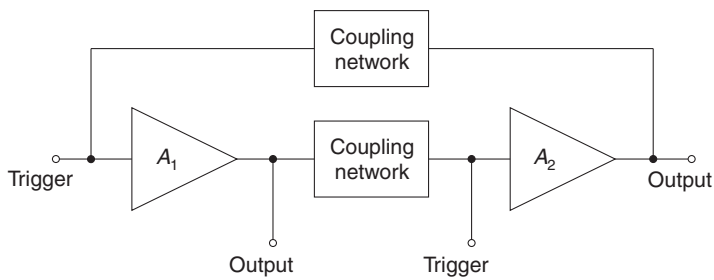


Fig. 9.6 *Basic configuration of a multivibrator*

Ordinarily, square waves can be generated by introducing a switch in a dc circuit, as shown in Fig. 9.7(a). Square wave will be generated by turning the switch *on* and *off* at regular intervals. Hence a multivibrator is essentially a switching circuit.

The circuit operates in two states *viz.*, *on* and *off* controlled by circuit condition. The operation of the circuit is such that when on amplifier is cut-off (*off*) the positive feedback loop maintains the other amplifier in a conducting

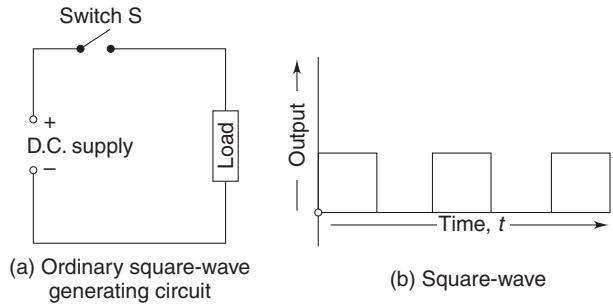


Fig. 9.7

or *on state*. When a trigger causes one amplifier to change *state*, the coupling network acts to change the state of the second amplifier. The outputs are opposite in sense, one indicating a transition from an "*off*" state to an *on* state and the other the opposite transistion. The condition in which the multivibrator may remain indefinitely until the circuit is triggered by some external signal is termed as the *stable state*.

There are only two possible states of a multivibrator and are as follows:

First state: Transistor Q_1 *on* and transistor Q_2 *off*.

Second state: Transistor Q_1 *off* and transistor Q_2 *on*.

Depending upon the type of coupling network employed, the multivibrators are classified into the following three categories:

1. Astable or free running multivibrator.
2. Monostable or single-shot multivibrator.
3. Bistable or flip-flop multivibrator.

The first one is the *non-driven* type whereas the other two are the *driven type* (also called the *triggered oscillators*).

Multivibrators are used for various purposes such as generation of nonsinusoidal waveforms (square, rectangular, sawtooth etc.) and pulses occurring periodically, frequency division, synchronized generation of pulses and extended waveforms, generation of time delays, storage of binary bit of information etc.

1. **Astable or free running multivibrator** In an astable multivibrator, both coupling networks provide ac coupling through coupling capacitors. Each amplifier stage provides phase shift of 180° in the midband so as to provide an overall phase shift of 360° or 0° and thus a positive feedback. The circuit, therefore, oscillates, provided that the total loop gain is equal to or greater than unity. It has no stable state. The two states of operation of astable multivibrators are quasi-stable (temporary) states. The astable multivibrator, therefore, makes successive transistions from one quasi-state

to the other one after a predetermined time interval, without the aid of an external triggering signal. The periodic time depends upon circuit time constants and parameters. Thus it is just an oscillator as it does not need any external pulse for its operation. Since its output oscillates in between *off* and *on* states freely, it is called a *free running* multivibrator. It is also named as *square-wave generator* based on its application. AMV may also be used as a *synchronized oscillator* and for driving sweep generators.

2. **Monostable or single-shot multivibrator** It is also called a *single swing*, or *delay multivibrator*. In this circuit, one coupling network provides a.c. coupling while the other provides d.c. coupling. It has only one stable (standby) state and one quasi stable state. In this circuit, a triggering signal is required to induce a transition from stable to quasi stable state. The circuit remains in the quasi stable state for a period determined by the circuit components. After this period, the circuit returns to its initial stable state and no external signal is required to induce this reverse transition. Thus, the monostable multivibrator supplies a single output pulse of a desired time duration for every input trigger pulse.

   Since the circuit vibrates once for a trigger, it is called as *univibrator* or *monostable multivibrator*. As only one triggering signal is required to induce a transition from a stable state to quasi stable state and the circuit returns to its initial state automatically after a definite period, it is called *single-shot multivibrator*.

3. **Bistable or flip-flop multivibrator** The bistable multivibrator, also called a *two-shot multivibrator*, requires application of two triggers to return the circuit to its original state. The first trigger causes the conducting transistor to be cut-off, and the second trigger causes a transistor back to the conducting state. Because two triggers are required, bistable circuits are sometimes called *flip-flops*. Obviously, such a circuit does not oscillate. In this circuit, both coupling networks provide d.c. coupling and *no energy storage element* is used.

A bistable multivibrator is employed for performing many digital operations such as counting and storing of binary informations. This circuit also finds extensive use in the generation and processing of pulse type waveforms.

Input-output relations for the three types of multivibrators are shown in Fig. 9.8.

9.5 ASTABLE MULTIVIBRATORS

The astable or free running multivibrator also called the *collector-coupled astable multivibrator*, is widely employed for the generation of square waves or pulses. A typical astable multivibrator is shown in Fig. 9.9. It is essentially a two-stage

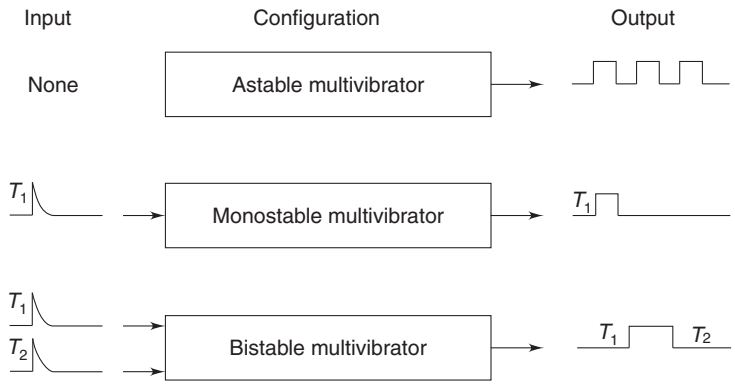


Fig. 9.8

R-C coupled amplifier with the output of first stage coupled to the input of second stage; and the output of second stage coupled to the input of the first stage. As the phase of a signal is reversed when amplified by a single stage common emitter amplifier, it comes back to its original phase when passed through two stages. Thus, the signal feedback to the base of either transistor is in the same phase as the original signal at its output. It amounts to a positive feedback. The amount of feedback is so large that the transistors are driven between cut-off and saturation almost instantaneously. A transistor remains in either saturation or cut-off for a period depending on the time constant of the base-circuit elements.

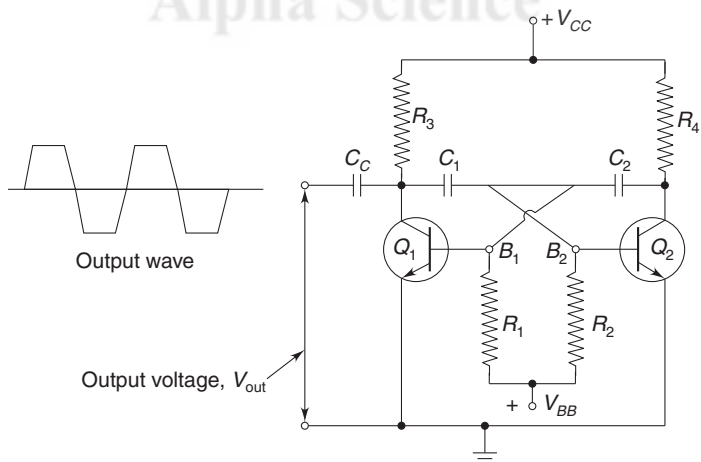


Fig. 9.9 *Collector-coupled astable multivibrator*

Operation The operation of the multivibrator circuit illustrated in Fig. 9.9 is as follows:

When power is first applied to the circuit, both transistors start conducting. Because of small differences in their operating characteristics, one of the transistors will conduct slightly more than the other. This starts a series of events. Assume arbitrarily that Q_1 initially conducts more than Q_2. This causes the collector voltage of Q_1 to drop more rapidly than that of Q_2. The resulting negative signal is fed to the base of Q_2 through capacitor C_1 and drives it towards cut-off. As a result the collector voltage of Q_2 rises towards V_{CC}.

This change in collector voltage of Q_2 is fed to the base of Q_1 through capacitor C_2. It causes transistor Q_1 to go into saturation. This happens so quickly that capacitor C_1 does not get a chance to discharge, and the reduced voltage at the collector of Q_1 appears across resistor R_2.

The switching action now begins, capacitor C_1 begins to discharge exponentially through R_2. When the voltage on C_1 reaches zero volt, C_1 attempts to charge up to the value of + V_{BB}, the base supply voltage. But this action immediately places a forward bias on Q_2 and thus transistor Q_2 begins to conduct. As soon as Q_2 starts conducting, its collector current causes collector voltage to fall. This negative-going voltage signal is fed to the base of transistor Q_1 through capacitor C_2 which begins to conduct less i.e., it comes out of saturation. This resulting increased voltage at the collector of Q_1 is coupled through C_1 and appears across R_2. The collector current of Q_2, therefore, increases. This process continues rapidly until transistor Q_1 finally cuts-off and transistor Q_2 conducts heavily. At this instant, the collector voltage of transistor Q_1 attains its maximum value of V_{CC}, capacitor C_1 charges to the full value of V_{CC}, and a full cycle of operation has been completed. The cycle then repeats itself. The output of the multivibrator can be taken from the collector of either transistor. The output is a squarewave, as shown in Fig. 9.10 and whose frequency is determined by R_1, R_2, C_1 and C_2 in the circuit.

Frequency of Oscillation From the above discussion it is found that as soon as astable multivibrator is switched *on*, transistor Q_1 goes *on* while transistor Q_2 goes *off*. Let this instant be $t = 0$.

At time = 0 for transistor Q_1 (in On-state)

Collector voltage,

$$v_{C_1} = V_{CE\,(\text{sat})} = 0 \tag{9.2}$$

Collector current,

$$i_{C_1} = I_{C_{1\,(\text{sat})}} = \frac{V_{CC} - V_{CE\,(\text{sat})}}{R_3}$$

$$= \frac{V_{CC}}{R_3} \tag{9.3}$$

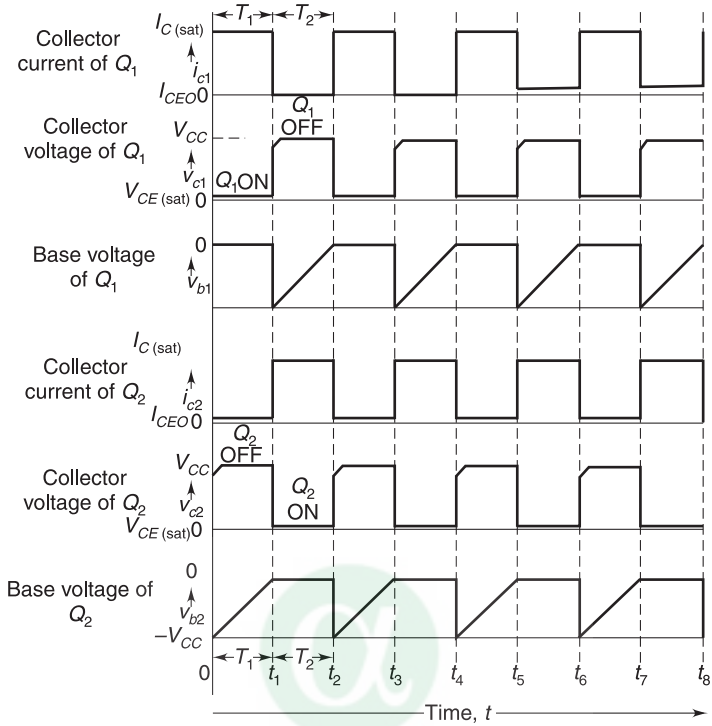


Fig. 9.10 *Waveforms for astable multivibrator*

Base voltage,

$$V_{B_{1\,(on)}} = 0 \tag{9.4}$$

Base current,

$$I_{B_{1\,(in)}} = \frac{V_{BB} - V_{BE\,(sat)}}{R_1}$$

$$\cong \frac{V_{BB}}{R_1} \tag{9.5}$$

For transistor Q_2 (in off-state)

Collector voltage,

$$v_{C_2} = V_{CC} - V_{CE\,(sat)}$$

$$\cong V_{CC} \tag{9.6}$$

Collector current,

$$i_{C_2} = I_{CEO} = 0 \tag{9.7}$$

At the instant of switching on the power supply but before the commencement of regenerative feedback sequence, the voltage on capacitor C_1 is V_{CC}. This implies that the collector side voltage of the capacitor C_1 is V_{CC} and base side voltage is $V_{B_{2\,(on)}}$. At the end of the regenerative cycle, the collector side voltage of the capacitor C_1 has decreased from V_{CC} to $V_{CE\,(sat)}$ (almost zero). But the capacitor C_1 cannot discharge instantaneously. Thus the voltage at the base of the transistor Q_2 is given as

$$V_{\substack{B_{2\,(off)} \\ (t=0)}} = V_{B_{2\,(on)}} - V_{CC} \tag{9.8}$$

After commencement of regenerative feedback but before Q_1 goes off (i.e., $0 < t < T_1$), voltage across capacitor C_1 increases from $V_{B_{2\,(off)}}$ $(t = 0)$ toward V_{BB}. Changing path is provided by capacitor C_1 and resistor R_2 assuming that $V_{C_1} = 0$. At any instant t, voltage V_{B_2} is given as

$$v_{B_{2\,(t)}} = V_{BB} + \left[V_{\substack{B_{2\,(off)} \\ (t=0)}} - V_{BB} \right] e^{\frac{-t}{R_2 C_1}} \tag{9.9}$$

Thus, the base voltage of transistor Q_2, $V_{B_{2\,(t)}}$ rises exponentially with time constant $C_1 R_2$. However, as soon as V_{B_2} becomes equal to $V_{B_{2\,(on)}}$ at time T_1, the transistor Q_2 commences conducting and $V_{B_{2\,(t)}}$ remains constant at $V_{B_{2\,(on)}}$.

At time T_1

$$V_{B_{2\,(on)}} = V_{BB} + \left[V_{\substack{B_{2\,(off)} \\ (t=0)}} - V_{BB} \right] e^{\frac{-T_1}{R_2 C_1}} \tag{9.10}$$

Solving Eqs. (9.8) and (9.10), we get

$$T_1 = R_2\, C_1 \log_e \left[\frac{V_{BB} + V_{CC} - V_{B_{2\,(on)}}}{V_{BB} - V_{B_{2\,(on)}}} \right] \tag{9.11}$$

If $V_{BB} >> V_{B_{2\,(on)}}$, Eq. (9.11) becomes

$$T_1 = R_2\, C_1 \log_e \left[1 + \frac{V_{CC}}{V_{BB}} \right] \tag{9.12}$$

and if $V_{BB} = V_{CC}$, then Eq. (9.12) reduces to

$$T_1 = R_2\, C_1 \log_e 2 = 0.693\, R_2\, C_1 \tag{9.13}$$

From Eq. (9.13) the *on-time* of transistor Q_1 or the *off-time* of transistor Q_2 is given as

$$T_1 = 0.693\, R_2\, C_1$$

Similarly, the *off-time* of transistor Q_1 or the *on-time* of transistor Q_2 is given as

$$T_2 = 0.693\ R_1\ C_2$$

Hence, total time period of square wave

$$T = T_1 + T_2 = 0.693\ (R_2\ C_1 + R_1\ C_2) \tag{9.14}$$

For a symmetrical multivibrator $R_1 = R_2 = R$ (say) and $C_1 = C_2 = C$ (say) then total time period is given as

$$T = 2 \times 0.693\ \text{RC} = 1.386\ \text{RC} \tag{9.15}$$

Frequency of the square wave,

$$f = \frac{1}{T} = \frac{1}{1.386\ \text{RC}} = \frac{0.72}{\text{RC}} \tag{9.16}$$

9.6 MONOSTABLE MULTIVIBRATORS

As already mentioned in Section 9.4, monostable multivibrator is a two-stage amplifier with two states—one stable state and another quasi stable state. The circuit remains in the quasi-stable state for a period determined by the circuit components and then returns to its initial stable state. In the stable state one transistor is in the *on* or *conducting state* while the other is in *off* or *non-conducting state*.

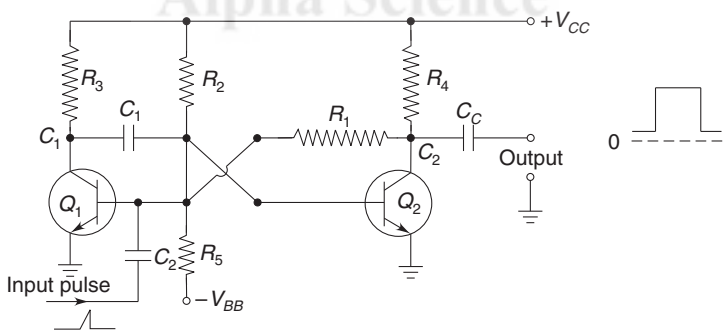


Fig. 9.11

The circuit arrangement of a transistor monostable multivibrator is shown in Fig. 9.11. It consists of two similar transistors Q_1 and Q_2 with equal collector resistances R_3 and R_4 (i.e. $R_3 = R_4$). The circuit is designed for operation with transistor Q_2 normally conducting and transistor Q_1 cut-off. The output of transistor Q_1 is coupled to the base of transistor Q_2 via capacitor C_1 as in case of astable multivibrator. However, the other coupling is direct one through base resistor R_1.

Transistor Q_2 is forward biased by power supply V_{CC} and base resistor R_2 and remains saturated in stable state. Operation in this stable state is ensured if the biasing resistor R_2 is selected to permit sufficient base current in transistor Q_2 for operation in saturation region. Transistor Q_1 is reverse biased by power supply V_{BB} and resistor R_5 remains cut-off in the stable state. The input pulse is applied through capacitor C_2. The output may be taken from either of the transistors Q_1 or Q_2.

Operation The operation of the circuit is explained below.

When the power supply is switched on but with no input pulse applied transistor Q_2 is *on* and is operating in its saturation region. It is being forward biased by supply V_{CC} and base resistor R_2. The collector of transistor Q_2 is virtually at ground potential. The base of transistor Q_2 is 0.7 V above ground

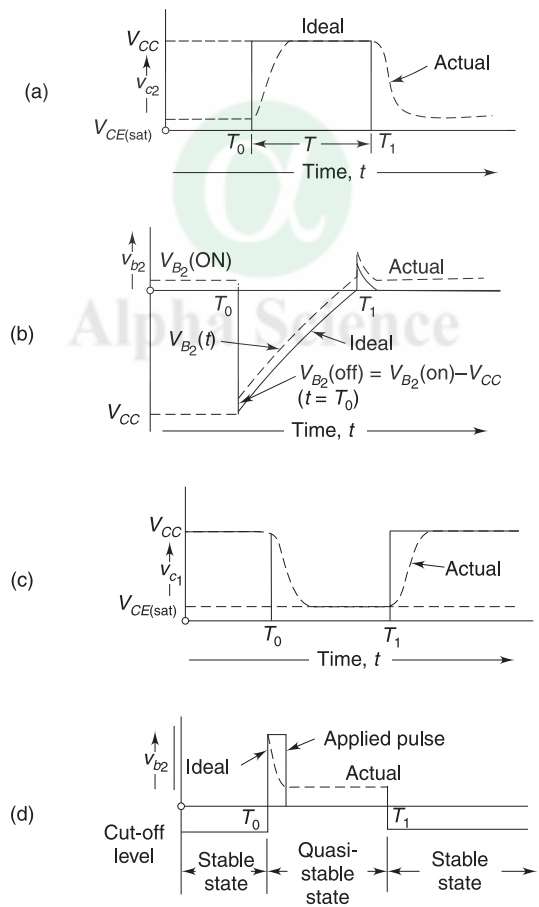


Fig. 9.12 *Waveforms for monostable multivibrator*

potential because of the forward bias of base-emitter junction. This is illustrated in Figs. 9.12(a) and (b).

Transistor Q_1 is cut-off being reverse biased by supply V_{BB} and resistor R_5. Its base resistor R_1 is connected to the collector of transistor Q_2, which is at zero potential. Thus collector resistor R_3 is completely disconnected from the grounded emitter of transistor Q_1 and is free to carry current to charge capacitor C_1. Since capacitor C_1 is connected to the base of transistor Q_2, which is close to ground potential, it will charge up almost to the supply voltage V_{CC}. The polarity of the charge on capacitor C_1 is plus (+) on the left and minus (–) on the right, as shown.

When a positive trigger pulse of short duration and sufficient magnitude is applied to the base of transistor Q_1 through capacitor C_2, it (input pulse) overrides the reverse bias of emitter-base junction of transistor Q_1 and gives it a forward bias. Thus transistor Q_1 starts conducting and the potential of collector of transistor Q_1 comes down to ground. Since charge on capacitor C_1 cannot disappear instantly, the voltage across the capacitor plates is maintained. With the positive side of the capacitor C_1 pulled down to zero volt by transistor Q_1, the negative side goes to a voltage far below ground potential. Thus a negative bias is applied to the base of transistor Q_2 and transistor Q_2 is cut-off. The collector of transistor Q_2 rises toward V_{CC} and is now capable of supplying base current to transistor Q_1 through its base resistor R_1. Thus transistor Q_1 remains turned *on* even after the positive spike from the transistor Q_1 is removed.

At time passes charging current flows onto the plates of capacitor C_1. The flow path is down through R_2, through capacitor C_1 and through collector to emitter of Q_1 into ground. As can be seen, this path seeks to charge capacitor C_1 to the opposite polarity, what happens is that the voltage across capacitor C_1 gets reduced. When the voltage across the capacitor crosses through zero and attains 0.7 V in the opposite polarity, as shown in Fig. 9.12(b), it bleeds a small amount of current into the base of transistor Q_2. This small base current causes collector current to flow in transistor Q_2, lowering its collector voltage. The reduced collector voltage causes a reduction in base current to transistor Q_1. This in turn causes a reduction in collector current of Q_1. The Q_1 collector voltage rises slightly, thereby raising the base of Q_2 higher yet. This action is regenerative; once it begins, it avalanches. In the end, transistor Q_2 is saturated once again and transistor Q_1 is cut-off. Thus the circuit reverts back to its original stable state. The circuit remains in this state until another triggering pulse causes the circuit to switch over the state and the other cycle repeats itself.

The width of the output pulse is determined by the time constant of C_1 R_2. The multivibrator generates one output pulse for every input trigger pulse

and that is why it is sometimes called the *one-shot multivibrator*. The width or duration of the pulse is given as $T = 0.693\ C_1\ R_2$.*

Applications

1. The falling part of the monostable multivibrator output is often used for triggering another pulse generator circuit, thus producing a pulse delayed by a time T with respect to the input pulse.
2. Monostable multivibrator is also employed for regenerating or rejuvenating old and wornout pulses. Various pulses used in computers and telecommunication systems become somewhat distorted during use. A monostable multivibrator can be employed to generate new, clean and sharp pulses from these distorted (used) pulses.

9.7 BISTABLE MULTIVIBRATOR

A bistable multivibrator, as its name implies, has two stable states. It can stay in either of the two states indefinitely (as long as d.c. power supply remains on). Only an application of a suitable trigger pulse makes the circuit to change to the other stable state. The circuit can be brought back to its original stable state only by applying another suitable trigger pulse. Thus one trigger pulse causes the multivibrator to *flip* from one state to another and the next pulse causes it to *flop* back to its original state. This is the very reason that it is called the *flip-flop* multivibrator.

The bistable multivibrator is employed in many digital operations such as counting and storing of binary information, as a frequency divider in timing circuits and can also be used for generation of pulses.

The circuit of a bistable multivibrator is shown in Fig. 9.13. It consists of two identical transistors Q_1 and Q_2 with equal collector resistances R_3 and R_4 and with output of one supplied to the input of the other. The feedback is coupled through resistors R_1 and R_2 shunted by capacitors C_1 and C_2 respectively. These capacitors are known as the *commutating capacitors* and are provided to improve

---

* From Fig. 9.12(b)

$$V_{B_2(t)} = V_{CC} + \left[V_{\underset{(t=T0)}{B_2(\text{off})}} - V_{CC}\right] e^{-(t-T_0)/R_2 C_1} = V_{CC} + \left[V_{B_2(\text{on})} - 2V_{CC}\right] e^{-(t-T_0)/R_2 C_1}$$

$\because \quad V_{B_2(\text{off})(t=T0)} = V_{B_2(\text{on})} - V_{CC}$

Now when $t = T_0 + T$

$$V_{B_2(T0+T)} = V_{B_2(\text{on})} = V_{CC} + \left[V_{B_2(\text{on})} - 2V_{CC}\right] e^{-T/R_2 C_1}$$

or $$T = R_2\ C_1 \log_e \frac{2V_{CC} - V_{B_2(\text{on})}}{V_{CC} - V_{B_2(\text{on})}} = R_2\ C_2 \log_e 2 = 0.693\ C_1\ R_2$$

if $V_{B_2(\text{on})}$ is negligible compared with V_{CC}.

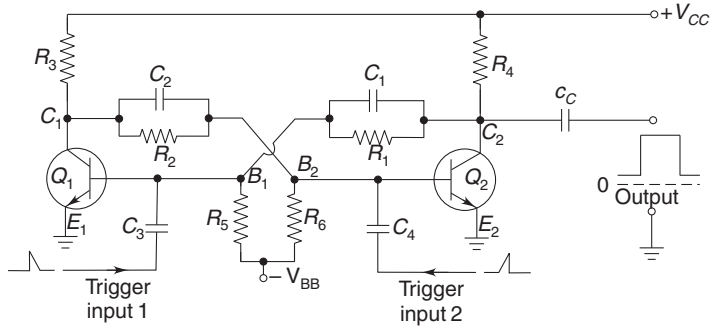


Fig. 9.13 *Bistable multivibrator circuit*

the switching characteristics of the circuit by allowing flow of high frequency components of square wave through them. Thus transition time (rise and fall times) is reduced and distortion less square wave output is obtained. The output may be taken from either of the two transistors Q_1 or Q_2.

Working When the power supply V_{CC} is switched *on* to the transistors, one transistor starts conducting slightly more than the other due to small differences in the characteristics of the transistors. This drives one transistor (say Q_1) to the saturation and the other transistor Q_2 to the cut-off in a manner explained in case of astable multivibrator. This is the *first stable state* of the multivibrator. The multivibrator cab be driven from the *first stable state* (Q_1 *on* and Q_2 *off*) to the other stable state (Q_1 off and Q_2 *on*) by applying either a negative trigger pulse to the base of transistor Q_1 or a positive trigger pulse to the base of transistor Q_2.

Let a negative pulse of short duration and sufficient magnitude be applied to the base of transistor Q_1 through capacitive C_3. This reduces the forward bias on transistor Q_1 and causes a reduction in collector current and, thereby, increase in potential of its collector terminal C_1. The rising collector voltage appears across the emitter-base junction of transistor Q_2 as it is connected to collector terminal C_1 via $R_2 - C_2$ parallel circuit. As a result collector current of transistor Q_2 increases and, therefore, its collector voltage falls. The decreasing collector voltage appears across the emitter-base junction of transistor Q_1 where it further reverse biases the emitter-base junction of transistor Q_1 to make collector current to fall. After few cycles, the transistor Q_2 is driven into saturation and transistor Q_1 to cut-off. This is the *second stable state* of the multivibrator. The circuit will now remain in this second stable state (Q_1 *off* and Q_2 on) until a positive pulse is applied to the base of transistor Q_1 or a negative pulse to the base of transistor Q_2.

The noteworthy points are:

1. Under cut-off condition nearly the full supply voltage V_{CC} appears across the transistor and this restricts the supply voltage V_{CC} to the order of several tens of volts (smaller than the transistor breakdown voltage V_{CE}).

2. Under saturation condition, the collector current I_C is maximum. Hence R_3 and R_4 must be chosen so as to restrict the collector current to the permissible value. The values of resistors R_1, R_2, R_5 and R_6, and V_{BB} must be so selected that in one state the base current is large enough to drive the transistor into saturation while in the second state the emitter-base junction must be below cut-off value.
3. The values of R_1, R_2, R_5 and R_6 are selected so that they are much greater than collector resistances R_3 and R_4 and consequently have no loading effects on the amplifier circuit.

If transistor Q_1 is in conducting state and transistor Q_2 is in non-conducting state, the voltages V_{B1} between point B_1 and ground (points E_1 and E_2) and V_{C1} between point C_1 and ground are both small. The base-bias voltage V_{B2} on transistor Q_2 is then

$$V_{B2} = - V_{BB} \frac{R_2}{R_2 + R_6} \tag{9.17}$$

which being negative, keeps transistor Q_2 in its cut-off state. Since R_4 is much smaller than R_1 and the collector current of transistor Q_2 is assumed zero, the collector voltage of transistor Q_2 is $+ V_{CC}$. The base current I_{B1} is then

$$I_{B1} = \frac{V_{CC} - V_{B1}}{R_1} - \frac{V_{BB} - V_{B1}}{R_5} \tag{9.18}$$

and if V_{B1} can be considered negligibly small,

$$I_{B1} = \frac{V_{CC}}{R_1} - \frac{V_{BB}}{R_5} \tag{9.19}$$

For keeping transistor Q_1 conducting, the value of I_{B1}, must be sufficiently positive to ensure continued current i.e.,

$$I_{B1} \geq \frac{I_{C1}}{h_{FE}} \tag{9.20}$$

Since

$$I_{C1} = \frac{V_{CC} - V_{CE\,(\text{sat})}}{R_3} \tag{9.21}$$

$$I_{B1} \geq \frac{V_{CC} - V_{CE\,(\text{sat})}}{h_{FE}\, R_3} \tag{9.22}$$

The values of V_{CC}, V_{BB}, R_1, R_2, R_3 and R_4 are so selected that Eq. (9.22) is satisfied and consequently a stable state exists with Q_2 cut-off and Q_1 in saturation. For the symmetry of circuit ($R_3 = R_4$; $R_1 = R_2$ and $R_5 = R_6$), it is evident that a second stable state exists with Q_1 cut-off and Q_2 in saturation.

Two points should be noted about the switching process. First the pulse used for switching need only be applied for a time sufficient for the changeover; it can then be removed. Second, if a negative pulse is used to affect switching, a positive pulse is required for the next switching, a negative pulse for the one following etc.

Idealized output waveforms for the bistable multivibrator are shown in Fig. 9.14 as solid curves. The dotted waveforms are those of an actual circuit, the deviations from the idealized result from switching transients.

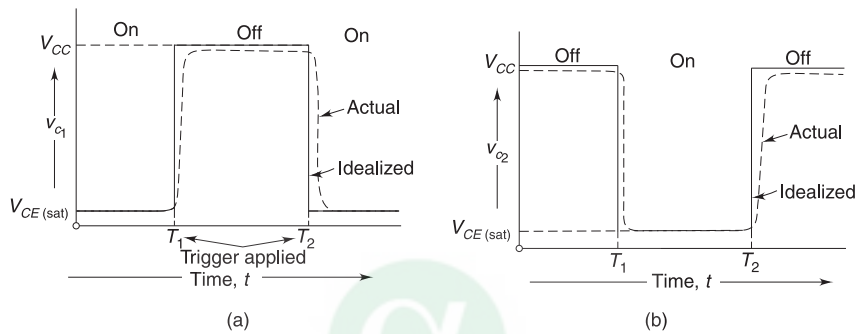


Fig. 9.14

Questions

9.1 Explain the distinguishing features of astable, monostable and bistable multivibrators and give the operational details of any one of them.

9.2 Draw the circuit of an astable multivibrator using *PNP* transistors, where output is a square wave with steep edges. Also draw the labeled waveforms at the collectors and bases of two transistors. Determine the frequency of the output waveform.

9.3 Describe with circuit diagram and waveforms working of an astable multivibrator.

9.4 Describe the operation of a monostable multivibrator.

9.5 Explain the working of a bistable multivibrator with the help of a neat circuit diagram and waveforms.

9.6 Write short notes on:

(i) Astable multivibrator

(ii) Bistable multivibrator

10 Three Terminal Devices

10.1 INTRODUCTION

In this chapter we will consider basically three terminal devices *viz*., UJT (Uni Junction Transistor) and FETs (Field Effect Transistors). We will first consider UJT. It is a triggering device that may be used to fire thyristors (i.e. solid state devices having two or more *P-N* junction which can be switched from a non conducting to a conducting state). (A thyristor is the solid state counterpart of a thyratron). The UJT has a negative resistance region (as we shall see) which may be used to advantage in a relaxation oscillator whose frequency may easily be varied to provide an excellent method of motor speed control.

The FETs may be thought of as the solid state counterpart of a pentode, in that, their characteristics are very similar. There are two types of FET as shown in Fig. 10.5. The first was described in 1952 by Shockley, but not made available until ten years later because of fabrication problems. It is the junction field effect transistor (JFET), which is a three terminal device and is the one, which we shall discuss after UJT. The second type of FET is the insulated gate field effect transistor (IGFET), more commonly known as the metal-oxide semiconductor (MOS) transistor or MOSFET. The IGFET is a four terminal device, yet, it is used basically as a three terminal device in most applications (Its fabrication is very closely related to integrated circuit technology).

10.2 UNIJUCTION TRANSISTOR (UJT)

Unijunction transistor (abbreviated as UJT), also called the *double-base diode*, is a 2 layer, 3 terminal solid-state (silicon) switching device. The device has a unique characteristic that when it is triggered, its emitter current increases regeneratively (due to negative resistance characteristic) until it is restricted by emitter power supply. The low cost per unit, combined with its unique characteristics, have warranted its use in a wide variety of applications. A few include oscillators, pulse generators, sawtooth generators, trigger circuits, phase control, timing circuits, and voltage- or current-regulated supplies. The device is, in general, a low-power absorbing device under normal operating conditions and provides tremendous aid in the continual effort to design relatively efficient systems.

10.2.1 Construction

The basic structure of a unijunction transistor is shown in Fig. 10.1(a). It essentially consists of a light-doped *N*-type silicon bar with a small piece of heavily doped *P*-type material alloyed to its one side to produce single *P-N* junction.

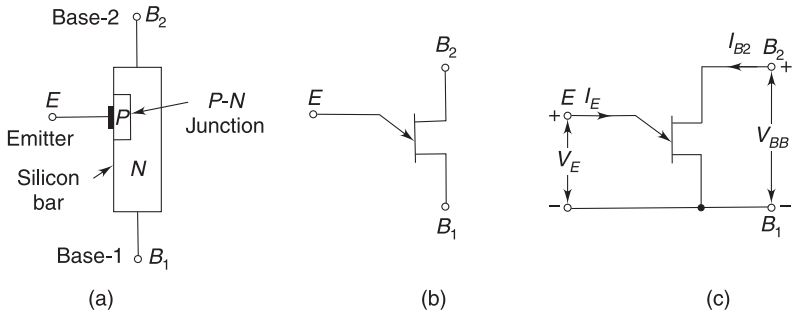


Fig. 10.1 *Unijunction transistor*

The single *P-N* junction accounts for the terminology *unijunction*. The silicon bar, at its ends, has two ohmic contacts designated as base-1 (B_1) and base-2 (B_2), as shown and the *P*-type region is termed the emitter (E). The emitter junction is usually located closer to base-2 (B_2) than base-1 (B_1) so that the device is not symmetrical, because symmetrical unit does not provide optimum electrical characteristics for most of the applications.

The symbol for unijunction transistor is shown in Fig. 10.1(b). The emitter leg is drawn at an angle to the vertical line representing the *N*-type material slab and the arrowhead points in the direction of *conventional current* when the device is forward biased, active or in the conducting state. The basic arrangement for the UJT is shown in Fig. 10.1(c).

A complementary UJT is formed by diffusing an *N*-type emitter terminal on a *P*-type base. Except for the polarities of voltage and current, the characteristics of a complementary UJT are exactly the same as those of a conventional UJT.

The noteworthy points about UJT are given below:

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

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

10.2.2 Equivalent Circuit

The equivalent circuit of a UJT is shown in Fig. 10.2. The *P-N* junction is represented in the emitter by a diode.

Silicon bar being lightly doped has a high resistance and can be represented as two resistors connected in series – R_{B_1} from base-1 to point *A* and R_{B_2} from base-2 to point *A*, as shown in Fig. 10.2. The resistance R_{B_1} is shown as a variable resistor since its magnitude varies with the emitter current I_E. In fact, for a typical UJT, R_{B_1} may vary from 5 kΩ down to 50 Ω for a corresponding change of emitter current I_E from 0 to 50 μA. The total resistance of silicon bar from one end to the other end, i.e., the resistance between B_2 and B_1 with emitter terminal *E* open is called the *interbase resistance* R_{BB}. From the equivalent circuit shown in Fig. 10.2, it is obvious that

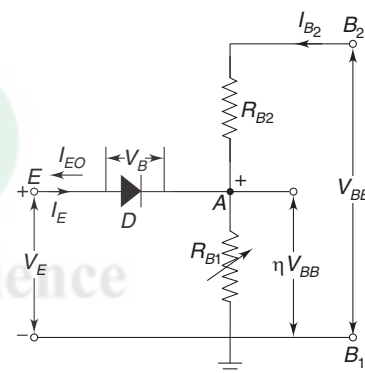


Fig. 10.2 *Equivalent circuit of a UJT*

$$R_{BB} = R_{B_1} + R_{B_2} \tag{10.1}$$

with emitter terminal *E* open

R_{BB} is typically within the range of 4 kΩ and 10 kΩ. Usually R_{B_1} is greater than R_{B_2} and R_{B_1} = 60 per cent of R_{BB}.

With a voltage V_{BB} applied between B_2 and B_1, as shown in the figure, the voltage at the junction of R_{B_1} and R_{B_2} (point *A*) is

$$V_A = V_{BB} \times \frac{R_{B_1}}{R_{B_1} + R_{B_2}} + \frac{R_{B_1}}{R_{BB}} V_{BB} \tag{10.2}$$

$$= \eta\ V_{BB} \tag{10.3}$$

The Greek letter η (eta) is called the *intrinsic stand-off* ratio of the device and is defined as

$$\eta = \frac{R_{B_1}}{R_{B_1} + R_{B_2}} \quad \text{with terminal } E \text{ open}$$

$$= \frac{R_{B_1}}{R_{BB}} \tag{10.4}$$

The intrinsic stand-off ratio is the property of a UJT and is always less than unity. The typical range of η is from 0.5 to 0.8.

Voltage V_A is also the voltage at the cathode of the diode representing the *P-N* junction. With the emitter terminal open circuited, the only current flowing is

$$I_{B_2} = \frac{V_{BB}}{R_{BB}} \tag{10.5}$$

If the emitter terminal is grounded, the *P-N* junction is reverse biased and a small *emitter reverse current* I_{E_0} flows.

10.2.3 Operation

Imagine that the emitter supply voltage is turned down to zero. Then the intrinsic stand-off voltage reverse biases the emitter diode, as mentioned above. If V_B is the barrier voltage of the emitter diode, then the total reverse bias voltage is $V_A + V_B = \eta\ V_{BB} + V_B$. For silicon $V_B = 0.7$ V.

Now let the emitter supply voltage V_E be slowly increased. When V_E becomes equal to $\eta\ V_{BB}$, I_{E_0} will be reduced to zero. With equal voltage levels on each side of the diode, neither reverse nor forward current will flow. When emitter supply voltage is further increased, the diode becomes forward biased as soon as it exceeds the total reverse bias voltage ($\eta\ V_{BB} + V_B$). This value of emitter voltage V_E is called the *peak point voltage* and is denoted by V_P. When $V_E = V_P$, emitter current I_E starts to flow through R_{B_1} to ground (i.e., B_1). This is the minimum current that is required to trigger the UJT. This is called the *peak point emitter current* and denoted by I_P. I_P is inversely proportional to the interbase voltage, V_{BB}. Now when the emitter diode starts conducting, charge carriers are injected into the R_{B_1} region of the bar. Since the resistance of a semiconductor material depends upon doping, the resistance of region R_{B_1} decreases rapidly due

to additional charge carriers (holes). With this decrease in resistance, the voltage drop across R_{B_1} also decreases, causes the emitter diode to be more heavily forward biased. This, in turn, results in larger forward current, and consequently more charge carriers are injected causing still further reduction in the resistance of the R_{B_1} region. Thus the emitter current goes on increasing until it is limited by the emitter power supply. Since V_A decreases with the increase in emitter current, the UJT is said to have *negative resistance characteristics*.

It is seen that the base-2 (B_2) is used only for applying external voltage V_{BB} across it. Terminals E and B_1 are the active terminals. UJT is usually triggered into conduction by applying a suitable positive pulse to the emitter. It can be turned off applying a negative trigger pulse.

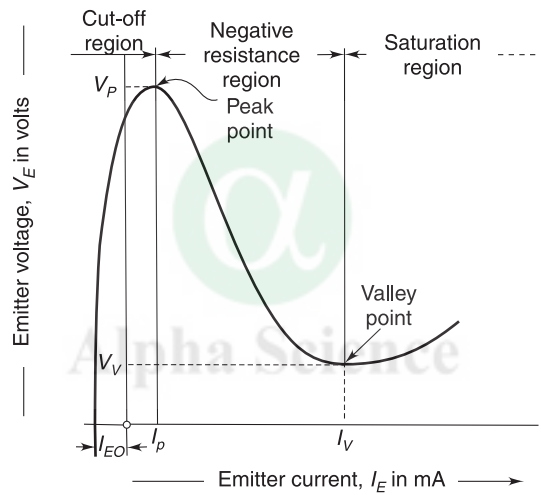


Fig. 10.3 *Static emitter-characteristic for a UJT*

10.2.4 Characteristics

The static emitter characteristic (a curve showing the relation between emitter voltage V_E and emitter current I_E) of a UJT at a given interbase voltage V_{BB} is shown in Fig. 10.3.* From Fig. 10.3, it is noted that for emitter potentials to the left of peak point, emitter current I_E never exceeds I_{EO} (measured in microamperes). The current I_{EO} corresponds very closely to the reverse leakage current I_{CO} of the conventional BJT. This region, as shown in the figure, is called the *cut-off* region. Once conduction is established at $V_E = V_P$ the emitter potential V_E starts decreasing with the increase in emitter current I_E. This corresponds

* *V-I* characteristics of UJT is one of the very few curves in which current *I* is taken on *X*-axis and voltage *V* is taken on *Y*-axis.

exactly to the decrease in resistance R_{B_1} for increasing current I_E. This device, therefore, has a *negative resistance region* which is stable enough to be used with a great deal of reliability in the areas of applications listed earlier. Eventually, the valley point reaches, and any further increase in emitter current I_E places the device in the saturation region, as shown in figure. In this region characteristic approaches that of a semiconductor diode in the equivalent circuit shown in Fig. 10.2. Three other important parameters for the UJT are I_P, V_V and I_V (shown in the Fig. 10.3) and are defined below:

Peak Point Emitter Current I_P It is the emitter current at the peak point. It represents the minimum current that is required to trigger the device (UJT). It is inversely proportional to the interbase voltage V_{BB}.

Valley Point Voltage V_V The valley point voltage is the emitter voltage at the valley point. The valley voltage increases with the increase in interbase voltage V_{BB}.

Valley Point Current I_V The valley point current is the emitter current at the valley point. It increases with the increase in interbase voltage V_{BB}.

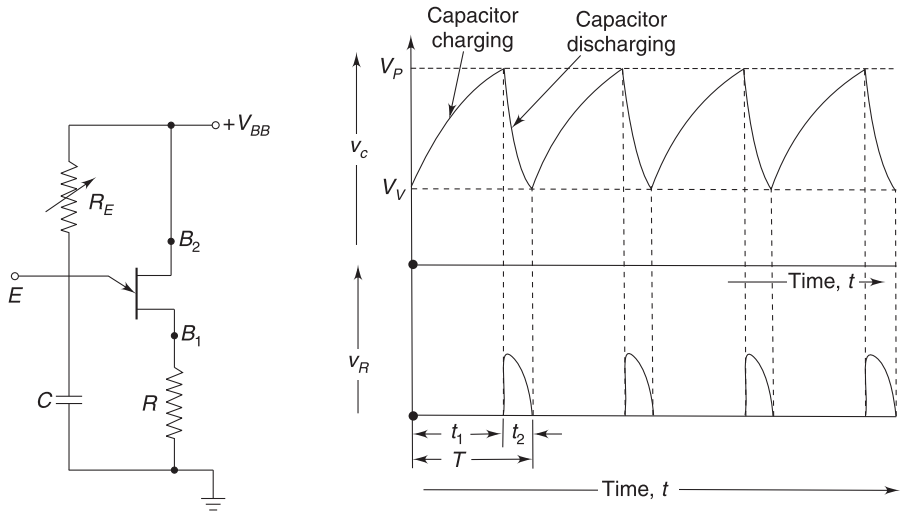
10.2.5 Special Features of UJT

The special features of a UJT are:

1. A stable triggering voltage (V_P)–a fixed fraction of applied interbase voltage V_{BB}.
2. A very low value of triggering current.
3. A high pulse current capability.
4. A negative resistance characteristic.
5. Low cost.

10.3 UJT RELAXATION OSCILLATOR

The relaxation oscillator shown in Fig. 10.4(a) consists of UJT and a capacitor C which is charged through resistor R_E when interbase voltage V_{BB} is switched *on*. During the charging period, the voltage across the capacitor increases exponentially until it attains the peak point voltage V_P. When the capacitor voltage attains voltage V_P, the UJT switches *on* and the capacitor C rapidly discharges via B_1. The resulting current through the external resistor R develops a voltage spike, as illustrated in Fig. 10.4(b) and the capacitor voltage drops to the value V_V. The device then cuts-off and the capacitor commences charging again. The cycle is repeated continually generating a sawtooth waveform across capacitor C. The resulting waveforms of capacitor voltage V_C and the voltage across resistor $R(V_R)$ are shown in Fig. 10.4(b). The frequency of the output sawtooth wave can be varied by varying the value of resistor R_E as it controls the time constant



(a) Basic circuit (b) Voltage waveforms across C and R

Fig. 10.4 *UJT relaxation oscillator*

($T = R_E C$) of the capacitor charging circuit. The discharge time t_2 is difficult to calculate because the UJT is in its negative resistance region and its resistance is continually changing. However, t_2 is normally very much less than t_1 and can be neglected for approximation.

For satisfactory operation of the above oscillator, the following two conditions for the turn-on and turn-off of the UJT must be met

$$R_E < \frac{V_{BB} - V_P}{I_P}$$

and

$$R_E > \frac{V_{BB} - V_V}{I_V} \tag{10.6}$$

i.e., the range of resistor R_E should be as given below

$$\frac{V_{BB} - V_P}{I_P} > R_E > \frac{V_{BB} - V_V}{I_V} \tag{10.7}$$

The time period and, therefore, frequency of oscillation can be derived as below.

During charging of capacitor, the voltage across the capacitor is given as

$$v_C = V_{BB}\left(1 - e^{-t/R_E C}\right)$$

where $R_E C$ is the time constant of the capacitor charging circuit and t is the time from the commencement of the charging.

The discharge of the capacitor commences at the end of charging period t_1, when the voltage across the capacitor v_C becomes equal to V_P i.e., ($\eta\ V_{BB} + V_B$)

$$V_P = \eta\ V_{BB} + V_B = V_{BB}\left(1 - e^{-t_1/R_EC}\right) \tag{10.8}$$

Neglecting V_B in comparison to $\eta\ V_{BB}$, we have

$$\eta\ V_{BB} = V_{BB}\left(1 - e^{-t_1/R_EC}\right) \tag{10.9}$$

or $$e^{-t_1/R_EC} = 1 - \eta$$

or $$e^{t_1/R_EC} = \frac{1}{1-\eta}$$

or $$t_1 = R_EC \log_e \frac{1}{1-\eta} \tag{10.10a}$$

So charging time period,

$$t_1 = 2.3\ R_EC \log_e \frac{1}{1-\eta} \tag{10.10b}$$

Since discharging time duration t_2 is negligibly small as compared to charging time duration t_1, so taking time period of the wave, $T \cong t_1$

Time period of the sawtooth wave,

$$T \equiv 2.3\ R_EC \log_{10} \frac{1}{1-\eta} \tag{10.11}$$

and frequency of oscillation

$$f = \frac{1}{T} = \frac{1}{2.3\ R_EC \log_{10} \frac{1}{1-\eta}} \tag{10.12}$$

10.4 FIELD-EFFECT TRANSISTOR

A BJT, made as an *NPN* or *PNP*, is a current-controlled device in which the movements of both the electrons and holes are involved to yield the net current. The field-effect transistor (FET), on the other hand, is a unipolar voltage-controlled device operating with only the majority carriers with the current either due to the moving electrons (in an *N*-channel FET) or due to the moving holes (in a *P*-channel), the motions being controlled by voltage at one of the terminals. The FETs are constructed in two different ways and are accordingly known as (a) JFETs and (b) MOSFETs or IGFETs.

A few general comparisons between FETs and BJTs can be made as follows:

(a) The FET has an extremely high input impedance, about 100 MΩ and above.

(b) The FET has no offset voltage when used as a switch or a chopper unlike BJT.

(c) The FET is relatively immune to radiation but the BJT is very sensitive (β is affected).

(d) The FET is less noisy than a BJT and thus more suitable for input stages of low-level amplifiers (extensively used in Hi-Fi FM receivers).

(e) The FET provides greater thermal stability than the BJT.

The main disadvantages of FETs are the following:

(a) The FET has a relatively small gain-bandwidth product compared to the BJT.

(b) The FET suffers from greater susceptibility to damage and thus needs careful handling.

The physical structure of a JFET is shown in Fig. 10.5(a) and 10.5(b). The *N*-channel JFET is constructed using a bar of *N*-type material into which a pair of *P*-type regions are diffused. The *P*-channel JFET is constructed the other way. The circuit symbols of the two types of JFETs are shown in Figs. 10.6(a) and 10.6(b).

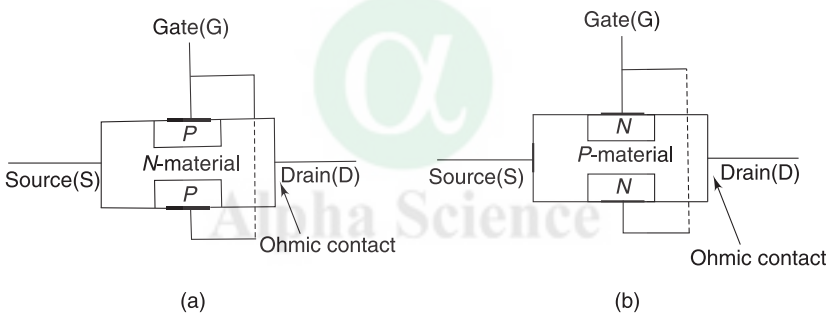


Fig. 10.5 *Construction of a FET (a) N-channel and (b) P-channel*

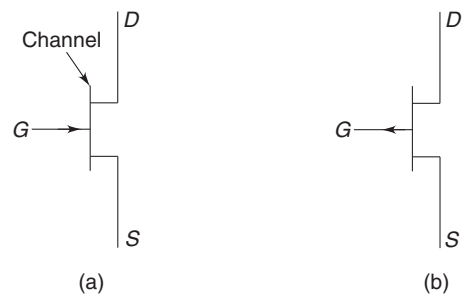


Fig. 10.6 *Circuit symbols of JFET: (a) N-channel and (b) P-channel*

10.4.1 Working and Characteristics

To examine how a JFET operates, let us consider the *N*-channel JFET shown in Fig. 10.7, with the applied bias voltages. The supply voltage V_{DD} provides

a potential difference between drain and source, V_{DS}, which results in a current I_D, from drain to source (electrons in an *N*-channel actually move from the source, hence the name 'to drain'). This drain current passes through the channel surrounded by the *P*-type gate. A voltage between gate and source, V_{GS} is set up by the supply voltage V_{GG}. Since this voltage reverse biases the gate-source junction, no gate current will result. The effect of the gate voltage will be to create a depletion region in the channel and thereby reduce the channel width to increase the drain-source resistance resulting in less drain current.

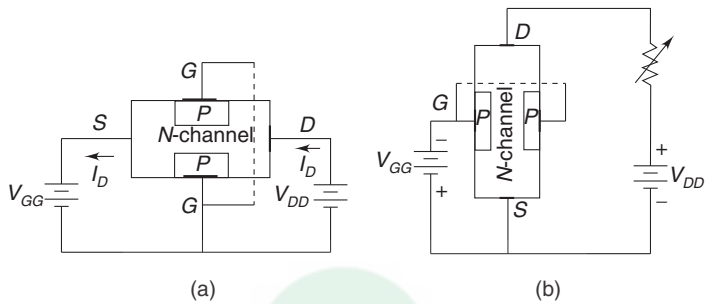


Fig. 10.7 *Basic operation of an N-channel JFET*

Let us first consider the device operating with V_{GS} = 0, i.e., with gate and source shorted, and then by increasing the reverse bias voltage V_{GS}. Figure 10.8 shows that I_D through the *N*-type material of the drain-source junction provides a voltage drop along the channel, which is more positive at the gate-drain junction than at the gate-source junction.

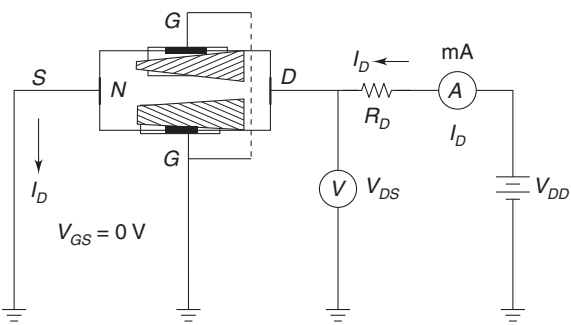


Fig. 10.8 *A JFET circuit with V_{GS} = 0 V. The device is always operated with the gate-to-source junction reverse biased. The depletion layer is wider towards the drain end because the reverse bias potential between the gate and drain is greater than that between the gate and source*

As V_{DD} (and therefore V_{DS}) is increased gradually from 0 V, I_D increases proportionately (Fig. 10.9 between *A* and *B*). in this region the channel resistance

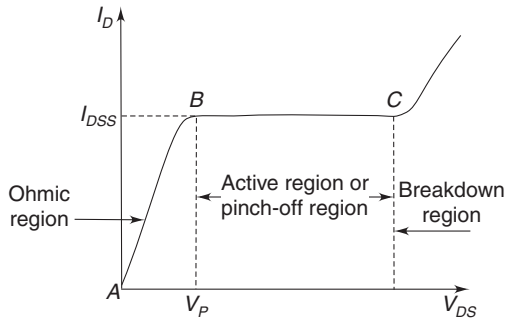


Fig. 10.9 *Drain characteristic curve for V_{GS} = 0 V*

is essentially constant because the depletion layer is not large enough to have any significant effect. This is called the ohmic region, because V_{DS} and I_D are related by Ohm's law.

At *B*, the curve levels off and I_D is relatively constant called I_{DSS}, the drain-to-source saturated current or the drain-to-source current under shorted gate condition. At this point, the reverse bias voltage between the gate and drain junction (V_{GD}) produces a depletion region which narrows down the channel width such that its resistance begins to increase sufficiently. The value of V_{GD} at this point is called the *pinch-off voltage* V_P. As $V_{DSP} = V_{DB} + V_{GS}$, we may write

$$V_{GD} = V_{GS} - V_{DSP}$$

Therefore with $V_{GS} = 0$,

$$V_{GD} = - V_{DSP}.$$

It is noted that V_P is constant for a given JFET and represents a fixed parameter. The value of V_{DS} at pinch-off (V_{DSP}) is a variable depending upon V_{GS}. This means that as $|V_{GS}|$ is increased (more negative for *N*-channel), the point corresponding to V_{DSP} of Fig. 10.9 shifts towards the origin. If V_{DS} is further increased beyond the point *B*, I_D remains constant to a specific value I_{DSS}, the maximum value of I_D when $V_{GS} = 0$. At the point *C*, breakdown occurs and I_D increases very rapidly. The JFETs are always operated below the breakdown point and within the active region.

The mechanism of current control clearly depends upon the degree to which the electric field due to the charge in the depletion region extends into the channel and provides the effect of decreasing the conductivity through the transistor, and hence the name 'Field Effect Transistor'.

Now let us consider a negative gate bias, for example, $V_{GS} = - 1$ V, as shown in Fig. 10.10. As V_{DD} is increased from 0 V, pinch-off occurs at a lower value of

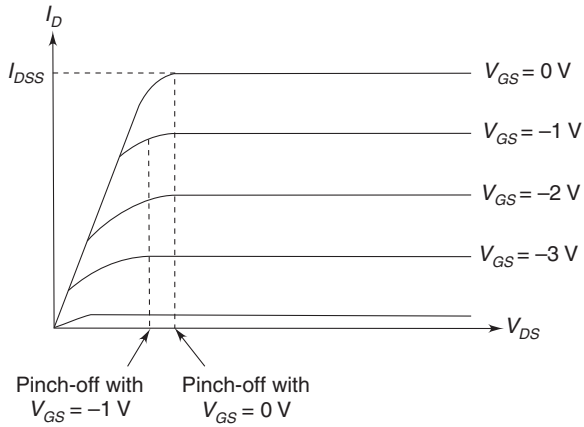


Fig. 10.10 *Family of drain curves for N-channel JFET*

V_{DS} as shown. The reason for this is that the pinch-off voltage V_P is constant and therefore for a certain negative gate voltage, the drain voltage must only reach a value sufficient to make the sufficient to make the gate-to-drain voltage equal to V_P in order to produce pinch-off. As V_{GS} is set to increasingly negative values, a family of curves is produced.

In the ohmic region ($r_{DS} = V_{DS}/I_D$), the JFET is useful as a voltage variable resistor (VVR) or voltage dependent resistor (VDR) as r_{DS} is related to V_{GS} in almost a linear manner. The pinch-off region is used for high impedance voltage amplifier.

The pinch-off voltage V_P is the drain voltage above which I_D becomes almost constant for the shorted gate condition. When V_{DS} equals V_P, the conducting channel becomes extremely narrow and the depletion layers almost touch. Because of the small passage between the depletion layers, any further increase in V_{DS} produces only a very small increase in I_D.

10.4.2 Gate-Source Cut-Off Voltage

The family of drain curves in Fig. 10.10 resembles the collector curves of a BJT. The highest curve is for $V_{GS} = 0$, the shorted-gate condition. The pinch-off voltage is V_P. As the reverse biasing V_{GS} is made more and more negative, for a value, $V_{GS} = V_{GS\,(off)}$, the depletion layers almost touch, cutting off the drain current. Since V_P is the drain voltage that also pinches off the drain current for the shorted-gate condition, $V_P = + V_{GS\,(off)}$. Since V_P is constant for a device, $V_{DSP} = 0$, since $I_D = 0$. Some data sheets do not list V_P, but they almost always list $V_{GS\,(off)}$, for example, $V_{GS\,(off)} = -4$ V on a data sheet means that $V_P = +4$ V. Let us not confuse cut-off with pinch-off. The pinch-off voltage V_P is the gate-to-drain voltage V_{GD} at which I_D reaches a constant value for a given V_{GS}. Whereas

the cut-off voltage $V_{GS\,(off)}$ is that value of V_{GS} at which $I_D = 0$, and I_D will remain zero when $|V_{GS}| \geq V_P$ and will be non-zero for less negative values of V_{GS} (I_D = 0 corresponds to off condition).

10.4.3 Transconductance Curve

From Fig. 10.10 for a given value of V_{DS}, we can read off the values of I_D for different values of V_{GS}. Then a curve for I_D versus V_{GS} can be plotted for a JFET with V_{DS} as a parameter. The curve so obtained is known as the transfer characteristic or the transconductance curve for the device and is shown in Fig. 10.11(a). Since this curve is part of a parabola, from calculus, the equation for this curve is of the form.

$$I_D = I_{DSS}\left[1 - \frac{V_{GS}}{V_{GS\,(off)}}\right]^2 \tag{10.13}$$

This is an ideal formula and can be used as an approximation for any JFET. This is why JFETs are sometimes called *square law devices*.

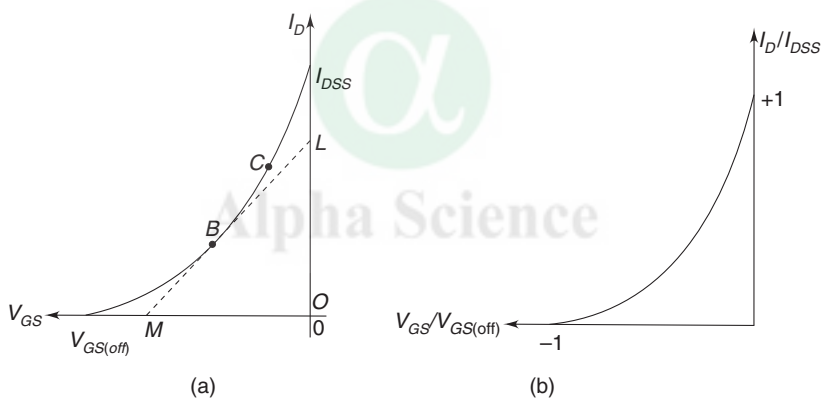


Fig. 10.11 *For N-channel JFET: (a) the transconductance curve and (b) the normalized transconductance curve*

10.4.4 Normalized Transconductance Curve

Equation (10.13) can be written as

$$\frac{I_D}{I_{DSS}} = \left[1 - \frac{V_{GS}}{V_{GS\,(off)}}\right]^2 \tag{10.14}$$

From the curve of Fig. 10.11(a), we can find out the values of $V_{GS\,(off)}$ and I_{DSS}, and then compute $V_{GS}/V_{GS\,(off)}$ and I_D/I_{DSS} and plot the curve as in Fig. 10.11(b) which is known as the normalized transconductance curve. This can also be done by taking the value of $V_{GS\,(off)}$ from the data sheet and using different values of V_{GS} in Eq. (10.14).

10.5 THE PINCH-OFF EFFECT IN FET

The control of drain current by V_{GS}, through the channel is not a repelling effect but a 'narrowing of the channel' effect. Consider what happens to the *P-N* junction between gate and source under reverse bias (Fig. 10.12).

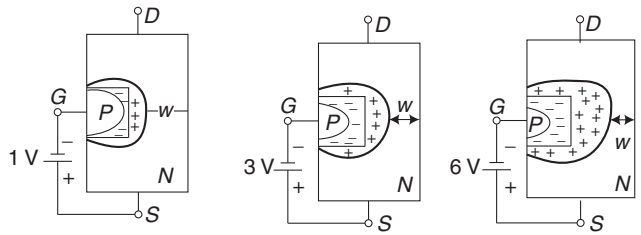


Fig. 10.12 *Effect of reverse bias on the width of the channel*

Assume, initially that no voltage is applied between *D* and *S*. With a small reverse bias, say V_{GS} = 1 V, the depletion region at the junction is quite narrow. Electrons have been drawn from the junction in the *N*-type material, leaving bound, immobile positive ions in that region of the channel, effectively reducing the width and conductivity of the channel. Similarly, negatively charged ions are left in the *P*-type material near the junction. As the reverse bias is increased to – 3 V and – 6 V, the depletion region or space charge region widens reaching ever further into the channel and restricting the passage of electrons from *S* to *D* when a voltage is applied. Finally, a value of voltage from gate to source can be applied that *pinches off* channel conduction altogether and cuts-off drain current. This value of gate voltage is called the pinch off voltage V_p and is referred to in the specifications as $V_{GS\ (off)}$.

10.5.1 Pinch-Off for V_{GS} = 0 V

In Fig. 10.13, a positive voltage V_{DS} (= 2 V) has been applied across the (*n*) channel and the gate has been connected directly to the source, to establish the condition V_{GS} = 0 V. Electrons will flow from source to the drain (and I_D will be equal to I_S). It is important to note that the depletion region is wider near the top of both *p*-type materials (i.e. nearer to the drain). The reason for the change in width of the depletion region can be understood as follows.

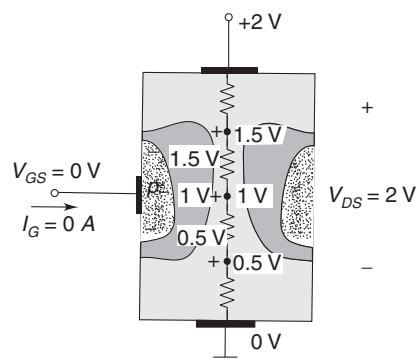


Fig. 10.13 *JFET with V_{GS} = 0 V and V_{DS} = + 2 V. As a result, there is varying reverse bias potentials across the P-N junction of this n-channel JFET*

Assuming a uniform-resistance in the *N*-channel, the resistance of the channel can be divided into divisions shown. Thus the midpoint will be at 1 V and the upper region of the *P*-type material will be reverse biased by ~1.5 V and the lower region, reverse-biased by ~0.5 V. Since the greater the applied reverse bias, the wider the depletion region – hence the distribution of the depletion region will be as shown in the figure.

10.5.2 Pinch-Off Voltage for V_{GS} = 0 V

As the voltage V_{DS} is increased from 0 to a few volts, the current will increases as determined by Ohm's law and the plot of I_D versus V_{DS} will appear as shown in Fig. 10.14. The relative straightness of the plot reveals that for the region of low value of V_{DS}, the resistance is essentially constant.

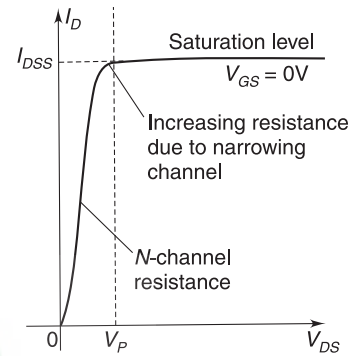


Fig. 10.14 *I_D versus V_{DS} for V_{GS} = 0 V*

As V_{DS} increases and approaches a level referred to as V_P in Fig. 10.14, the depletion region will widen, causing a resistance to increase and the curve in the graph of Fig. 10.14 to occur. The more horizontal the curve, the higher the resistance, suggesting that the resistance is approaching "infinite" ohms in the horizontal region (Fig. 10.14). If V_{DS} is increased to a level where it appears that the two depletion regions "touch" as shown in Fig. 10.15, "pinch-off" results. The level of V_{DS} that establishes this condition is the "pinch-off" voltage and is denoted by V_P as shown in Fig. 10.15.

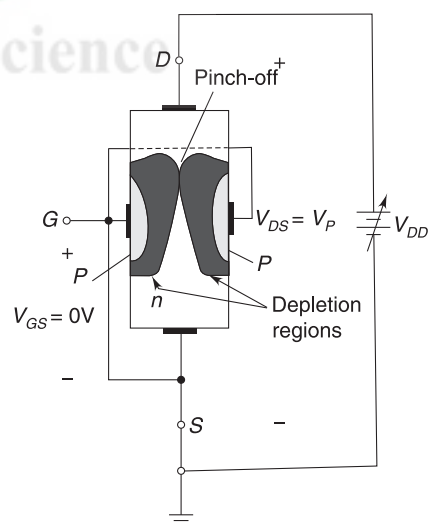


Fig. 10.15 *Pinch-off (V_{GS} = 0 V, V_{DS} = V_P)*

In actuality, the term "pinch-off" is a misnomer in that it suggests that the current I_D is pinched-off and drops to zero amperes. As shown in Fig. 10.14, however, this is not the case – I_D maintains a saturation level defined as I_{DSS} in Fig. 10.16.

10.5.3 $V_{GS} < 0$ V

The voltage from gate to source i.e., V_{GS} is the controlling voltage of the JFET. Thus curves of I_D versus V_{DS} for various levels of V_{GS} can be obtained. For the *N*-channel device, the controlling voltage V_{GS} is made more and more negative from its V_{GS} = 0V level. Consequently, the resulting saturation level for I_D reduces and will continue to decrease as V_{GS} is made more and more negative, as shown in Fig. 10.16, and ultimately, the level of V_{GS} that results in I_D = 0 mA is defined by $V_{GS} = V_P$, with V_P being a negative voltage for *N*-channel device, and a positive voltage for *P*-channel JFET.

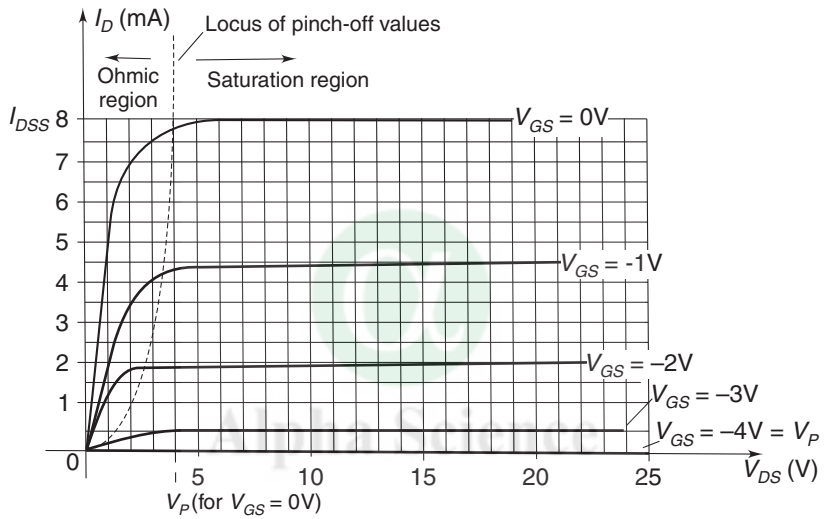


Fig. 10.16 *N-channel JFET characteristics with I_{DSS} = 8 mA and V_P = – 4 V*

10.6 JFET PARAMETERS

As we have seen above the drain current I_D in a FET depends on the drain voltage V_{DS} and the gate voltage V_{GS}. If any one of these variable is kept fixed the relation between the other two can be found. These relations are determined by three parameters of the FET.

Mutual Conductance or Transconductance The transconductance of an FET is the ratio of a small change in the drain current to the corresponding change in the gate-source voltage at constant drain-source voltage. The changes are to be considered in the straight portion of the characteristics. Thus

$$\text{Transconductance} = g_m = \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS}} \tag{10.15}$$

The transconductance is measured in the unit mho or siemens(s).

The transconductance is given by the slope of the transfer characteristic. For example, in Fig. 10.11, the value of g_m at the point B is given by OL in ampere/OM in volt.

Drain Resistance The drain resistance of an FET is the ratio of a small change in drain voltage to the resulting change in the drain current at constant gate voltage. Its unit is then ohm. Thus,

$$r_d = \left(\frac{\partial V_{DS}}{\partial I_D}\right)_{V_{GS}} \tag{10.16}$$

The drain resistance is given by the reciprocal of the slope of the output characteristic. It is clear from Fig. 10.16 that the drain resistance in the pinch-off region is very high and fairly constant. In the ohmic region the drain resistance is less than in the pinch-off region and also depends upon the gate voltage.

Amplification Factor The amplification factor is defined as the ratio of the change in the drain voltage to the corresponding decrease in gate voltage at constant drain current. Thus,

$$\mu = -\left(\frac{\partial V_{DS}}{\partial V_{GS}}\right)_{I_D} \tag{10.17}$$

The minus sign indicates that when the drain voltage is increased the gate voltage has to be decreased in order to keep the drain current constant.

10.6.1 Relation between the FET Parameters

As the drain current I_D depends on the drain voltage V_{DS} and the gate voltage V_{GS}, we can write

$$I_D = f(V_{DS}, V_{GS}) \tag{10.18}$$

Hence, we get

$$dI_D = \left(\frac{\partial I_D}{\partial V_{DS}}\right)_{V_{GS}} dV_{DS} + \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS}} dV_{GS}$$

Dividing both sides by dV_{GS} we get

$$\frac{dI_D}{dV_{GS}} = \left(\frac{\partial I_D}{\partial V_{DS}}\right)_{V_{GS}} \left(\frac{dV_{DS}}{dV_{GS}}\right) + \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS}}$$

If I_D is constant then $\dfrac{dI_D}{dV_{GS}} = 0$.

$$\therefore \quad 0 = \left(\frac{\partial I_D}{\partial V_{DS}}\right)_{V_{GS}} \left(\frac{\partial V_{DS}}{\partial V_{GS}}\right) + \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS}}$$

Substituting for the partial derivatives from Eqs. (10.16), (10.17) and (10.18) we get

$$0 = \left(\frac{1}{r_d}\right)(-\mu) + g_m$$

$$\therefore \quad \mu = g_m\, r_d \tag{10.19}$$

Thus the amplification factor is equal to the product of the transconductance and the drain resistance.

10.7 VARIATION OF TRANSCONDUCTANCE WITH DRAIN CURRENT AND GATE VOLTAGE

Differentiating I_{DS} in Shockley's equation with respect to V_{GS}, we get

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

But $$\left(\frac{\partial I_D}{\partial V_{DS}}\right) = g_m$$

$$\therefore \quad g_m = -\frac{2I_{DSS}}{V_P}\left(1 - \frac{V_{GS}}{V_P}\right) \tag{10.20}$$

Now from Eq. (10.14) of Section 10.44, we have

$$\left(1 - \frac{V_{GS}}{V_P}\right) = \sqrt{\frac{I_{DS}}{I_{DSS}}}$$

Substituting this value in Eq. (10.21)

$$g_m = -2\,\frac{\sqrt{I_{DS}\, I_{DSS}}}{V_P} \tag{10.21}$$

Let $g_m = g_{mo}$ when $V_{GS} = 0$. Then from Eq. (10.21), we get

$$g_{mo} = -\frac{2I_{DSS}}{V_P} \tag{10.22}$$

Combining Eqs (10.21) and (10.22), we get

$$g_m = g_{mo}\left(1 - \frac{V_{GS}}{V_P}\right) \tag{10.23}$$

Equation (10.21) shows that the transconductance varies as the square root of the saturation drain current I_{DS}. Equation (10.23) shows that the transconductance decreases linearly with increase of gate voltage V_{GS}.

10.8 CHANNEL JFET

We have so far confined our discussion of FETs to the *N*-channel type. *P*-channel JFETs are also available. The above discussion can be applied to the *P*-channel

JFETs by interchanging *N* and *P* and reversing the voltages and currents. The following table gives the comparison of the *N*-channel and *P*-channel JFETs.

	N-channel JFET	*P*-channel JFET
1.	Current carriers are electrons	Current carriers are holes
2.	Mobility of electrons large	Mobility of holes poor
3.	Input noise low	Input noise large
4.	Transconductance large	Transconductance small

The above facts show that *N*-channel FETs serve better than *P*-channel FETs.

10.9 ESSENTIALS OF FET MODELLING

The field effect transistor (FET), like the bipolar junction transistor (BJT), is a three-terminal semiconductor device having the circuit symbol shown in Fig. 10.17 with the terminals marked as source (S), gate (G), and drain (D). An FET is called a unipolar device because the current through it results from the flow of only one of the two kinds of charge carriers, namely holes or electrons. Essentially, an FET is a voltage-operated device in that a constant voltage at the gate-source terminals affects the current flow through the device. There are two main types of field effect transistors: the junction field-effect transistor (JFET) and the metal-oxide-semiconductor FET (MOSFET). In the following section, we will mainly focus our attention on developing the small signal equivalent model of a JFET.

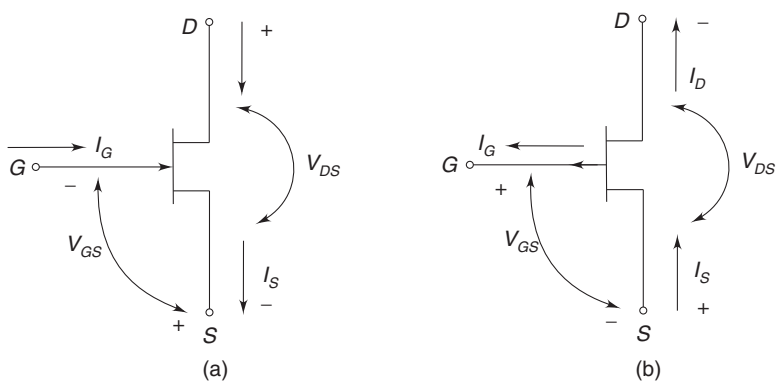


Fig. 10.17 *FET symbol: (a) N-channel JFET and (b) P-channel JFET*

The field effect transistor is always operated with the gate-source junction reverse-biased. The *N*-channel JFET transfer and output characteristics are displayed in Fig. 10.18. For analytical purposes, the volt-ampere characteristics

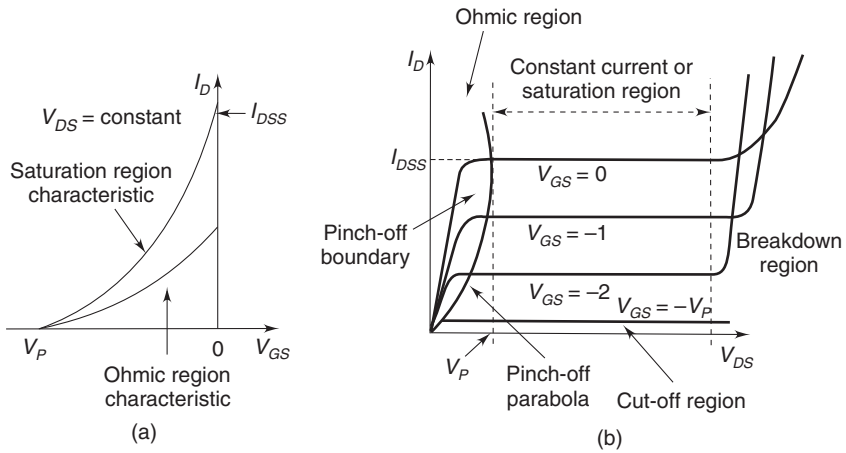


Fig. 10.18 *N-channel JFET characteristics: (a) transfer characteristic and (b) output characteristic*

of JFET as illustrated in Fig. 10.18, may be conveniently divided into four regions of operation. These are:

1. Region below Pinch-off: Ohmic Region

This region is marked on the output characteristic curves as shown in Fig. 10.18(b). The ohmic region is defined by $V_{DS} \leq |V_P|$, i.e., the drain-source voltage is less than the pinch-off voltage. In this region, the volt-ampere relationship is closely approximated by

$$I_D = I_{DSS}\left[2\left(1 - \frac{V_{GS}}{V_P}\right)\frac{V_{DS}}{V_P} - \left(\frac{V_{DS}}{V_P}\right)^2\right] \tag{10.24}$$

For small values of V_{DS}, Eq. (10.24) becomes

$$I_D \equiv \frac{2I_{DSS}}{V_P}\left(1 - \frac{V_{GS}}{V_P}\right)V_{DS} \tag{10.25}$$

2. Region between Pinch-off and Breakdown: Saturation Region

This region as marked on the characteristic curves of Fig. 10.18(b) is characterized by

$$(V_{GS} - V_P) < V_{DS} < V_{DO}$$

where V_{DO} is the drain-source breakdown voltage.

In other words, we say that this is the region where the drain-source voltage is greater than the pinch-off voltage and the drain current as illustrated in Fig. 10.18(b) is almost constant and independent of V_{DS}. In this region, the drain current strongly depends upon the gate-source voltage V_{GS} and is closely approximated by the equation of the form

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{10.26}$$

3. Cut-off Region

The cut-off region is defined by $|V_{GS}| > |V_P|$. Under this condition $I_D = I_D/(\text{OFF})$ which is in the order of a few picoamperes and hence we assume that $I_D = 0$.

4. Breakdown Region

This region is characterized by the fact that the drain-source voltage V_{DS} exceeds the drain-source breakdown voltage V_{DO}.

Note that as marked in Fig. 10.18(b), the pinch-off parabola which separates the ohmic region from the saturation region is dictated by the equation

$$V_{GS} - V_{DS} = V_P \tag{10.27}$$

The finite slope of the FET output curves (in the saturation region) can be accommodated by introducing an additional term to the FET transfer characteristic Eq. (10.26) as

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2\left(1 + \frac{V_{DS}}{V_A}\right) \tag{10.28}$$

where V_A is the equivalent to the Early voltage for the BJT. This Earlt effect is due to the shortening of the channel length as more voltage is applied across the region where the depletion regions make contact.

For computer simulation, the transfer characteristic of the JFET operating in saturation region is defined by

$$I_D = \beta\,(V_{GS} - V_{TO})^2\,(1 + \lambda V_{DS}) \tag{10.29}$$

with the following relationships

$$\beta = \frac{I_{DSS}}{V_P^2}, \quad \lambda = \frac{1}{V_A}, \quad V_{TO} = V_P \tag{10.30}$$

Note that Eqs. (10.28) and (10.29) are equivalent. Also in the ohmic region, $0 < V_{DS} < (V_{GS} - V_P)$, the characteristic curve is approximated by the relation of the form

$$I_D = \beta\left[2(V_{GS} - V_{TO})\,V_{DS} - V_{DS}^2\right](1 + \lambda V_{DS}) \tag{10.31}$$

Note that Eqs. (10.24) and (10.31) are also equivalent.

Normally when accuracy is not required, the JFET transfer characteristic is defined by Eq. (10.26).

The JFET is normally biased to operate in the saturation region, where the I_D remains constant with V_{DS}. Figure 10.19 shows a simple biasing scheme. For the circui in Fig. 10.19(a), using KVL in the drain-source loop, we may write

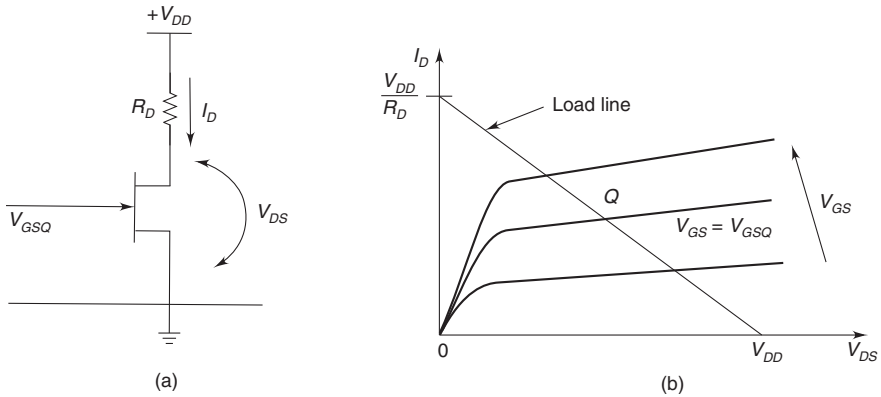


Fig. 19.19 *(a) N-channel JFET symbol with biasing arrangement and terminal voltages and currents and (b) output characteristic curves with load line and the operating point Q*

$$I_D R_D + V_{DS} = V_{DD}$$

or
$$\frac{I_D}{\frac{V_{DD}}{R_D}} + V_{DS}/V_{DD} = 1 \tag{10.32}$$

which is an equation of straight line.

Equation (10.32) is plotted on V_{DS} versus I_D output characteristic of the JFET. Because the JFET output characteristics and Eq. (10.32) must be satisfied simultaneously, this gives an intersection point Q as marked in Fig. 19.19(b) for a constant $V_{GS} = V_{GSQ}$. This point Q is called the Q-point or the quiescent point. A close inspection of Fig. 10.19(b) in the close vicinity of the Q-point shows that for small signals the output characteristics of an N-channel JFET may be taken as linear, parallel, and equidistant. This is shown in Fig. 10.20. We may then look for a model to represent the small-signal behaviour of JFET in the close vicinity of the Q-point. The model so developed will essentially going to be a linear model capable of being analyzed by algebraic methods.

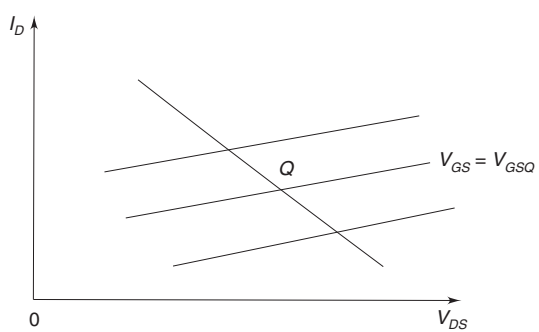


Fig. 10.20 *JFET output characteristics, with load line, in the close vicinity of the Q-point*

10.10 D.C. LOAD LINE AND BIAS POINT

The dc load line for a FET circuit is drawn upon the output characteristics of the device in exactly the same way as for a BJT circuit.

Consider the FET circuit and characteristics shown in Figs. 10.22 and 10.21 respectively. The drain-source voltage is given as

$$V_{DS} = V_{DD} - I_D R_D \tag{10.33}$$

where V_{DD} is the supply voltage and $I_D R_D$ is the voltage drop across R_D.

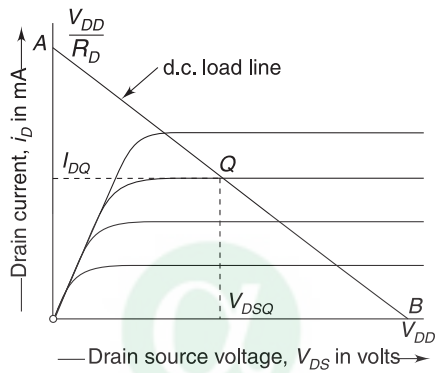


Fig. 10.21

Consider the following two particular situations:

(i) When $I_D = 0$, $V_{DS} = V_{DD}$ … cut-off point B

(ii) When $V_{DS} = 0$, $I_D = \dfrac{V_{DD}}{R_D}$ … saturation point A

By joining these two points A and B, d.c. load line is obtained.

A *d.c. bias point or quiescent point* (Q-point) is selected on the FET d.c. load line. This point defines the dc conditions that exist in a circuit when no input signal is applied. The bias point is selected to give maximum possible variations in output (drain-source) voltage when the drain current is varied by an input signal. In this case the positive-going and negative-going output voltage swings should be closely equal as possible. Where maximum output voltage swings is not required, the bias point may be at any convenient position on the load line.

For the load line in Fig. 10.21, the bias conditions are $\dfrac{V_{DD}}{2}$ and $\dfrac{V_{DD}}{2R_D}$.

For a FET amplifier, V_{DS} must remain in the pinch-off region of the characteristics. This mean that it must not be allowed to go below the level of the pinch-off voltage (V_P). Thus, in the design of a FET bias circuit, the drain source voltage should always be a minimum of ($V_P + 1$) volt.

i.e., $$V_{DS\,(\min)} = V_P + 1 \text{ volts} \tag{10.34}$$

When an external bias voltage V_{GS} is included, the pinch-off voltage for that bias level on the drain characteristics is $V_P - V_{GS}$, where V_P is the pinch-off voltage at $V_{GS} = 0$. Consequently, the minimum drain-source voltage may be reduced to

$$V_{DS\,(\min)} = V_P - V_{GS} + 1 \text{ volts} \tag{10.35}$$

For the operation of FET as a switch, the device will either be biased *off* or *on* into the channel ohmic region of the drain characteristics.

10.11 FET BIASING

Unlike BJTs, thermal runaway does not occur with FETs. However, the wide differences in maximum and minimum transfer characteristics make I_D levels unpredictable with simple fixed-gate bias voltage. To obtain reasonable limits on quiescent drain current I_D and drain-source voltage V_{DS}, source resistor and potential divider bias techniques must be used. With few exceptions, MOSFET bias circuits are similar to those used for JFETs. Various FET biasing circuits are discussed below:

10.11.1 Fixed Bias

D.C. bias of a FET device needs setting of gate-source voltage V_{GS} to give desired drain current I_D. For a JFET drain current is limited by the drain-source saturation current I_{DSS}. Since the FET has such a high input impedance that no gate current flows and the d.c. voltage of the gate set by a voltage divider or a fixed battery voltage is not affected or loaded by the FET.

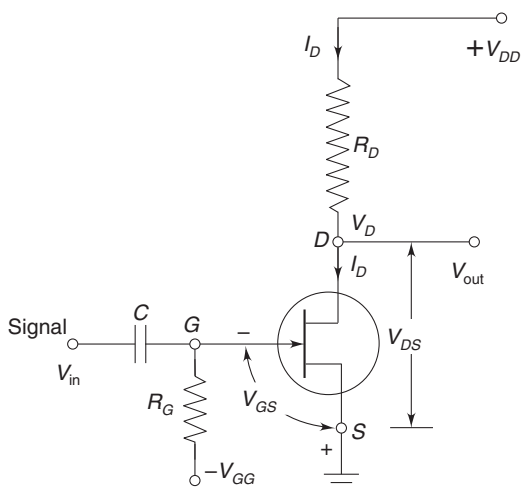


Fig. 10.22 *Fixed bias circuit for JFET*

Fixed d.c. bias is obtained using a battery V_{GG}. This battery ensures that the gate is always negative w.r.t. source and no current flows through resistor R_G and gate terminal i.e., $I_G = 0$. The battery provides a voltage V_{GS} to bias the *N*-channel JFET, but no resulting current is drawn from the battery V_{GG}. Resistor R_G is included to allow any a.c. signal applied through capacitor C to develop across R_G. While any a.c. signal will develop across R_G, the d.c. voltage drop across R_G is equal to $I_G R_G$ i.e., 0 volt.

The gate-source voltage V_{GS} is then

$$V_{GS} = - V_G - V_S = - V_{GG} - 0 = - V_{GG} \tag{10.36}$$

The drain-source current I_D is then fixed by the gate-source voltage as determined by Eq. (12.56) Section 12.9.

This current then causes a voltage drop across the drain resistor R_D and is given as

$$V_{R_D} = I_D R_D$$

and output voltage,

$$V_{\text{out}} = V_{DD} - I_D R_D \tag{10.37}$$

Since V_{GG} is fixed value of d.c. supply and the magnitude of gate-to-source voltage V_{GS} is also fixed, hence this circuit is named as *fixed bias circuit*. Since this bias circuit uses two batteries V_{DD} and V_{GG}, it is also known as *two battery bias circuit*.

A FE has a high input impedance. To make advantage of it, R_G should be as large as possible so that the input impedance of the circuit remains high. If R_G is extremely large a charge accumulated on the gate may take a long time to leak off. A reasonable upper limit is 1 MΩ. Normally R_G should not exceed this value.

10.11.2 Self-bias

This is the most common method for biasing a JFET. Self-bias for an *N*-channel JFET is shown in Fig. 10.23. This circuit eliminates the requirement of two d.c. supplies i.e., only drain supply is used and no gate supply is connected. In this circuit, a resistor R_S, known as *bias resistor*, is connected in the source leg. The dc component of drain current I_D flowing through R_S makes a voltage drop across resistor R_S. The voltage drop across R_S reduces the gate-to-source reverse voltage required for FET operation. The capacitor C_S by-passes the a.c. component of the drain current I_D. The addition of resistor R_G in the circuit does not disturb the dc bias. This is because gate current flowing through it is zero and the gate leakage current is also almost zero. Thus, gate is essentially at d.c. ground. The resistor R_G is inserted in the circuit to avoid the short circuiting of the a.c. input voltage.

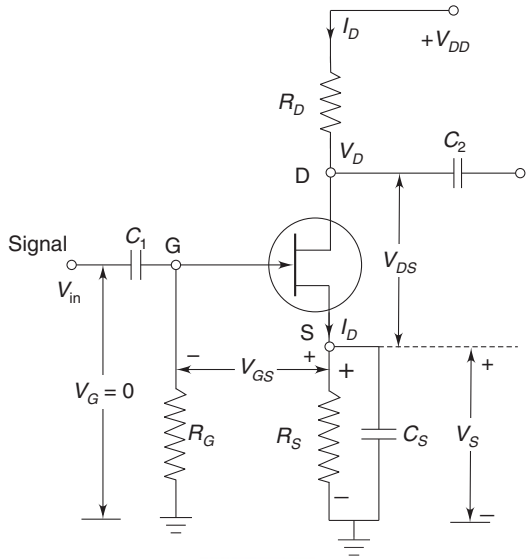


Fig. 10.23 *Self-bias circuit for N-channel JFET*

Further, if there is a leakage, the resistor R_G will provide to it an escape route. Otherwise, the leakage current would build up static charge (voltage) at the gate which could change the bias. The resistor R_S, the feedback resistor, also helps in preventing any variation in FET drain current.

Since no gate current flows through the reverse biased gate, the gate current $I_G = 0$ and, therefore,

$$V_G = I_G R_G = 0 \text{ V}$$

With a drain current I_D the voltage at the S is

$$V_S = I_D R_S \tag{10.38}$$

The gate-to-source voltage is then

$$V_{GS} = V_G - V_S = 0 - I_D R_S = - I_D R_S \tag{10.39}$$

So voltage drop across resistance R_S provides the biasing voltage V_{GS} and no external source is required for biasing and this is the reason that it is called *self biasing*.

The operation point (i.e. zero signal I_D and V_{DS}) can easily be determined from Eq. (12.56) Section 12.9 and equation given below:

$$V_{DS} = V_{DD} - I_D (R_D + R_S) \tag{10.40}$$

Thus d.c. conditions of JFET amplifier are fully specified.

Self-biasing of a JFET stabilizes its quiescent operating point against any change in its parameters like transconductance. Let the given JFET be replaced by another JFET having the double conductance then drain current will also try to be double but since any increase in voltage drop across R_S, therefore, gate-source voltage, V_{GS} becomes more negative and thus increase in drain current is reduced.

10.11.3 Potential Divider Bias

A slightly modified form of dc bias is provided by the circuit shown in Fig. 10.24. The resistor R_{G_1} and R_{G_2} form a potential divider across drain supply V_{DD}. The voltage V_2 across R_{G_2} provides the necessary bias. The additional gate resistor R_{G_1} from gate to supply voltage facilitates in larger adjustment of the dc bias point and permits use of larger valued R_S.

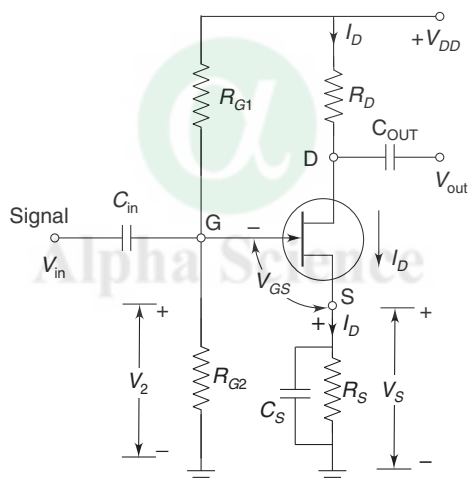


Fig. 10.24 *Potential divider bias circuit for N-channel JFET*

The gate is reverse biased so that $I_G = 0$ and gate voltage

$$V_G = V_2 = \frac{V_{DD}}{R_{G_1} + R_{G_2}} \times R_{G_2} \tag{10.41}$$

and

$$V_{GS} = V_G - V_S = V_G - I_D R_S \tag{10.42}$$

The circuit is so designed that $I_D R_S$ is larger than V_G (or V_2) so that V_{GS} is negative. This provides correct bias voltage.

The operating point can be determined as

$$I_D = \frac{V_2 - V_{GS}}{R_S}$$

and $$V_{DS} = V_{DD} - I_D\,(R_D + R_S) \tag{10.43}$$

Maximum gain is achieved by making resistance R_D as large as possible and for a given level of I_D it needs maximum voltage drop across resistor R_D. However, greatest bias stability is achieved by making R_S as large as possible.

If the gate voltage V_G is very large as compared to gate-to-source voltage V_{GS}, the drain current is approximately constant. In practice, the voltage divider bias is less effective with JFET than BJT. This is because in BJT, V_{BE} = 0.7 V (silicon) with only minor variation from one transistor to another transistor. But in a JFET, the V_{GS} can vary several volts from one JFET to another.

10.11.4 Current-Source Bias

When the drain supply voltage V_{DD} is not large, there may not be enough gate voltage to swamp out the variations in V_{GS}. In such a case current-source bias (Fig. 10.25) may be used. In this arrangement, the BJT pumps a fixed current through the JFET. The drain current is given by

$$I_D = \frac{V_{EE} - V_{BE}}{R_E} \tag{10.44}$$

Figure 10.26 illustrates how effective current-source bias is. Both Q points have the same value of drain current. Although V_{GS} is different for each Q point, V_{GS} no longer has any effect on the value of drain current.

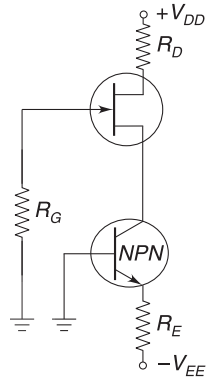


Fig. 10.25 *Current-source bias*

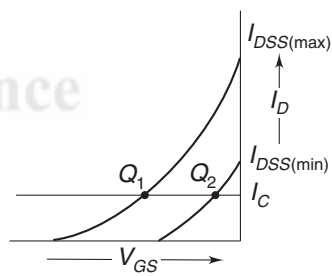


Fig. 10.26

10.12 FET SMALL SIGNAL MODELS

10.12.1 Low Frequency Model

In a FET, instantaneous drain current i_D is a function of the instantaneous gate source voltage v_{GS} and instantaneous drain-source voltage v_{DS} and is, therefore, expressed as

$$i_D = f\,(v_{GS},\, v_{DS}) \tag{10.45}$$

If both the gate and drain voltages are varied, the change in drain current is given approximately by the first two terms in the Taylor's series expansion of Eq. (10.44) or

$$\Delta i_D = \left.\frac{\partial i_D}{\partial v_{GS}}\right|_{V_{DS}} \Delta v_{GS} + \left.\frac{\partial i_D}{\partial v_{DS}}\right|_{V_{GS}} \Delta v_{DS} \tag{10.46}$$

Using the conventional small signal notations, Δi_D, Δv_{GS} and Δv_{DS} may be replaced respectively by time varying components i_d, v_{gs}, and v_{ds}. Now Eq. (10.46) becomes

$$i_d = g_m\, v_{gs} + \frac{1}{r_d}\, v_{ds} \tag{10.47}$$

where

$$g_m = \left.\frac{\partial i_D}{\partial v_{GS}}\right|_{V_{DS}} = \left.\frac{\Delta i_D}{\Delta v_{GS}}\right|_{V_{DS}} = \left.\frac{i_d}{v_{gs}}\right|_{V_{DS}} \tag{10.48}$$

Parameter g_m is the mutual conductance, or transconductance

and

$$\frac{1}{r_d} = \left.\frac{\partial i_D}{\partial v_{DS}}\right|_{V_{GS}} = \left.\frac{\Delta i_D}{\Delta v_{DS}}\right|_{V_{GS}} = \left.\frac{i_d}{v_{ds}}\right|_{V_{GS}} \tag{10.49}$$

The reciprocal of r_d is the drain conductance g_d.

Circuit shown in Fig. 10.27 is made satisfying Eq. (10.47) giving the incremental drain current i_d in form of g_m, r_d, v_{gs} and v_{ds}. This circuit forms the low frequency small signal model for FET: This model consists of one dependent current generator whose current $g_m\, v_{gs}$ is proportional to the time varying gate source voltage v_{gs} and proportional constant is the transconductance g_m. The arrow points from drain to source to indicate a phase change of 180° (or phase reversal) between output and input voltages as will occur in actual operation. In low frequency model, as the gate is reverse biased, the gate current becomes zero. So, in low frequency model, gate to source junction is represented by an open circuit and no current is drawn by the input terminal of the FET. The reason is that input resistance, (gate-to-source resistance) is very large. The output resistance is represented by the resistance r_d (resistance from drain to source). If r_d is sufficiently large as compared to those of other circuit elements, this is ignored. Now the equivalent circuit becomes simply a current source whose magnitude is controlled by v_{gs} and g_m—clearly a voltage controlled current source.

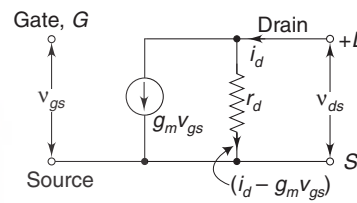


Fig. 10.27 *Low frequency small signal model for FET*

Small signal model for sinusoidal input voltage of *rms* value V_{gs} may be drawn, as shown in Fig. 10.28 where v_{gs}, i_d, v_{ds} have been replaced respectively by rms values V_{gs}, I_d and V_{ds}. The voltage-controlled current source ($g_m\ V_{gs}$) can be replaced by

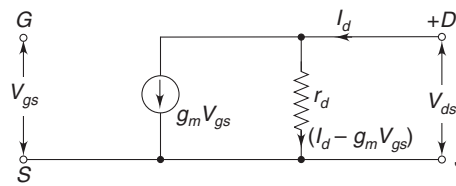


Fig. 10.28 *Low frequency small signal model for sinusoidal input of FET*

equivalent voltage-controlled voltage source (μV_{gs}) as depicted in Fig. 10.29.

Now the equivalent circuit consists of a voltage source (μV_{gs}) in series with a drain resistance r_d. Here μ is the amplification factor of the FET and is equal to the product of transconductance g_m and drain resistance r_d i.e.,

$$\mu = g_m r_d.$$

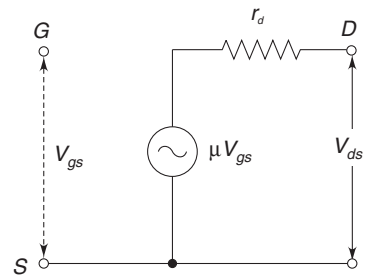


Fig. 10.29

Comparison of Low Frequency Models of FET and BJT

1. Both FET and BJT models have a dependent current generator in the output circuit.
2. In FET models, the generator current is proportional to the input voltage V_{gs} while in BJT models, the generator current is proportional to the input current.
3. In FET models input impedance is very high (theoretically infinite at low frequencies) while in common emitter BJT model the input impedance is of the order of 800 Ω.
4. In FET, there is no feedback from output (drain) to the input (source) while in BJT models it is. Thus it can be safety said that at low frequencies, FET forms a more ideal amplifier than BJT amplifier.

10.12.2 High Frequency Model

At low frequency, reactance offered by interelectrode capacitance is very large and, therefore, interelectrode capacitances do not appear in low frequency equivalent circuit. At higher frequencies device capacitances between terminals cause reduction in amplifier gain as capacitive impedance decreases with the increase in supply frequency. Such circuit capacitances resulted in due to device construction or stray wiring are shown in Fig. 10.30 to indicate that they are not capacitances that are connected into the circuit intentionally but arise as a result of the circuit and device construction.

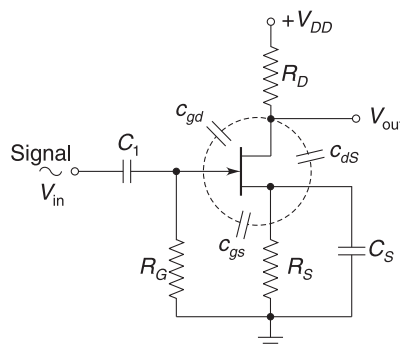


Fig. 10.30

Figure 10.31 depicts the high frequency model of the FET which is identical with Fig. 10.27 except that the capacitances between the gate-source and gate-

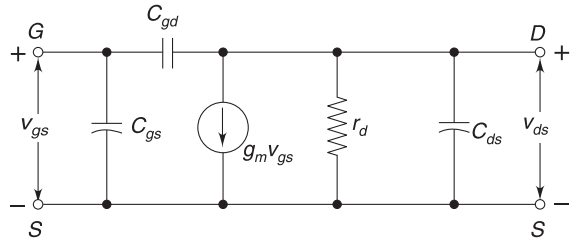


Fig. 10.31 *High frequency model of FET*

drain nodes are shown in the figure. The capacitor C_{gs} represents the barrier capacitance between gate and source, and C_{gs} is the barrier capacitance between gate and drain. The element C_{ds} represents the drain to source capacitance of the channel. Because of these internal capacitances feedback exists between the input and output circuits of the FET and voltage amplification drops drastically with increase in frequency.

10.13 COMMON SOURCE JFET AMPLIFIER

The common source configuration is the most widely employed configuration. It is because this configuration provides high input impedance, good voltage gain and moderate output impedance.

The circuit of a common source *N*-channel JFET amplifier using self bias is shown in Fig. 10.32. The signal source, V_{in} is connected to JFET gate through coupling capacitor C_1 and external load R_L is connected to the drain terminal D via coupling capacitor C_2. R_G is used to provide leakage path to the gate current, R_S for developing gate bias and C_S for providing ac ground to the iput signal.

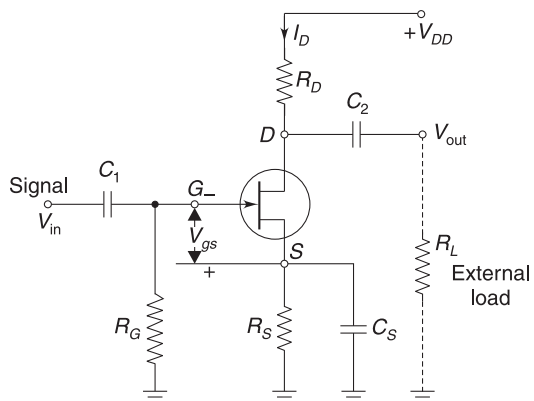


Fig. 10.32 *Common source amplifier circuit*

Note that the polarity of the capacitors in Fig. 10.32, the negative terminal of each (polarized) capacitor must always be at a lower d.c. potential than the positive terminal. Incorrectly connected capacitors can affect the bias conditions and ac performance of the circuit.

For ac considerations, by pass capacitor C_S effectively ties the FET source terminal to the ground level. Thus, all of the input voltage V_{in} is developed across the gate source terminals. With the increase or decrease in input voltage, gate-source voltage V_{gs} changes and corresponding change in V_{gs} makes the drain current I_D to increase or decrease correspondingly.

The phase relationship between the circuit input and output voltages can be determined by considering the effect of a positive going input signal. When input voltage V_{in} increases in positive direction, it reduces the negative gate-source voltage $-V_{gs}$. The reduction in $-V_{gs}$ raises the level of drain current I_D and consequently increases the voltage drop across drain resistance R_D. Since $V_D = V_{DD} - I_D R_D$, the increase in I_D results in a drop in drain (output) voltage V_D. Thus as V_{in} increases in a positive direction. V_{out} goes in a negative direction. Similarly, when V_{in} goes negative, the resultant increase in $-V_{gs}$ reduces the drain current I_D which, in turn, reduces the voltage drop across R_D and develops a positive going output voltage V_{out}. Thus, the amplifier output voltage is 180° out of phase with the input voltage.

Analysis The first step in a.c. analysis of the common source circuit is to draw the ac equivalent circuit. This is done by replacing all the capacitors with short circuits and reducing d.c. supply voltages to zero. A.C. equivalent circuit for common source amplifier is given in Fig. 10.33.

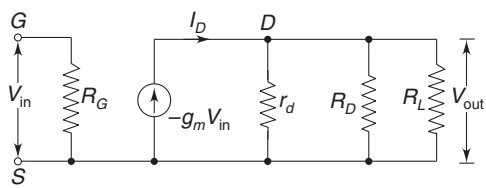


Fig. 10.33 *A.C. equivalent circuit for common source JFET amplifier*

Input Resistance In an ideal JFET R_{gs} is infinite because $I_G = 0$. However, in actual device R_{gs} is not infinite but extremely high (100 MΩ or so) in comparison to R_G. Thus, input resistance

$$R_{in} = R_G \parallel R_{gs} = R_G \tag{10.50}$$

Output Resistance Looking into the drain and source terminals, the large drain resistance r_d is seen. Thus,

$$Z_d = r_d \tag{10.51}$$

Z_d is the device output impedance; the circuit output impedance is R_D in parallel with Z_d, so

$$Z_{\text{out}} = R_D \parallel Z_d = R_D \parallel r_d \tag{10.52}$$

Since usually $r_d >> R_D$, the circuit output impedance is taken to be R_D.

Voltage Gain Output voltage,

$$V_{\text{out}} = I_D\,(r_d \parallel R_D \parallel R_L)$$
$$= -\,g_m\,V_{in}\,(r_d \parallel R_D \parallel R_L) \qquad \because I_D = -\,g_m\,v_{in}$$

and voltage gain,

$$A_V = \frac{V_{\text{out}}}{V_{\text{in}}} = -\,g_m\,(r_d \parallel R_D \parallel R_L) \tag{10.53}$$

usually $r_d >> R_D \parallel R_L$

So voltage gain,

$$A_V = -\,g_m\,(R_D \parallel R_L) \tag{10.54}$$

The minus sign indicates that output voltage V_{out} is 180° out of phase with input voltage V_{in}.

10.14 COMMON DRAIN JFET AMPLIFIER, OR SOURCE FOLLOWER

In the common drain circuit (also called the *source follower*), the output voltage is developed across source resistor R_S. External load resistor R_L is connected to FET source terminal S through coupling capacitor C_2, and gate bias voltage V_G is derived from V_{DD} by means of potential divider R_1 and R_2. No resistor is connected in series with the drain terminal, and no source bypass capacitor is used. The input signal is applied to the gate through coupling capacitor C_1 (Fig. 10.34).

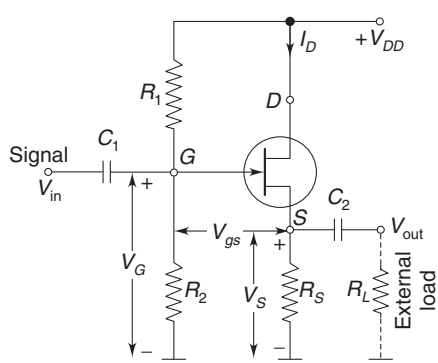


Fig. 10.34 *Common drain amplifier circuit*

For understanding the operation of the common drain circuit, note that gate voltage V_G is a constant quantity, as in any potential divider bias circuit. When a signal is applied to the FET gate through C_1, gate voltage increases and decreases as the input signal goes positive and negative respectively. Also V_{gs} remains substantially constant, therefore, the source voltage V_S increases or decreases with the increase or decrease in gate voltage. Since the FET source is the output terminal, it is seen that the output voltage from a common drain circuit is approximately the same as the input voltage. Thus, the common drain circuit can be said to have approximately unity gain. Because the output voltage at the FET source terminal follows variations in the signal voltage applied to the gate, the common drain circuit is also known as a *source* follower.

The common drain a.c. equivalent circuit is shown in Fig. 10.35. The current generator is $g_m V_{gs}$ where $V_{gs} = V_{\text{in}} - V_{\text{out}}$. Moreover $R_G = R_1 \parallel R_2$.

$$V_{\text{out}} = I_D (r_d \parallel R_S \parallel R_L) \qquad \text{where } I_D = g_m V_{gs}$$

or
$$V_{\text{out}} = g_m V_{gs} (r_d \parallel R_S \parallel R_L)$$

and
$$V_{\text{in}} = V_{gs} + V_{\text{out}}$$
$$= V_{gs} + V_{gs} g_m (r_d \parallel R_S \parallel R_L) \qquad \because V_{gs} = V_{\text{in}} - V_{\text{out}}$$

Voltage gain

$$A_v = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{g_m (r_d \parallel R_S \parallel R_L)}{1 + g_m (r_d \parallel R_S \parallel R_L)}$$

Normally $r_d >> R_S \parallel R_L$

$\therefore$
$$A_v = \frac{g_m (R_S \parallel R_L)}{1 + g_m (R_S \parallel R_L)} \tag{10.55}$$

If $g_m (R_S \parallel R_L) >> 1$ then $A_v = 1$

So value of voltage gain becomes approximately equal to unity (slightly less than unity).

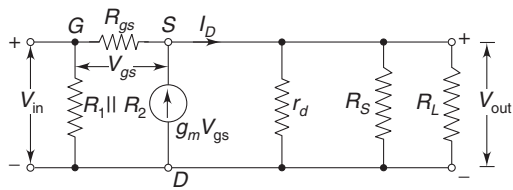


Fig. 10.35 *A.C. equivanelnt circuir for common drain JFET amplifier*

Input Resistance The circuit input impedance is

$$R_G = R_1 \parallel R_2.$$

Output Resistance In a common drain circuit, any variation in output voltage V_{out} affects upon the gate-source voltage V_{gs} and V_{gs} is given as

$$V_{gs} = V_{out} \times \frac{R_{gs}}{(R_S \parallel R_G) + R_{gs}}$$

Normally $R_{gs} >> R_S \parallel R_G$,

so $$V_{gs} = V_{out}$$

and $$I_D = g_m V_{gs}$$

Also $$Z_S = \frac{V_{out}}{I_D} = \frac{V_{gs}}{g_m V_{gs}} = \frac{1}{g_m} \tag{10.56}$$

Actually, r_d is in parallel with $\frac{1}{g_m}$, but since $r_d >> \frac{1}{g_m}$, it is neglected, R_S is the output impedance of the device. The circuit output impedance

$$Z_{out} = R_S \parallel \frac{1}{g_m} \tag{10.57}$$

10.15 COMMON GATE JFET AMPLIFIER

The circuit of a simple common gate amplifier is shown in Fig. 10.36 and its ac equivalent circuit is shown in Fig. 10.37. In this amplifier circuit, input signal is applied to the source terminal S and output is taken from the drain terminal D, as illustrated in Fig. 10.36. The gate is grounded and external load resistor R_L is connected to the drain terminal D through coupling capacitor C_2.

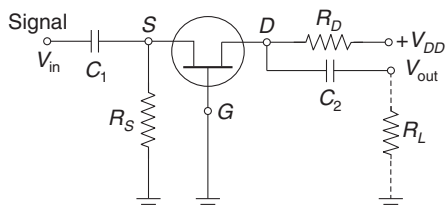


Fig. 10.36 *Common gate amplifier circuit*

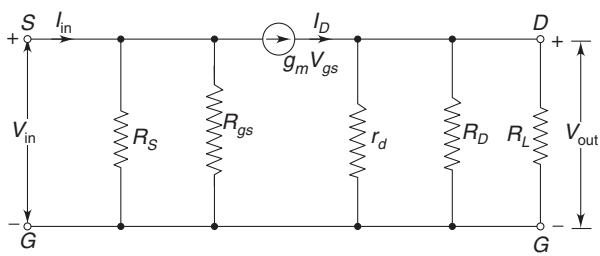


Fig. 10.37 *A.C. equivalent circuit for common gate JFET amplifier*

In a.c. equivalent circuit, current source is connected between drain and source terminals, as usual. However, since source and drain are the input and output terminals respectively, $g_m V_{gs}$ appears between the input and output.

Input Impedance Ignoring current through R_S and R_{gs}, the input current

$$I_{in} = I_D = g_m V_{gs} = g_m V_{in}$$

So input impedance of FET,

$$Z_S = \frac{V_{in}}{I_{in}} = \frac{V_{in}}{g_m V_{in}} = \frac{1}{g_m} \tag{10.58}$$

The input impedance of the circuit is $Z_S \parallel R_S$. So

$$Z_{in} = \frac{1}{g_m} \parallel R_S \tag{10.59}$$

Actually speaking r_d and R_D are also involved in Z_{in} but their effect is negligible.

Output Impedance The output of a common gate circuit is taken from the drain terminal, just as in the case of a common source circuit. So the output impedance at the drain terminal is

$$Z_d = r_d \tag{10.60}$$

Again this equation gives the impedance looking into the device drain terminal. The circuit output impedance

$$Z_{out} = R_D \parallel Z_d = R_D \parallel r_d \tag{10.61}$$

$$V_{out} = I_D (r_d \parallel R_D \parallel R_L) \qquad \text{where } I_D = g_m V_{gs}$$

$$= g_m V_{gs} (r_d \parallel R_D \parallel R_L)$$

$$= g_m V_{in} (r_d \parallel R_D \parallel R_L) \tag{10.62}$$

$\because V_{gs} = V_{in}$

Voltage gain,

$$A_v = \frac{V_{out}}{V_{in}} = g_m (r_d \parallel R_D \parallel R_L) \tag{10.63}$$

Usually $r_d >> R_D \parallel R_L$

So $$A_v = g_m (R_D \parallel R_L) \tag{10.64}$$

The above equation for voltage gain is the same as that in case of a common source circuit (Eq. (10.54)) except that there is no minus sign. It means that output voltage and input voltage are in phase.

10.16 APPLICATIONS OF FETS

FET has a very high input impedance (100 MΩ in case of JFETs and 10^4 to 10^9 MΩ in case of MOSFETs), the major shortcomings of an ordinary transistor i.e., low input impedance with consequent loading of signal source is eliminated in FET. Hence FET is an ideal device for use in almost every application in which transistors can be used.

1. FETs are widely used as input amplifiers in oscilloscopes, electronic voltmeters and other measuring and testing equipment because of their high input impedance.
2. FETs are used in RF amplifiers in FM tuners and communication equipments for the low noise level.
3. FETs are used as voltage variable resistors (VVRs) in operational amplifiers and tone controllers etc. because it is a voltage controlled device.
4. FETs are used in mixer circuits in FM and TV receivers and communication equipments because of their low intermodulation distortion.
5. FETs are used in low frequency amplifiers in hearing aids and inductive transducers because of the small coupling capacitors.
6. FETs are used in digital circuits in computers, LSI and memory circuits because of very small size.

10.17 MOSFET

In the previous sections, we examined the junction field effect transistor (JFET), of which there is only a depletion type. We now look at the other family, the MOSFET, in which, there are two types. Also known as the insulated – gate FET and IGFET, the MOSFET derives its name from the fact that its metal gate is insulated by a very thin oxide layer from the semiconductor channel (Note the underlined letters)

10.17.1 The Depletion MOSFET

As shown by the diagram in Fig. 10.38, there is no *P-N* junction between gate and channel but rather a capacitor, consisting of the metal gate, contact, a dielectric of silicon dioxide and the channel. It is this construction which accounts for the very large input resistance of 10^{10} to 10^{15} ohms and is the major difference from the JFET.

The symbol clearly shows the gate insulated from the channel. Some manufacturers internally connect the bulk to the source, but there are times when a circuit requires the two to be separated. An *N*-channel MOSFET is shown in Fig. 10.38. The symbol clearly shows where the *P-N* junction is

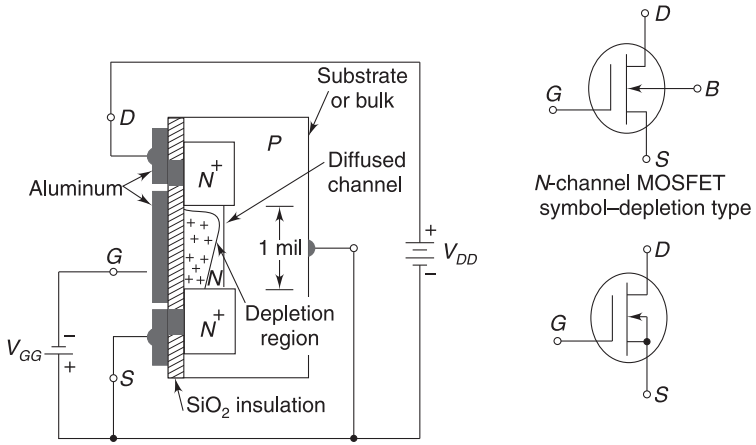


Fig. 10.38 *Depletion type of MOSFET and alternative symbols for N-channel*

located. (A *P*-channel would have the arrow reversed). The drain characteristics and transfer curve for the depletion MOSFET are shown in Fig. 10.39. When $V_{GS} = 0$ V, a significant current flows for a given V_{DS} (like the JFET). When the gate (one plate of the capacitor) is made negative, the other plate, the channel, has a positive charge induced in it opposite the gate. This serves to deplete the channel of majority carriers (electrons) so the conductivity decreases, giving rise to characteristic curves like the JFET. However, unlike the JFET, there is no *P*-*N* junction to keep reverse biased, so the gate is allowed to be made positive. Under this condition, a negative charge is induced in the channel, thereby increasing carriers and *enhancing* current flow.

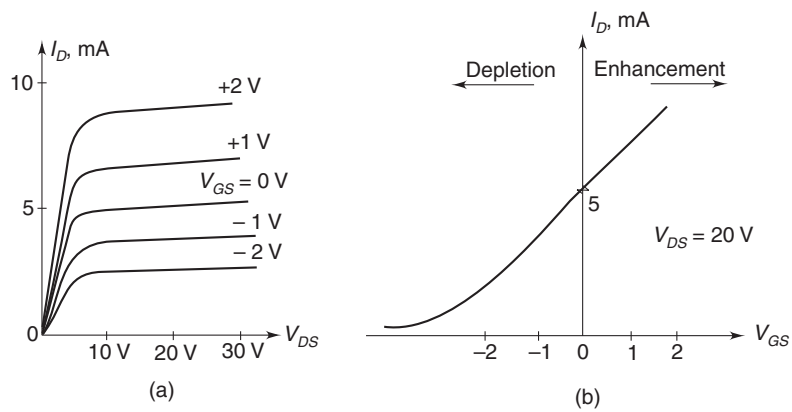


Fig. 10.39 *(a) Drain characteristics for a depletion-type MOSFET; (b) transfer curve for a depletion MOSFET showing how depletion and enhancement operation is possible*

Thus, this device can be operated in depletion and enhancement modes, although it is termed a depletion type to differentiate it from the second MOSFET type that can only work in the enhancement mode (A JFET is inherently a depletion type, but since no enhancement JFET is available, no distinction is necessary).

The depletion MOSFET can thus be operated at zero bias in an amplifier with no change in input resistance on the positive and negative excursions of the input signal voltage.

10.17.2 The Enhancement MOSFET

The enhancement MOSFET differs from the depletion type in that no continuous channel exists between source and drain, with the result that no current flows at zero gate voltage. The symbol's "broken channel" represents this condition, as can be seen in Fig. 10.40.

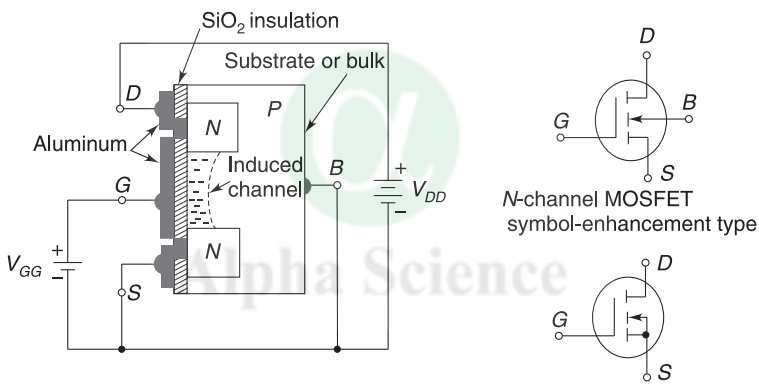


Fig. 10.40 *Enhancement type of MOSFET with positive gate voltage inducing a channel*

The only way in which current can flow from source to drain is for a positive voltage on the gate to *induce* a channel by drawing the minority carrier electrons in the *P*-type bulk into a concentrated layer. As can be seen by the transfer characteristics in Fig. 10.41, a minimum or threshold gate voltage is required before drain current will flow.

This type of FET is very useful in switching applications, since no gate voltage is required to hold it off, and when the device is to be switched on, the application of the same polarity as the drain will achieve this condition. Also, the threshold voltage provides some "noise immunity" to prevent false switching.

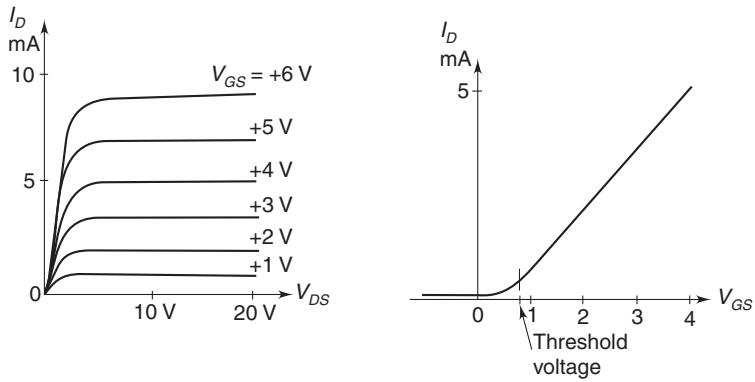


Fig. 10.41 *Typical drain characteristics and transfer curve for an N-channel enhancement type of MOSFET*

10.18 COMPARISON BETWEEN NMOS AND PMOS

1. *P*-channel is much easier and cheaper to produce than the *N*-channel device.
2. The drain resistance of *P*-channel MOSFET is three times higher than that of an identical *N*-channel device.
3. The *N*-channel MOSFET is smaller for the same complexity than the *P*-channel MOSFET.
4. The *N*-channel MOSFET has the higher packing density which makes it faster in switching applications due to the smaller junction areas and low inherent capacitances.
5. A *P*-channel MOSFET occupies a larger area than the *N*-channel MOSFET for the given drain current rating. This is because the electron mobility is 2.5 times the mobility of a hole.
6. The *N*-channel MOSFET has higher false turn-on possibility than the *P*-channel device because of the positively charged contaminants.

MOSFET Handling Caution *MOSFETs are delicate devices and can get readily destroyed. So they are to be handled carefully. Furthermore, they should never be connected or disconnected while the power is ON. Finally, before picking up a MOSFET device, get your body grounded by touching the chasis of the equipment you are working on.*

10.19 COMPARISON BETWEEN JFETS AND MOSFETS

JFETs and MOSFETs are quite similar in their operation principles and in their electrical characteristics. However, they differ in certain aspects, as detailed below on next page:

1. In MOSFETs (D-MOSFETs as well as E-MOSFETs), the transverse electric field induced across an insulating layer deposited on the semiconductor material controls the conductivity of the channel whereas in the JFETs the transverse electric field across the reverse biased *P-N* junction controls the conductivity of the channel.
2. JFETs can only be operated in the depletion mode whereas MOSFETs can be operated in either depletion or in enhancement mode. In a JFET, if the gate is forward biased, excess-carrier injunction occurs and the gate current is substantial. Thus channel conductance is enhanced to some degree due to excess carriers but the device is never operated with gate forward biased because gate current is undesirable.
3. MOSFETs have input impedance much higher than that of JFETs because of small leakage currents. The input impedance of the MOSFET is of the order of 10^{10} to 10^{15} Ω whereas the input impedance of JFET is of the order of 10^{8} Ω.
4. JFETs have characteristic curves more flatter than those of MOSFETs indicating a higher drain resistance.

   The drain resistance of a JFET is of the order of 10^{5} to 10^{6} Ωwhereas the drain resistance of a MOSFET is of the order of 1 to 50 kΩ.
5. MOSFET is very susceptible to overload voltage and needs special handling during installation. It may get damaged easily if not properly handled.
6. Capacitive effects may be considerably lower in dual gate MOSFETs.
7. Special digital CMOS circuits are available which involve near zero power dissipation and very low voltage and current requirements. This makes them most suitable for portable systems.
8. When JFET is operated with a reverse bias on the junction, the gate current I_G is larger than it would be in a comparable MOSFET. The current caused by minority carrier extraction across a reverse biased junction is greater, per unit area, than the leakage current that is supported by the oxide layer in a MOSFET. Thus MOSFET devices are more useful in electrometer application than are the JFETs.

For the above reasons, and also because MOSFETs are somewhat easier to manufacture, they are more widely used than are the JFETs.

Table 10.1 shows comparison between the parameters of JFETs and MOSFETs.

Table 10.1 *Comparison of JFETs amd MOSFETs*

	JFET	MOSFET
g_{fs}	1000 to 25,000 µmhos	1000 to 20,000 µmhos
r_d	0.1 to 1 M ohm	1 to 50 K ohms
I_{GSS}	0.1 to 10 nA	0.1 to 10 pA
r_{gs}	> 10^9 ohms	> 10^{13} ohms
C_{rss}	1 to 4 pF	0.005 to 1 pF

In Table 10.1, g_{fs} is the *transconductance* (similar to the mutual conductance g_m of a tube), given by

$$g_{fs} = \frac{\Delta I_D}{\Delta V_{GS}},$$

V_{DS} being held constant (usually 15 V) and $V_{GS} = 0$ V

We know (as in case of JFET), the device is controlled by a reverse biased *P-N* junction (i.e. gate and source). Thus the gate current (which determines the input resistance of the FET) is understandally small. The *gate cut-off current* is represented by I_{GSS} and for a JFET is a few nanoamperes and a few pico-amperes ($\sim 10^{-12}$ A) for a MOSFET. In Table 10.1 r_{gs} represents the *input resistance from gate to source* and is also referred to as the gate-to-source leakage resistance)

$$r_{gs} = \frac{V_{GS}}{I_{GSS}}$$

(usually at $V_{GS} = -15$ V and $V_{DS} = 0$)

r_d is the *drain resistance* (similar to plate-resistance of a tube) and is defined as

$$r_d = \frac{\Delta V_{DS}}{\Delta I_D},$$

V_{GS} held constant (usually at $V_{GS} = 0$)

10.20 SILICON CONTROLLED RECTIFIER

The silicon controlled rectifier (SCR) is a three terminal semiconductor switching device which is probably the most important circuit element after the diode and the transistor. Invented in 1957, an SCR can be used as a controlled switch to perform various functions such as rectification, inversion and regulation of power flow. The SCR is a unidirectional power switch and is being extensively used in switching d.c. and a.c., rectifying a.c. (to give controlled d.c. output), converting d.c. into ac etc. Thus, SCR has assumed paramount importance in electronics because it can be produced in versions to handle currents up to several thousand amperes and voltages up to more than 1 kV.

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

Why not Germanium Controlled Rectifier?

The device is made of Si because leakage current in silicon is very small as compared to Ge. Since the device is used as a switch, it will carry leakage current in the off condition which should be as small as possible.

10.20.1 Construction

It is a four layered device shown in Fig. 10.42, having a cathode, an anode and a control gate. It provides current only in one direction and hence it can be described as unidirectional, reverse-blocking thyristor. [Thyristors are those solid state devices which have two or more junctions and can be switched from ON to OFF or from OFF to ON, i.e., between two conducting states]. The SCR can be switched from OFF to ON states by a positive trigger pulse of right character applied to the gate.

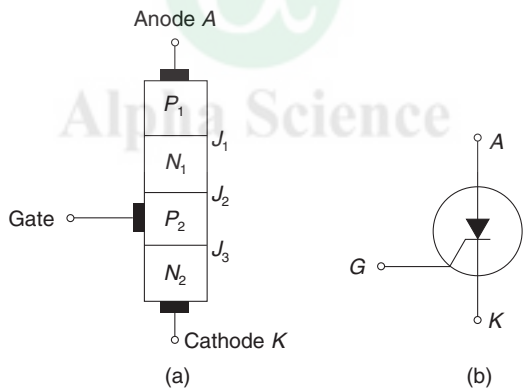


Fig. 10.42 *Silicon controlled rectifier (SCR): (a) structure; (b) symbol*

10.20.2 Operation

The operation of the SCR may be explained by considering it to be divided into two parts, each being a transistor: a *PNP* transistor T_1 and another *NPN* transistor T_2 as shown in Fig. 10.43(a). Figure 10.43(b) shows its electrical equivalent circuit.

First, let it be assumed that the gate terminal G is not connected anywhere and the gate current I_G is zero. A voltage supply V_{SS} is connected in series

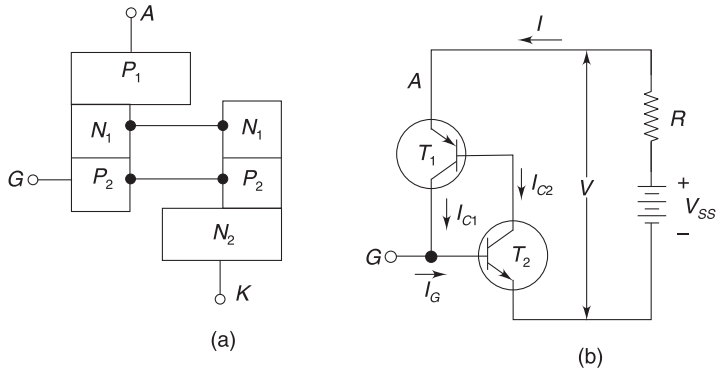


Fig. 10.43 *SCR equivalent circuit*

with a protecting resistor R between the anode and the cathode, as shown in Fig. 10.43(b). The junctions J_1 and J_3 are forward biased and the junction J_2 is reverse biased by the applied voltage V_{SS}. This biasing of the junctions is just correct for the operation of the two transistors T_1 and T_2. The collector-currents of the two transistors are then given by

$$I_{C1} = \alpha_1 I + I_{CO1}$$

and

$$I_{C2} = \alpha_2 I + I_{CO2} \tag{10.65}$$

where α and I_{CO} are the current gain and collector reverse saturation current respectively, pertaining to the two transistors.

The total current I in the anode terminal is then given by

$$I = I_{C1} + I_{C2} = (\alpha_1 + \alpha_2) I + I_{CO1} + I_{CO2}$$

$$I = \frac{I_{CO1} + I_{CO2}}{1 - (\alpha_1 + \alpha_2)} \tag{10.66}$$

Though for an ordinary transistor, the value of α is very near to but less than unity but the SCR is designed so that $(\alpha_1 + \alpha_2)$ is maintained less than unity for normal voltages between anode and cathode, and gate open circuited; the values of α_1 and α_2 being individually less than 0.5. It is the characteristic property of silicon junctions that the value of α in such junctions is quite small (less than 0.5) at very small current levels, and it increases as the current level is increased. Such property is not exhibited by germanium and, hence, the controlled rectifiers are always made of silicon.

The volt-ampere characteristic of a SCR for $I_G = 0$ is shown in Fig. 10.44(a).

At low currents in the region OA of the curves, the value of $(\alpha_1 + \alpha_2) < 1$ and junction J_2 behaves as a reverse bias P-N junction. As the applied voltage

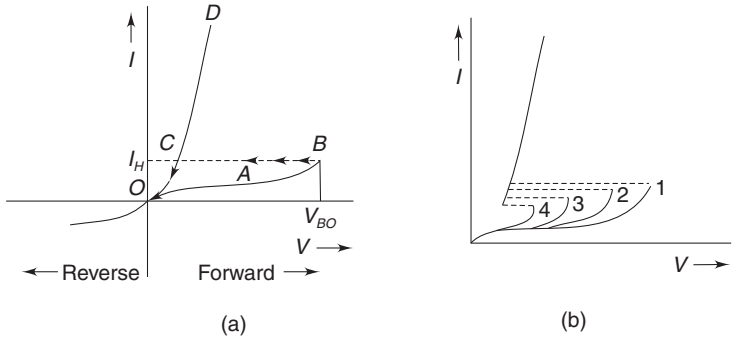


Fig. 10.44 *Volt-ampere characteristic of a SCR for (a) $I_G = 0$, (b) Varying value of I_G; $I_{G1} < I_{G2} < I_{G3} < I_{G4}$*

V is increased, the current slowly increases and the value of $(\alpha_1 + \alpha_2)$ also goes on increasing till it becomes unity at point B corresponding to the voltage V_{BO}, called the *breakover* or *triggering voltage* and current is I_H. The current I increases enormously, so that in practice the current is limited only by the protecting resistance R in the external circuit. The junction J_2 also becomes forward biased with the result that the whole device behaves as a forward-biased PN junction. This part of operation described by region CD in the figure, is associated with high current and low voltage drop.

Once the SCR is fired to bring it in its 'ON' condition, the conduction can be stopped only by reducing the current below the value I_H called the *holding current*. As soon as the current falls below this value, conduction ceases and the operating point goes directly from C to O as indicated.

In the above discussion, the effect of the gate current I_G was not taken into account. Now, suppose that the supply voltage V_{SS} is less than that needed to fire the SCR. If a gate current I_G is now supplied, the junction J_3 becomes more forward biased, and it is possible that the value of α_2 increases to such an extent so as to satisfy the condition $\alpha_1 + \alpha_2 = 1$. The SCR would then fire even with much lower anode voltage by merely introducing small amount of the gate current I_G. This is the principle on which the operation of SCR is based upon. The additional base current I_G is effective in raising α_2 independently of the applied voltage V. As shown in Fig. 10.44(b) higher the value of the triggering gate current I_G, the lower is the value of applied voltage V at which breakover occurs.

Once the SCR has fired, the gate losses its controlling action and the SCR cannot be turned off by gate signals. The only way to bring the SCR back to its normal OFF condition is to reduce the anode voltage sufficiently for anode current to drop below the holding current I_H.

The four-layer devices, which have a control mechanism, are commonly referred to as *thyristors*. However, the term is most frequently applied to SCR's. (Note that SCR has a gate terminal for control purposes)

SCR's find application in regulated power supplies, time-delay circuits, motor controls, inverters, battery chargers, heater controls and phase controls.

Questions

10.1 Describe the construction and working of UJT. Also draw its characteristics.

10.2 Describe the principle of operation of UJT. Describe the intrinsic stand-off ratio with regard to UJT.

10.3 Explain the working principle of UJT. Also explain why UJT has a negative resistance region.

10.4 Explain the working of a UJT relaxation oscillator and derive an expression for frequency of oscillations.

10.5 With the help of neat diagram, describe the construction and operation of a JFET.

10.6 Describe construction, working and characteristics of *N*-channel JFET.

10.7 State the advantages of FET over BJT.

10.8 Prove that the transconductance g_m of a JFET is given by

$$g_m = \frac{2}{|V_P|}\sqrt{I_{DS}\, I_{DSS}}$$

where V_P = pinch-off voltage, I_{DS} = drain current I_{DSS} = maximum I_{DS} corresponding to $V_{GS} = 0$ V.

10.9 Define transconductance (g_m), output resistance (r_p) and gain of a FET and derive the relation between them.

10.10 Explain the meaning of pinch-off voltage in a JFET. How does the current flow in a JFET after pinch-off?

10.11 Discuss the three configuration of FET biasing and explain any one in detail.

10.12 Sketch and explain the small-signal model of an FET (i) at low frequencies and (ii) at high frequencies.

10.13 Draw the basic structure of an *N*-channel JFET. Draw and explain the drain characteristics qualitatively.

10.14 Explain the four distinct regions of the output characteristics of a JFET.

10.15 Explain the difference between enhancement and depletion type MOSFETs.

10.16 Draw and label cross-sectional view of depletion and enhancement type MOSFETs. Draw the drain characteristics of both.

10.17 Give a comparison between NMOS and PMOS.

10.18 Give a comparison between JFETs and MOSFETs.

10.19 Explain the concept of dc load line and bias point on FET characteristics.

10.20 By drawing suitable diagrams, explain the construction and working of a silicon controlled rectifier.

10.21 Draw two transistor model of an SCR and explain its turn-on mechanism.

10.22 Draw V-I characteristics of SCR and explain its modes of operation.

10.23 Write short notes on:

(i) SCR

(ii) Operation of UJT

10.24 BJTs are current operated devices, while FETs are voltage operated. Explain.

11 Modulation and Demodulation

11.1 INTRODUCTION

The speech or music consists of a series of compressions and rarefactions. When a speech or particular music is to be broadcast, the compressions and rarefactions, of which the speech or music is composed of, are translated into a tiny varying electric current with the help of a crystal microphone. The frequency of variations of the electric current produced so lies in the audio range and therefore, it is known as *audio frequency signal*. This low frequency signal is fed to an amplifier, called the *audio frequency amplifier*, so that its strength is raised to the desired level. At audio frequencies (15 Hz – 20 kHz), the signal power is very small and therefore, cannot be fed to the antenna for communication purpose. The radiation of electrical energy is possible only at high frequencies (exceeding 20 kHz). The high frequency signals can be transmitted over long distances (say thousands of kilometers) with comparatively small power. Apparently, for proper transmission of intelligence, some means must be devised which will permit transmission to occur at high frequencies while it simultaneously allows carrying the intelligence contained in the audio frequency signals. This is achieved by superimposing electrical audio signal on a high frequency wave called the *carrier wave* (carrier wave is high frequency undamped radio wave generated by the radio frequency oscillators). The resulting modulated wave (audio frequency signal superimposed on carrier wave) inherits all the characteristics of the audio signal. *The process of superimposition of audio signal over the carrier wave is called the modulation*. After modulation this wave can be fed to the antenna and the intelligence can be transmitted over a long distance. The carrier waves are of constant frequency and travel through space with a velocity of light i.e., 3×10^8 m/s.

The modulated wave i.e., audio frequency signal superimposed upon the carrier wave radiated out from the transmitter antenna after travelling through space strikes the receiving aerial at the receiving end and enters the receiver which separates the audio frequency signal from carrier wave. The audio frequency signal is converted back into the sound wave. The process of separating audio frequency signal from radio frequency carrier wave is known as *detection* or

demodulation. The detection or demodulation of radio frequency modulated wave is essential before it is allowed to pass through the headphones or loudspeakers as high frequency currents have no effect on any ordinary sound producing device. Radio frequency current produces no effect on the human ear too.

11.2 MODULATION

Any wave has three significant characteristics viz., amplitude, frequency and phase, and modulation is a process of impressing information to be transmitted on a high frequency wave, called the carrier wave, by changing its one of the characteristics (amplitude, frequency or phase angle).

Modulation may also be defined as the process of altering some characteristic (amplitude, frequency or phase angle) of the carrier wave in accordance with the instantaneous value of some other wave called the *modulating wave*.

11.2.1 Carrier Wave

Carrier wave is a high frequency, constant amplitude, constant frequency and non-interrupted wave generated by radio frequency oscillators. These waves are inaudible i.e., by themselves they are not able to produce any sound in the loudspeaker. As their name indicates, their function is to carry the audio or video signal from transmitting station to the receiving station. The resulting wave (audio signal superimposed on the carrier wave) is called the *modulated wave*.

11.2.2 Need for Modulation

Low frequency signals cannot be transmitted over long distances if radiated directly into the space. This is because of the following hurdles:

1. **Short Operating Range** The energy of any wave depends on its frequency—the larger the frequency of the wave, the greater the energy associated with it. Obviously the audio signal having small frequency and consequently small power cannot be transmitted over long distance when radiated directly into the space. However, modulated wave (audio signal with high frequency) can be transmitted over long distances.
2. **Poor Radiation Efficiency** At audio frequencies radiation is not practicable as efficiency of radiation is poor. However, electrical energy can be radiated efficiently at high frequencies (above 20 kHz).
3. **Mutual Interference** If low frequency signals are transmitted directly from different sources, all of them would be mixed up and completely blanket the air. However, by modulation, different messages of different frequency levels can be transmitted simultaneously without any interference.

4. **Huge Antenna Requirement** For efficient radiation of a signal, the length of transmitting and receiving antenna should be at least one quarter wavelength i.e.,

$$l = \frac{1}{4} \text{ wavelength} = \frac{1}{4} \frac{\text{Velocity}}{\text{frequency in Hz, } f} \text{ metres}$$

$$= \frac{\frac{1}{4} \times 3 \times 10^8}{f} = \frac{75 \times 10^6}{f} \text{ metres}$$

Thus, for transmitting a signal of frequency 2 kHz, an antenna of length 37.5 km will be required, practically impossible. On the other hand, for transmitting a signal of frequency 2 MHz, an antenna of about 37.5 metres would be required which can be easily constructed.

The carrier wave alone also cannot be employed for transmission of message, through it has high frequency and enormous energy. The reason is that an unmodulated carrier wave has all of its parameters constant (constant amplitude, constant frequency and constant phase relationship with respect to some reference) while any message has changing quantities.

Hence, the solution lies in modulation that enables a low frequency signal transmission over long distances through space with the help of a high frequency carrier wave. These carrier waves need reasonably sized antenna and produce no interference with other transmitters operating in the same area.

11.2.3 Types of Modulation

The sinusoidal carrier wave may be represented as

$$v_c = V_c \sin(\omega_c t + \theta) = V_c \sin(2\pi f_c t + \theta)$$

where V_c is the maximum value, f_c is the frequency and θ is the phase relation w.r.t. some reference of the carrier wave. Since there are three variables (viz. amplitude, frequency and phase) in the above equation, there are three types of sine-wave modulations known as *amplitude modulation* (*AM*), *frequency modulation* (*FM*) and *phase modulation* (*PM*).

In *amplitude modulation*, the amplitude of the carrier wave is varied in accordance with the modulating signal, keeping the frequency and phase of the carrier wave unchanged.

In *phase modulation*, the phase of the carrier wave is varied in accordance with the modulating signal, keeping the amplitude and frequency of the carrier wave unchanged.

However, the modulation may also be classified, according to the nature of carrier wave, into continuous wave modulation and pulse modulation.

In India, amplitude modulation and frequency modulation are employed in radio broadcasting while in television transmission frequency modulation is employed for audio signals and amplitude modulation for video signals. Hence, here our discussion will be limited to the amplitude and frequency modulations.

1. **Continuous Wave Modulation** When the carrier wave is continuous in nature, the modulation process is called the **continuous wave modulation** or **analog modulation**. Amplitude modulation and angle modulation (frequency modulation and phase modulation) fall under this category.
2. **Pulse Modulation** When the carrier wave is a pulse-type waveform, the modulation process is referred to as **pulse modulation**. In pulse modulation, the carrier wave consists of a periodic sequence of rectangular pulses. Pulse modulation may be of an analog or digital-type.

In analog pulse modulation, the amplitude, duration, or position of a pulse is varied in accordance with sample values of the message signal. The analog pulse modulation, therefore, may be of three types *viz*., pulse-amplitude modulation (PAM), pulse-duration modulation (PDM) and pulse-position modulation (PPM).

On the other hand, the digital form of pulse modulation is known as **pulse-code modulation (PCM)**.

11.3 AMPLITUDE MODULATION

The process of varying amplitude of the high frequency or carrier wave in accordance with the intelligence (code, voice or music) to be transmitted, keeping the frequency and the phase of the carrier wave unchanged, is known as the *amplitude modulation*. In this process the modulating signal (intelligence) is superimposed upon the radio frequency carrier by applying both to a nonlinear impedance (e.g. the modulator). The principle of amplitude modulation is illustrated in Fig. 11.1. In this process, the amplitudes of both positive and negative half cycles of carrier wave are varied in accordance with the intelligence. The carrier then consists of sine waves whose amplitudes follow the amplitude variations of the modulating wave. The carrier is contained in an envelope formed by the modulating wave. From Fig. 11.1, it is obvious that the amplitude variations of the carrier wave is at the signal frequency and the frequency of the amplitude modulated wave is the same as that of the carrier wave.

11.3.1 Analysis and Frequency Spectrum of Amplitude Modulated Carrier Wave

Let the carrier and modulating voltage waves be represented as

$$v_c = V_c \sin \omega_c t \tag{11.1}$$

and

$$v_m = V_m \sin \omega_m t \tag{11.2}$$

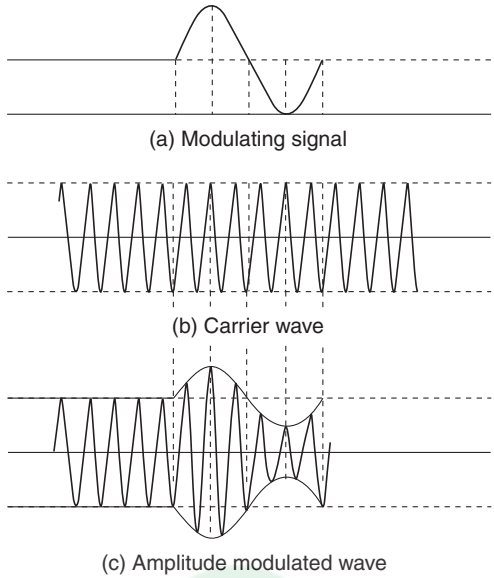


Fig. 11.1 *Principle of amplitude modulation*

where v_c, V_c and ω_c are the instantaneous value, peak value and angular velocity of the carrier and v_m, V_m and ω_m are the instantaneous value, maximum value and angular velocity of the modulating signal. Phase angle has been ignored in both equations as it remains unchanged in amplitude modulation process.

The amplitude of the carrier wave varies at a modulating signal frequency f_m. The amplitude of amplitude modulated wave is given as

$$A = V_c + v_m = V_c + V_m \sin \omega_m t = V_c \left[1 + \frac{V_m}{V_c} \sin \omega_m t\right]$$

$$= V_c (1 + m \sin \omega_m t) \tag{11.3}$$

where m is the ratio of peak values of modulating signal and carrier wave and is known as *modulation index*. Its value is restricted between zero and unity.

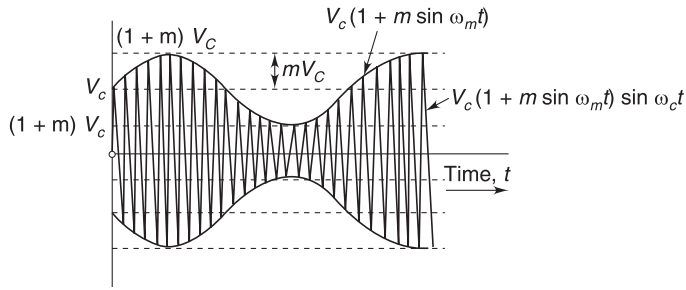
The instantaneous value of amplitude modulated wave is given by the following equation

$$v = A \sin \omega_c t$$

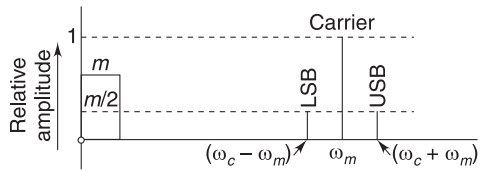
$$= V_c (1 + m \sin \omega_m t) \sin \omega_c t$$

$$= V_c \sin \omega_c t + mV_c (\sin \omega_m t \sin \omega_c t)$$

$$= V_c \sin \omega_c t + \frac{mV_c}{2} \cos (\omega_c - \omega_m) t - \frac{mV_c}{2} \cos (\omega_c + \omega_m) t \tag{11.4}$$



(a) Amplitude Modulated Sine wave with $m < 1$



(b) Frequency Spectrum of a Sinusoidally Modulated wave $m < 1$

Fig. 11.2

Inspection of Eq. (11.4) reveals that the amplitude modulated wave is equivalent to the summation of three sinusoid: one having amplitude V_c and frequency $\frac{\omega_c}{2\pi}$, the second having amplitude $\frac{mV_c}{2}$ and frequency $\frac{\omega_c - \omega_m}{2\pi}$ and the third having amplitude $\frac{mV_c}{2}$ and frequency $\frac{\omega_c + \omega_m}{2\pi}$. In practical radio transmission, ω_c may be many times greater than ω_m. Hence the frequency of the second and third terms on the right hand side of Eq. (11.4) is generally close to the carrier frequency. Figure 11.3 represents this situation graphically on the *frequency-spectrum* plot. The frequency components contained in the amplitude-modulated wave are shown by vertical lines approximately located along the frequency axis. The height of each vertical line is drawn in proportion to its amplitude. The lower frequency component $\frac{\omega_c - \omega_m}{2\pi}$ is called the *lower side frequency*. The upper frequency component $\frac{\omega_c + \omega_m}{2\pi}$ is called the *upper side frequency*. The amplitude of each side frequency is $\frac{mV_c}{2}$. The amplitude of sideband frequencies can never exceed half the carrier amplitude because $m \not> 1$.

Hence it can be concluded that during the process of amplitude modulation

(i) the original carrier frequency is not altered but two new frequencies $\frac{\omega_c + \omega_m}{2\pi}$ and $\frac{\omega_c - \omega_m}{2\pi}$, known as *sideband frequencies*, are produced. $\frac{\omega_c + \omega_m}{2\pi}$ is termed the upper sideband (USB) and $\frac{\omega_c - \omega_m}{2\pi}$ is termed as the lower sideband (LSB). For example, if the carrier is of 1 MHz and the signal is of 2 kHz then the modulated signal contains carrier of 1 MHz and upper sideband of 1,002 kHz and lower sideband of 998 kHz.

(ii) the carrier voltage component does not transmit any information because the signal frequency $\frac{\omega_m}{2\pi}$ is contained in sidebands only which is evident from Fig. 11.2(b).

(iii) in amplitude modulated wave the bandwidth is from $\frac{\omega_c - \omega_m}{2\pi}$ to $\frac{\omega_c + \omega_m}{2\pi}$ i.e., $\frac{2\omega_m}{2\pi}$ or twice the signal frequency. For example, if the frequency of modulating signal is 2 kHz, the total bandwidth of the modulated wave is 4 kHz, and this bandwidth must be passed by all transmitting and receiving circuits. When a carrier wave is amplitude modulated by a single frequency, two sideband frequencies are produced. When a modulating signal consists of more than one frequency, two sideband frequencies are produced. When a modulating signal consists of more than one frequency, two sideband frequencies are produced by every frequency. If the modulating signal contains two frequencies, four side band frequencies are produced, i.e., two upper side and two lower sideband frequencies.

The number of sideband frequencies higher than the carrier frequency are the same as the number of frequencies lower than it. All the sideband frequencies above the carrier frequency make up the USB and all those below the carrier frequency forms the LSB. The frequencies in the USB represent the sum of the individual modulating frequencies and the carrier frequency while those in LSB represent the difference between the modulating frequencies and the carrier frequency. The total bandwidth can be expressed in terms of the highest modulating frequency and is equal to twice this frequency.

11.3.2 Modulation Index (*m*)

The extent to which the amplitude of the carrier wave is varied by the modulating signal is called the *degree of amplitude modulation* or *modulation index* and is represented by *m*. Thus the ratio of change in amplitude of carrier wave to the amplitude of normal carrier wave is called the *modulation index*, *m*.

$$\text{i.e., } m = \frac{V_m \text{ (amplitude change of carrier wave)}}{V_c \text{ (amplitude of normal or unmodulated carrier wave)}} \tag{11.5}$$

The modulation index is a number lying between 0 and 0.8. When it is expressed as percentage, then it is called the *percentage modulation*.

The percentage modulation of a modulated carrier is determined by dividing the change in amplitude by the amplitude of the unmodulated carrier, multiplied by 100

$$\text{i.e., } \qquad \% \text{ modulation} = \frac{V_m}{V_c} \times 100 \tag{11.6}$$

Modulation index may also be defined in terms of the values referred to the modulated carrier wave and is given as

$$m = \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}}$$

where $V_{\max}$ and $V_{\min}$ are the maximum and minimum values of the amplitude of the modulated carrier wave. This is worked out as below.

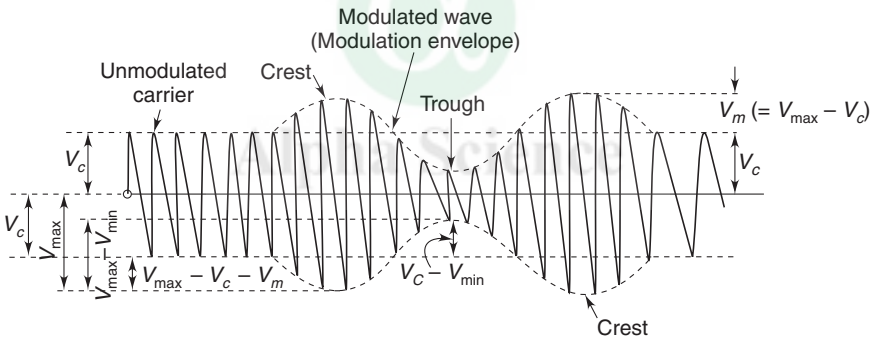


Fig. 11.3 *Amplitude modulated radio-frequency carrier wave*

Consider the modulated carrier wave shown in Fig. 11.3.

$$2V_m = V_{\max} - V_{\min}$$

or
$$V_m = \frac{V_{\max} - V_{\min}}{2} \tag{11.7}$$

and
$$V_c = V_{\max} - V_m$$

$$= V_{\max} - \frac{V_{\max} - V_{\min}}{2} = \frac{V_{\max} + V_{\min}}{2} \tag{11.8}$$

Dividing Eq. (11.7) by Eq. (11.8), we get

$$m = \frac{V_m}{V_c} = \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}} \tag{11.9}$$

For a *sinusoidal* modulating signal, where the modulation is symmetrical about the carrier, the modulation index is also given by the relations:

$$m = \frac{V_{\max} - V_{\min}}{2V_c} = \frac{V_{\max} - V_c}{V_c} = \frac{V_c - V_{\min}}{V_c} \tag{11.10}$$

Amplitude modulated waves for different values of '*m*' are shown in Fig. 11.4.

(a) *Case I* Smallest value of $m = 0$ i.e., when the amplitude of the modulating signal is zero. It means that $m = 0$ for an unmodulated carrier wave.

(b) *When the signal amplitude is one-half of the carrier amplitude, as shown in Fig. 11.4(a)*. In this case, the amplitude of the carrier wave varies between 1.5 A and 0.5 A where *A* is the amplitude of normal (or unmodulated) carrier wave.

So the change in the amplitude of carrier

$$= 1.5\text{ A} - \text{A} = 0.5\text{ A}$$

and $$m = \frac{0.5\text{ A}}{\text{A}} = 0.5 \text{ or } 50\%$$

The carrier is said to be 50% modulated.

(c) *When the signal amplitude equals the carrier amplitude, as shown in Fig. 11.4(b)*. In this case, the amplitude of carrier wave varies between 2 A and zero.

So the change in the amplitude of carrier = 2 A – A = A

and $$m = \frac{\text{A}}{\text{A}} = 1.0 \text{ or } 100\%$$

The carrier wave is 100% modulated.

(d) *When the signal amplitude is more than that of the carrier*. In such a case the modulation is called the *over-modulation* because *m* exceeds unity. In this case, the amplitude of the trough of the wave reaches zero amplitude and remains there for a finite time period, as illustrated in Fig. 11.4(c). Modulation in excess of 100% is undesirable because it produces severe distortion and interference, called spatter, in the transmitter output.

Modulation index '*m*' is a very important factor, as if determines the strength and quality of the transmitted signal. In an amplitude modulated wave, the signal

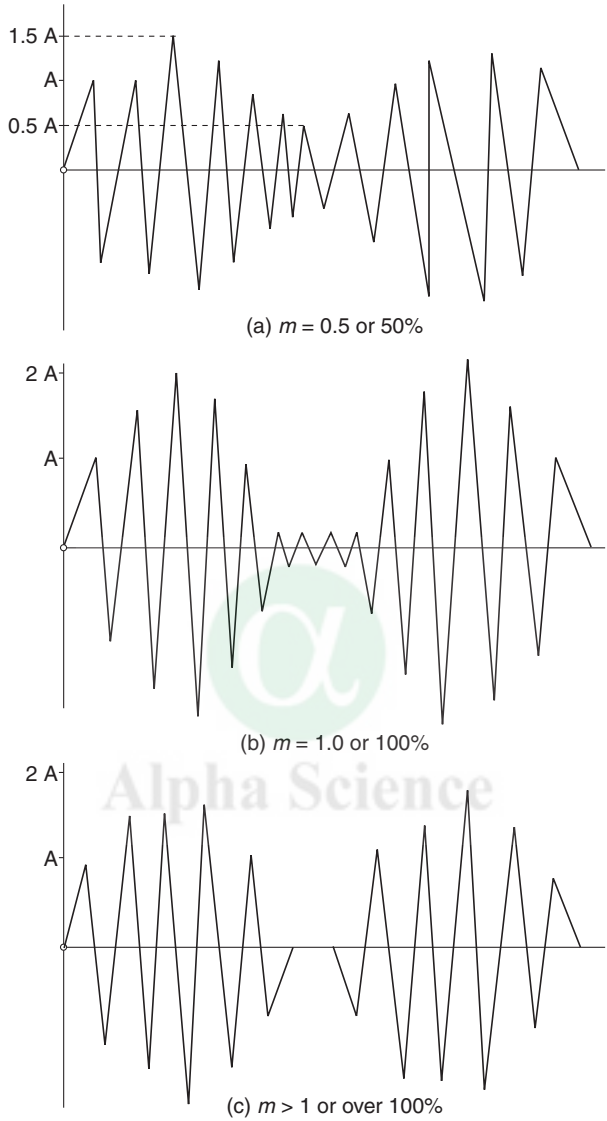


Fig. 11.4 *Degree of modulation*

is contained in the variations of the carrier amplitude. When the carrier wave is modulated to a small degree (i.e. $m << 1$), the amount of carrier amplitude variation is small and, therefore, the audio signal transmitted is weak. With the increase in value of modulation index, the audio signal becomes stronger and stronger. But if m exceeds unity, severe distortion and interference are produced in the transmitter output. Maximum undistorted power of radio transmission is obtained for unity modulation index (i.e. $m = 1$).

11.3.3 Power Relations in Amplitude Modulated Wave

From Eq. (11.4) it is obvious that the carrier component of the amplitude modulated wave has the same amplitude as the unmodulated carrier. However, two sideband components are also present in the modulated wave. Thus, the modulated wave has more power than that had by the carrier wave before modulation. Now, total power in the modulated wave is given as

$$P_{\text{total}} = P_{\text{carrier}} + P_{\text{LSB}} + P_{\text{USB}} \tag{11.11}$$

When an amplitude modulated wave is impressed upon some resistance (say antenna resistance) R then

$$P_{\text{carrier}} = \left(\frac{V_c/\sqrt{2}}{R}\right)^2 = \frac{V_c^2}{2R} \tag{11.12}$$

Each sideband has peak value of $\frac{m}{2} V_c$ and rms value of $\frac{m}{2} \frac{V_c}{\sqrt{2}}$. Hence power in each sideband

$$\begin{aligned} P_{\text{LSB}} &= P_{\text{USB}} \\ &= \frac{\left(\frac{m}{2} \frac{V_c}{\sqrt{2}}\right)^2}{R} = \frac{m^2 V_c^2}{8R} \\ &= \frac{m^2}{4} \cdot \frac{V_c^2}{2R} \\ &= \frac{m^2}{4} P_{\text{carrier}} \end{aligned} \tag{11.13}$$

Substituting values of P_{carrier}, P_{LSB} and P_{USB} from Eqs. (11.12) and (11.13) in Eq. (11.11) we have

$$\begin{aligned} P_{\text{total}} &= \frac{V_c^2}{2R} + \frac{m^2}{4} \cdot \frac{V_c^2}{2R} + \frac{m^2}{4} \cdot \frac{V_c^2}{2R} \\ &= \frac{V_c^2}{2R}\left(1 + \frac{m^2}{2}\right) = P_{\text{carrier}}\left(1 + \frac{m^2}{2}\right) \end{aligned} \tag{11.14}$$

Conclusions

1. From above Eq. (11.14) it is obvious that the maximum power in the amplitude modulated wave (without distortion) will occur for $m = 1$

   i.e., $P_{\text{total}} = 1.5\ P_{\text{carrier}}$ when $m = 1$

2. The ratio of P_{SB} and T_{total} is given as

$$\frac{P_{SB}}{P_{total}} = \frac{\frac{m^2}{4} \cdot P_{carrier} + \frac{m^2}{4} \cdot P_{carrier}}{P_{carrier}\left[1 + \frac{m^2}{2}\right]} = \frac{\frac{m^2}{2}}{1 + \frac{m^2}{2}} \qquad (11.15)$$

For $m = 1$, $\quad \frac{P_{SB}}{P_{total}} = \frac{1/2}{3/2} = \frac{1}{3}$

i.e., only one-third of the total power of the modulation wave is contained in the two sidebands and rest of the two-third power lies in the carrier component, which is of no use.

3. In most applications, carrier is simultaneously modulation by several sinusoidal modulating signals. In such a case the total modulation index is given as

$$m_t = \sqrt{m_1^2 + m_2^2 + m_3^2 + \ldots} \qquad (11.16)$$

4. If I_c and I_t represent the rms values of unmodulated (or carrier) current and total modulated current and R is the resistance through which these current flow, then

$$\frac{P_{total}}{P_{carrier}} = \frac{I_t^2 R}{I_c^2 R} = \left(\frac{I_t}{I_c}\right)^2 \qquad (11.17)$$

But from Eq. (11.14),

$$\frac{P_{total}}{P_{carrier}} = 1 + \frac{m^2}{2}$$

So $\quad \left[\frac{I_t}{I_c}\right]^2 = 1 + \frac{m^2}{2}$

or $\quad I_t = I_c \sqrt{1 + \frac{m^2}{2}} \qquad (11.18)$

11.3.4 Limitations of Amplitude Modulation

Amplitude modulation suffers from the following drawbacks:

1. **Low Efficiency** In amplitude modulation, useful power that lies in the sidebands, is quite small, so the efficiency of AM system is low.
2. **Limited Operating Range** Transmitters employing amplitude modulation have small operating range. This is due to low efficiency. Hence information cannot be transmitted over long distances.

3. **Noise Reception** In case of AM, the reception is generally noisy. This is because a radio receiver cannot distinguish between the amplitude variations that represent noise and those contain the desired signal.
4. **Poor Audio Quality** In order to attain high-fidelity reception, all audio frequencies up to 15 kHz must be reproduced and this necessitates the bandwidth of 30 kHz while the AM broadcasting stations are assigned bandwidth of only 10 kHz to minimize the interference from the adjacent broadcasting stations. Therefore, in AM broadcasting stations audio quality is usually poor.

Example 11.1 *A sinusoidal carrier wave of frequency 1 MHz and amplitude 100 V is amplitude modulated by a sinusoidal voltage of frequency 5 kHz producing 50% modulation. Calculate the frequency and amplitude of USB and LSB.*

Solution Frequency of carrier,

$$f_c = 1 \text{ MHz or } 1{,}000 \text{ kHz}$$

Frequency of modulating signal,

$$f_m = 5 \text{ kHz}$$

Modulation index,

$$m = 50\% \text{ or } 0.5$$

Lower sideband frequency

$$= f_c - f_m$$

$$= 1{,}000 - 5 = 995 \text{ kHz } \textbf{Ans.}$$

Upper sideband frequency

$$= f_c + f_m$$

$$= 1{,}000 + 5 = 1{,}005 \text{ kHz } \textbf{Ans.}$$

Amplitude of each sideband

$$= \frac{mV_C}{2} = \frac{0.5 \times 100}{2}$$

$$= 25 \text{ V } \textbf{Ans.}$$

11.4 TYPES OF AMPLITUDE MODULATION

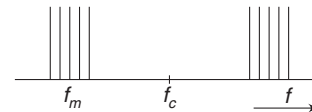
As already discussed in Section 11.3.1, one carrier and two sidebands are produced in AM generation. However, it is not essential that all these signals are transmitted to enable the receiver to reconstruct the original signal. One may attenuate or altogether remove the carrier or any of the sidebands without affecting the communication process.

Accordingly there are 3 possible suppressed component systems viz. DSBSC (double sideband suppressed carrier) system, SSBTC (single sideband transmitted carrier) system and SSBSC (single sideband suppressed carrier) system.

In **DSBSC system**, as its name indicates, carrier component is altogether removed resulting in saving of enormous amount of power. From Section 11.3.3, it is known that carrier signal contains two-third of total transmitted power for 100% modulation.

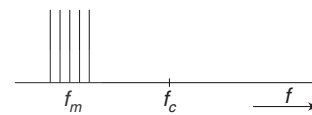
In **SSBTC system**, one sideband is suppressed and the carrier and other sideband are transmitted, as shown in Fig. 11.5(b). In this case one-sixth of total transmitted power is saved for 100% modulation.

SSBSC system consists in transmitting only one sideband and suppress the other sideband and the carrier, as shown in Fig. 11.5(c). It utilizes the fact that the intelligence or message is contained in each sideband and not in the carrier. Two sidebands being the exact images of each other, carry the same audio intelligence. Thus, all information is available in one sideband only, and one sideband along the carrier can be discarded with no loss of intelligence.



(a) DSBSC System

(b) SSBTC System



(c) SSBSC System

Fig. 11.5

However, for demodulation purpose at the receiving end, carrier is necessary. Hence in this system, carrier is re-inserted at the receiving end in proper phase, frequency and amplitude. The received sideband and the re-inserted carrier are then mixed in a demodulator to get back the original signal.

The advantages of SSBSC (or SSB) system are given below:

1. Smaller operating cost.
2. 9 to 12 dB improvement in signal/noise power ratio in the reproduced output at the receiving end. This is because SSB signal has narrower bandwidth, therefore, a narrower passband is permissible within the receiver, thereby limiting the noise pick-up.
3. Small sized power supply requirement—very important in a spacecraft.
4. Half bandwidth per channel multiplexing twice as many channels in a given frequency range.
5. 83.3% saving in total transmitted power.
6. Removal of distortion due to selective fading.
7. Reduction in carrier interference with other stations.
8. Some privacy automatically provided.

Of course, the receiver for the SSB system gets complicated and the use of SSB system is, therefore, limited to radio telephony. The standard or double-sideband full-carrier (DSBFC) transmission system is widely used in broadcasting because of its relative simplicity of its modulating equipment.

There are various methods of suppressing carrier and one sideband. However, the carrier is suppressed always by employing some form of balanced modulator. For suppression of one sideband in SSB system, the following three methods are commonly employed.

1. Filter method
2. Phase shift method
3. Weaver's method

11.5 VESTIGIAL SIDEBAND (VSB) MODULATION SYSTEMS

A vestigial sideband modulation system is actually a compromise between DSBSC and SSB modulation systems. In other words, it can be said that it is an optimum choice in which the advantages of DSBSC and SSB modulation systems have been exploited.

As a matter of fact, the generation of VSB modulation signals is easier than other modulated signals such as conventional AM, DSBSC and SSB signals. Its bandwidth is only slightly higher (approximately 25%) than SSB signals but considerably less than DSBSC signals. SSB modulation is rather most suited for the transmission of voice signals because of the energy gap that exists in the frequency spectrum of the voice signals between zero and few hundred Hz. On the other hand, when signals contain frequency components at extremely low frequencies (as in telegraph signals) the USB and LSB of the translated signal tend to meet at the carrier frequency. Under such circumstances, it becomes very difficult to isolate one sideband from the other. Hence, SSB scheme becomes unsuitable for handling such types of signals.

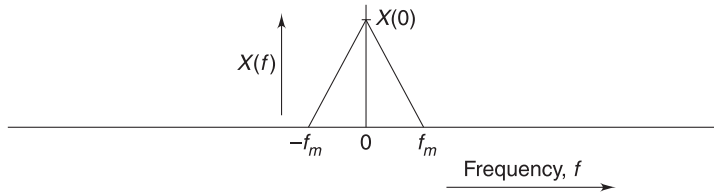
This difficulty has been overcome in a scheme known as VSB modulation. In VSB modulation instead of rejecting one sideband completely as in SSB modulation scheme, a gradual cutoff of one sideband is allowed. This gradual cut is compensated by a vestige or portion of the other sideband.

The technique has been illustrated in Fig. 11.6 which shows the frequency spectrum of the modulating signal and corresponding DSBSC, SSB and VSB signals.

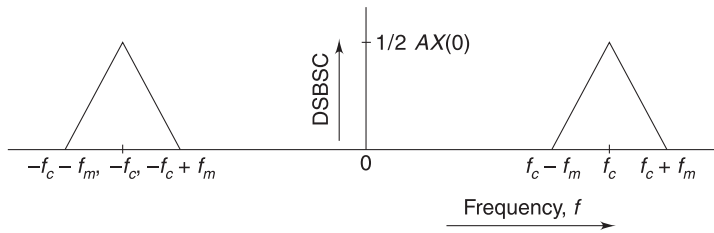
Obviously, the bandwidth of VSB signal is given by

$$BW = f_c + f_v - f_c + f_m = f_m + f_v \tag{11.19}$$

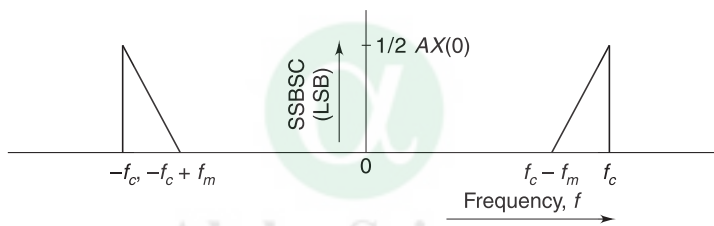
where f_m is the BW of the message signal, and f_v is BW of the VSB.



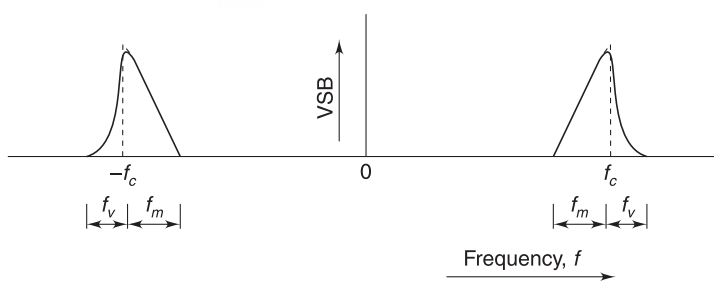
(a) Frequency spectrum of baseband signal $x(t)$



(b) Frequency spectrum of DSBSC signal



(c) Frequency spectrum of SSB signal



(d) Frequency spectrum of VSB signal

Fig. 11.6 *Illustration of frequency spectrum of VSB signal*

11.6 COMPARISON OF VARIOUS AM SYSTEMS

The relative merits and demerits of various forms of AM can be summarized below:

1. The demodulation or detection of AM signal is simpler than that of DSBSC and SSB systems. The conventional AM can be demodulated by rectifier

or envelope detector. Detection of DSBSC and SSB is rather difficult and expensive. Furthermore, it is quite easier to generate conventional AM signals at high power levels as compared to DSBSC and SSB signals. That is why conventional AM systems are used for broadcasting purpose.

2. The advantage of DSBSC and SSB systems over conventional AM system is that the former needs lesser power for transmission of given information. For sinusoidal modulation the carrier consumes about 2/3 of the total power for 100% modulation is conventional AM. However, only 1/3 of the total power is carried by sidebands which carry the information. This makes AM transmitters less efficient. On the other hand, the receivers of DSBSC and SSB systems, though efficient are much more complex and expensive too and, therefore, only find applications in point-to-point communication. In point-to-point communication only a few receivers and one transmitter are required. In public broadcast system, one transmitter caters to millions of receivers. It is obviously required to be simpler and cheaper.
3. SSB scheme requires only one-half of the bandwidth required in DSBSC system and less than that required in VSB system. Thus, it can be said that SSB modulation scheme is the most efficient among DSBSC and VSB schemes. SSB modulation scheme is used for long distance transmission of voice signals because it allows longer spacing between repeaters.

11.7 BASIC PRINCIPLE OF AM GENERATION

In amplitude modulation, the modulating circuit performs an operation on the radio frequency in such a way that the amplitude of the RF voltage and current in the circuit are changed (modulated) in accordance with an information signal (audio or video, for example). The result is that the envelope of the RF signal is a replica of the information signal. The modulating circuit is typically an amplifier. To accomplish this effect, the gain of the amplifier must be varied. The gain must be increased momentarily to obtain the increase in amplitude of the modulated signal. An amplifier with nonlinear transfer characteristics provides the desired operating feature. The modulating circuit is generally a nonlinear RF amplifier. The modulating signal shifts the operating point between points of higher gain and lower gain.

11.8 BLOCK DIAGRAM OF AM (AMPLITUDE MODULATOR) TRANSMITTER

Block diagram of typical amplitude modulation transmitter is shown in Fig. 11.7.

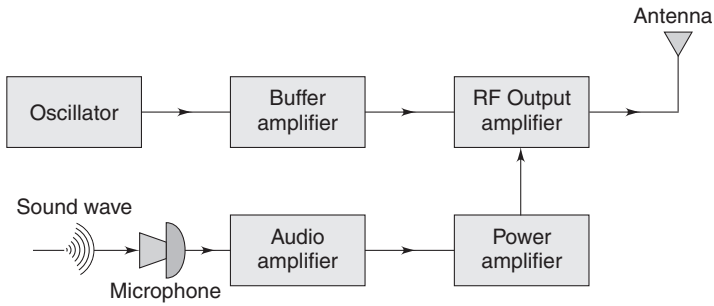


Fig. 11.7 *Block diagram of a AM transmitter*

Sound waves produced by speech or music strike the diaphragm of a microphone that converts them into a tiny varying current. The audio frequency output of the microphone is amplified by a low level audio amplifier and, finally, by a power amplifier.

On the other side, the carrier waves of radio frequency are generated by a crystal-controlled oscillator. These carriers waves are applied to a tuned buffer amplifier and finally to an RF output amplifier. Buffer amplifier is used to isolate the RF oscillator and the power amplifier stages.

The amplified audio signal is then combined with the carrier to give a modulated carrier wave which is fed to the transmitter antenna.

11.9 TRANSISTOR AM MODULATORS

Low power AM transmitters usually employ transistors for generating carrier power up to a few hundred watts. Three types of linear modulation are possible: (1) collector modulation (2) base modulation and (3) emitter modulation.

11.9.1 Collector Modulation

This is very popular method for AM generation. The circuit diagram for collector modulation is shown in Fig. 11.8. Here, the transistor Q_1 makes an RF class C amplifier. The carrier signal is applied to the base of Q_1. V_{CC} makes collector supply used for biasing purpose. Also, the transistor Q_2 makes a class B amplifier which is used to amplify the audio or modulating signal. The modulating or baseband signal appears across the modulation transformer after amplification. The amplified modulating signal appears in series with collector supply V_{CC}. The function of capacitor C is to offer low impedance path to high frequency carrier signal and, therefore, the carrier signal is prevented from flowing through the modulation transformer.

Operating Principle It is a known fact that in a class C amplifier, the magnitude of output voltage is a definite fraction of or at the most equal to supply voltage V_{CC}.

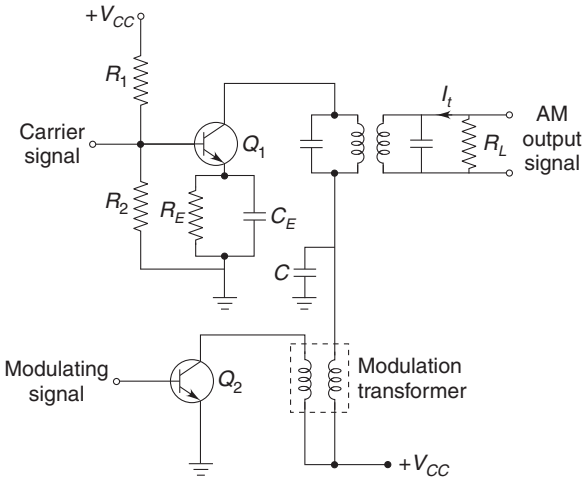


Fig. 11.8 *Collector modulation circuit*

In addition to this, a linear relationship exists between the output tank current I_t and the variable supply voltage V_c (here it has been assumed that supply voltage V_{CC} is a varying quantity and its varying value is denoted by V_c). This means that in class *C* amplifier, the output voltage will be an exact replica of the input voltage waveform and the magnitude of output voltage will be approximately equal to carrier supply voltage V_{CC}. Now if R is the resistance of the output tank circuit at resonance, then the magnitude of output voltage is given as

$$RI_t \approx V_{CC}$$

So, the unmodulated carrier is amplified by class *C* modulation amplifier using transistor Q_1 and its magnitude will remain constant at V_{CC} since there appears no voltage across the modulation transformer in the absence of modulating signal.

But now if a modulating voltage $v_m = V_m \cos \omega_m t$ appears across the modulation transformer, this signal will be added to the carrier supply voltage V_{CC}. This results in a quite slow variation in carrier supply voltage V_{CC}. This type of slow variation in carrier supply voltage changes the magnitude of the carrier signal voltage and the output of the modulated class *C* amplifier as shown in Fig. 11.9.

It may be observed that the envelope of the output voltage is identical with the modulating voltage and thus AM signal is produced.

Mathematical Analysis The slowly varying carrier supply voltage V_c may be expressed as

$$\begin{aligned} V_c &= V_{CC} + v_m \\ &= V_{CC} + v_m \cos \omega_m t \end{aligned} \qquad (11.20)$$

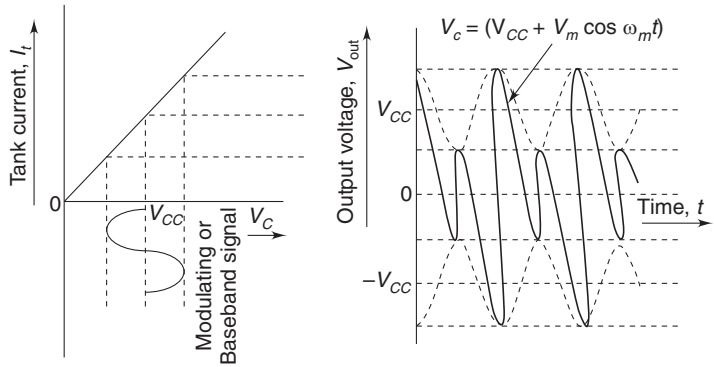


Fig. 11.9

$$= V_{CC} + mV_{CC} \cos \omega_m t$$

$$= V_{CC} (1 + m \cos \omega_m t) \qquad (11.21)$$

$$\therefore \text{ Modulation index } \quad m = \frac{\text{Maximum modulating voltage}}{\text{Maximum carrier voltage}}$$

$$= \frac{V_m}{V_{CC}}$$

Let the carrier voltage be

$$v_C = V_{CC} \cos \omega_c t$$

Then the modulated output voltage will be

$$v_{\text{out}} = V_c \cos \omega_c t$$

Substituting the value of V_c in above equation from Eq. (11.21), we have

$$v_{\text{out}} = V_{CC} (1 + m \cos \omega_m t) \cos \omega_c t \qquad (11.22)$$

which is required expression for AM wave.

The collector modulation is commonly used because of its following advantages:

1. Good linearity.
2. High collector circuit efficiency.
3. High power output per transistor.

However, it suffers from the following drawbacks:

1. Large modulating power requirement.
2. Hundred percent modulation cannot be achieved because of collector saturation.

11.9.2 Base Modulation

The circuit for base modulation is shown in Fig. 11.10. The carrier signal from a crystal oscillator is coupled to the base terminal of the amplifier through transformer T_1. The modulating signal is applied across resistor R_1 through capacitor C_1. The capacitor C_1 allows only audio frequency a.c. signal to pass through and blocks the dc components. The modulating signal develops a voltage across R_1, which, in turn, alters the quiescent point in the base circuit. With the increase in bias current, the signal at the collector terminal is correspondingly strengthened. Similarly, as the bias current falls during the reverse part of the cycle of the modulating signal, the output at the collector terminal is also reduced. Consequently, a plot of the collector output will be found to give a variation similar to that shown in Fig. 11.1. The resistor R_E is the self-bias emitter resistor and C_E is its bypass capacitor. Also, R_2 is a dropping resistor which limits the dc voltage drop across R_1, which, in turn, establishes the initial quiescent base current. Capacitor C_2 is a bypass capacitor for the carrier signal in the input circuit. In understanding the operation of this circuit, it is helpful to keep in mind the fact that ω_m is very small in comparison to ω_c. Hence for any given cycle of the carrier frequency the bias in the base-to-emitter circuit appears to be fixed. The capacitor C_3 is selected to provide series resonance with the primary winding of transformer T_2. The amplitude-modulated carrier of RF current in primary of transformer T_2 induces similar current in the secondary of transformer T_2. Thus, the amplitude modulated wave is radiated out in the space by transmitting antenna.

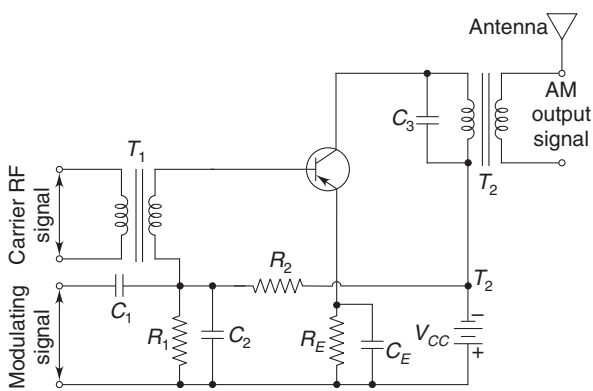


Fig. 11.10 *Base modulation circuit*

Special Features of Base Modulation

1. Power required from this modulator is small in comparison to that of collector modulator.

2. Comparatively low power output and poor efficiency. The power output and efficiency suffer because the unmodulated collector current peaks can be only about half as large as in the collector modulated circuit.
3. Critical adjustment of the base modulated amplifier.
4. High degree of linearity is more difficult to obtain.

Base modulation is used in TV transmission because it needs little power and can meet the power need of large bandwidth.

11.9.3 Emitter Modulation

The circuit for emitter modulation is shown in Fig. 11.11. It is essentially a CE amplifier. The carrier signal from a crystal oscillator is coupled to the base of amplifier through capacitor C_{in}. The modulating signal is applied to the emitter circuit. The capacitors C_{in} and C_C are so chosen as to by-pass carrier frequency only. The transformer is employed to provide better matching of the impedance. The voltage divider $R_1 - R_2$ and emitter resistance R_E provide the proper biasing. The tuned circuit at the collector of the modulator stage eliminates the undesirable signals. The capacitor C' keeps the RF out of power supply V_{CC}.

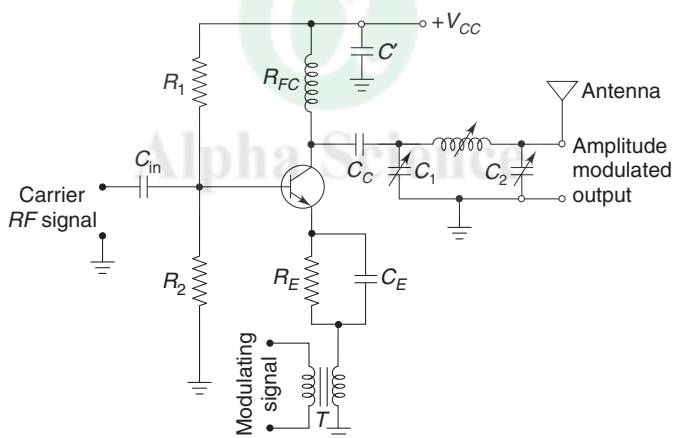


Fig. 11.11 *Emitter modulation circuit*

The modulating signal, being a part of the biasing circuit, produces low frequency variations in the emitter circuit. The result is that amplitude of the carrier varies in accordance with the strength of the modulating signal. The amplitude modulated output is obtained at the collector terminal of the amplifier, as shown in Fig. 11.11.

The characteristics of emitter modulation lies in between those of base modulation and collector modulation.

11.10 SQUARE LAW DIODE MODULATOR

Basic circuit of a square law diode modulator is shown in Fig. 11.12. It utilizes the nonlinear region of voltage-current dynamic characteristic of a *P-N* diode. This method is suited at low voltage levels because of the fact that current-voltage characteristics of a diode is highly nonlinear, particularly in the low voltage region, as shown in Fig. 11.13.

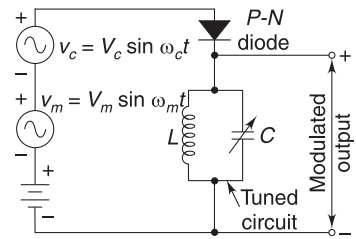


Fig. 11.12 *Basic circuit of square law diode modulator*

As shown in Fig. 11.12, carrier and modulating signals are simultaneously applied across the diode. D.C. supply V_{CC} is connected across the diode to provide a fixed operating point on the *v-i* characteristic of diode. The square law diode modulator operates on the principle that when two different frequencies are simultaneously passed through a non-linear device, such as diode, the process of amplitude modulation occurs. Hence, when carrier and modulating frequencies are simultaneously applied at the input of diode, then different frequency terms appear at the output of diode. These different frequency terms are applied across a tuned circuit which is tuned to the carrier frequency and has a narrow bandwidth just to pass two sidebands along with the carrier and reject other frequencies. Thus, at the output of the tuned circuit, carrier and two sidebands are obtained i.e., generation of AM wave.

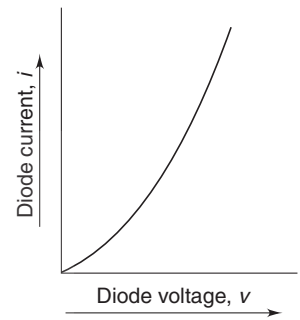


Fig. 11.13 *V-i characteristic of P-N-diode*

The ac current i may be expressed as a function of ac voltage v_S as given by the following Taylor series.

$$i = a_1 v_S + a_2 v_S^2 \tag{11.23}$$

where a_1 and a_2 are the Taylor series coefficients

and $$v_S = V_c \sin \omega_c t + V_m \sin \omega_m t \tag{11.24}$$

Substituting the value of v_S from Eq. (11.24) in Eq. (11.23), we have

$$i = a_1 V_c \sin \omega_c t + a_1 V_m \sin \omega_m t + a_2 [V_c^2 \sin^2 \omega_c t + 2 V_c V_m \sin \omega_c t \sin \omega_m t + V_m^2 \sin^2 \omega_m t]$$

The quantity in bracket derives from the square term in Eq. (11.23). A little thought reveals that it is this term which gives rise to product of the carrier

and modulating sinusoids, which, in turn, yields the upper and lower frequency sidebands thereby permitting the amplitude modulated signal to be identified.

Substituting $\sin^2 \omega_c t = \dfrac{1 - \cos 2\,\omega_c t}{2}$ in above equation we have

$$i = \frac{a_2}{2}\left(V_c^2 + V_m^2\right) + a_1\, V_m \sin \omega_m t - \frac{a_2}{2}\, V_c^2 \cos 2\,\omega_c t$$

$$- \frac{a_2}{2}\, V_m^2 \cos 2\,\omega_m t + a_1\, V_c \sin \omega_c t$$

$$+ \frac{a_2\, V_c\, V_m}{2} \cos (\omega_c - \omega_m)\, t - \frac{a_2\, V_c\, V_m}{2} \cos (\omega_c + \omega_m)\, t \qquad (11.25)$$

The last three terms of Eq. (11.25) are identical in form to those appearing in Eq. (11.4), so that together they yield in the output tuned circuit the desired amplitude modulated signal. The first four terms are of no use and can be filtered out by tuning *L-C* circuit to frequency ω_c. The output current is given as

$$i_{\text{out}} = a_2\, V_c\, (1 + m \cos \omega_m t) \cos \omega_c t \qquad (11.26)$$

11.11 GENERATION OF DSBSC SIGNAL

DSBSC signal can be obtained by simply multiplying the modulating signal with the carrier signal. By simple multiplication of $V_c \sin \omega_c t$ and $V_m \sin \omega_m t$, we have the lower and upper sidebands without carrier as shown below:

$$v_c \times v_m = V_c \sin \omega_c t \times V_m \sin \omega_m t$$

$$= \frac{V_c\, V_m}{2} \times 2 \sin \omega_c t \sin \omega_m t$$

$$= \frac{V_c\, V_m}{2} [\cos (\omega_c - \omega_m)\, t - \cos (\omega_c + \omega_m)\, t] \qquad (11.27)$$

It means that a DSBSC signal is obtained by simply multiplying modulating signal ($V_m \sin \omega_m t$) with carrier signal ($V_c \sin \omega_c t$). This is achieved by a product modulator. The block diagram of a product modulator is depicted in Fig. 11.14.

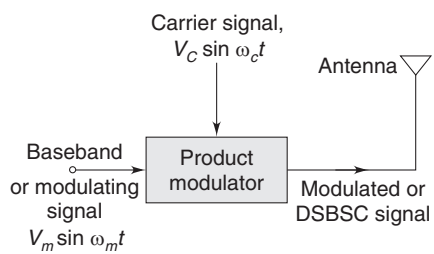


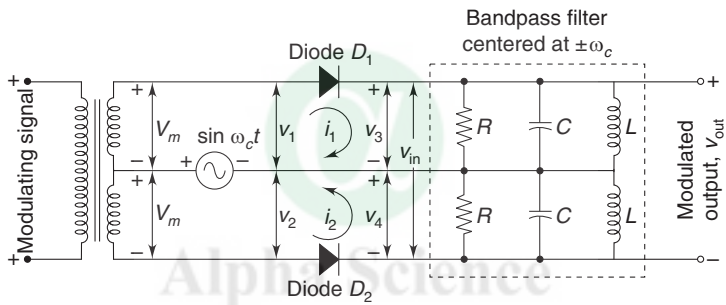
Fig. 11.14

Here we will discuss two types of product modulators namely the balanced and ring modulator.

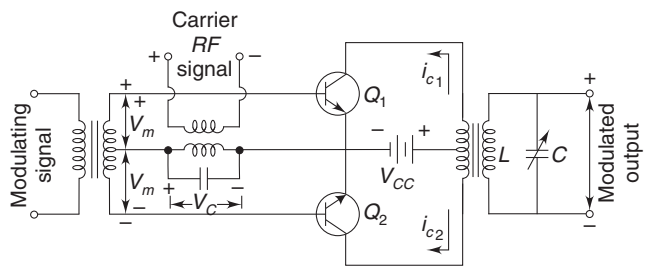
11.11.1 Balanced Modulation

We know that a non-linear resistance or a nonlinear device may be used to provide amplitude modulation, i.e., one carrier and two sidebands. However, a DSBSC signal contains only two sidebands. Thus, if two nonlinear devices such as diodes, transistors, etc. are connected in a balanced mode so as to suppress the carriers of each other, then only sidebands are left, i.e., a DSBSC signal is produced. Thus, a balanced modulator may be defined as a circuit in which two non-linear devices are connected in a balanced mode to generate a DSBSC signal.

Two balanced modulator circuits using diodes and transistors as nonlinear devices are shown in Figs. 11.15(a) and 11.15(b).



(a) Balanced modulator using diodes



(b) Balanced modulator using transistors

Fig. 11.15

It is assumed that both diodes and transistors used in circuits shown in Fig. 11.15 have identical characteristics and that the circuit is symmetrical with respect to the centre tap of the transformer. As the two halves have to be matched or balanced, such a modulator is known as a **balanced modulator**.

The modulating signal $V_m \sin \omega_m t$ is applied in push-pull and the carrier *RF* signal $V_c \sin \omega_c t$ is put in parallel to a pair of identical diodes D_1 and D_2 (Fig. 11.15(a)) or two matched transistors Q_1 and Q_2 (Fig. 11.15(b)). In transistor circuit the carrier *RF* signal is thus applied to the bases of the two transistors in phase, while the modulating signal appears 180° out of phase at the bases, as they are at the opposite ends of a centre-tapped transformer.

A bandpass filter used in diode circuit (Fig. 11.15(a)) is that type of filter which allows to pass a particular band of frequencies only. Here since the bandpass filter is centred around $\pm\omega_c$, it will pass a narrowband of frequencies centred at $\pm\omega_c$ with a small bandwidth of $2\omega_m$ to preserve the sidebands.

The modulated output currents of the two transistors Q_1 and Q_2 are combined in the centred-tapped primary of the push-pull output transformer. These currents, therefore, are opposite and subtract as indicated by the direction of arrows in the figure. For a symmetrical arrangement, the carrier frequency will be completely cancelled.

It may be proved that the output signal (or modulated signal) contains only sidebands with carrier component completely eliminated as below.

The expression for the modulated wave generated by transistor Q_1 may be given as

$$\begin{aligned} v_1 &= V_c (1 + m \sin \omega_m t) \sin \omega_c t \\ &= V_c \sin \omega_c t + mV_c \sin \omega_m t \sin \omega_c t \end{aligned}$$

The modulated wave generated by transistor Q_2 will be the same as the generated by Q_1 except that it will have a phase difference of 180° (i.e., reversed phase). So

$$v_2 = V_c \sin \omega_c t - mV_c \sin \omega_m t \sin \omega_c t$$

At the output transformer, the output voltage v_{out}, due to push-pull arrangement, is proportional to $(v_1 - v_2)$, so

Modulated output,

$$\begin{aligned} v_{out} &= v_1 - v_2 \\ &= 2m\, V_c \sin \omega_m t \sin \omega_c t \\ &= mV_c [\cos (\omega_c - \omega_m)\, t - \cos (\omega_c + \omega_m)\, t] \end{aligned}$$

Thus, it is proved that the output signal consists of only sidebands and the carrier component is eliminated.

11.12 SINGLE SIDEBAND GENERATOR

SSB can be obtained by passing the output of a carrier suppressor (product modulator) through filter circuits that are selective to transmit one sideband while

suppressing the other. However, this method is suitable only for obtaining SSB corresponding to low or moderate carrier frequencies. The other method is the phase shift method which avoids filter and instead uses two balanced modulators and two phase shifting networks. These two methods are described as below:

11.12.1 Frequency Discrimination Method

In a frequency discrimination method, firstly, a DSBSC signal is generated simply by using an ordinary product modulator or a balanced modulator. After this, from the DSBSC signal one of the two sidebands is filtered out by a suitable bandpass filter. The schematic diagram for this method is depicted in Fig. 11.16.

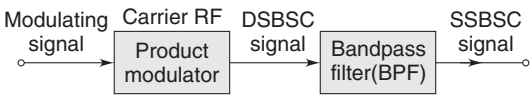


Fig. 11.16 *Frequency discrimination method for SSBSC generation*

In fact, the design of bandpass filter is quite critical and thus puts some limitations on the modulating or baseband and carrier frequencies. The frequency discrimination method for SSBSC generation have got following limitations.

1. This method is useful only if the baseband signal is restricted at its lower edge due to which the upper and lower sidebands are non-overlapping. For example, the filter method is used for speech communication where lowest spectral component is 70 Hz and it may be takes as 300 Hz without affecting the intelligibility of the speech signal. However, this system is not useful for video communication where the baseband signal starts from d.c. (0 Hz).
2. The another restriction of this method is that the baseband signal must be appropriately related to the carrier frequency. In fact, the design of the bandpass filter (BPF) becomes difficult if the carrier frequency is quite higher than the bandwidth (BW) of the baseband signal.

11.12.2 Phase Shift Method

This method uses two balanced modulators and two phase shifting networks as shown in Fig. 11.17. This method avoids the uses of filters. In Fig. 11.17, one of the modulators M_1 receives the carrier voltage shifted by 90° and the modulating voltage, whereas another balanced modulators M_2 receives the modulating voltage shifted by 90° and carrier voltage. Both balanced modulators give an output consisting only of sidebands.

Both the upper sidebands lead the input carrier voltage by 90°. One of the lower sideband leads reference voltage by 90° and the other lags behind it by

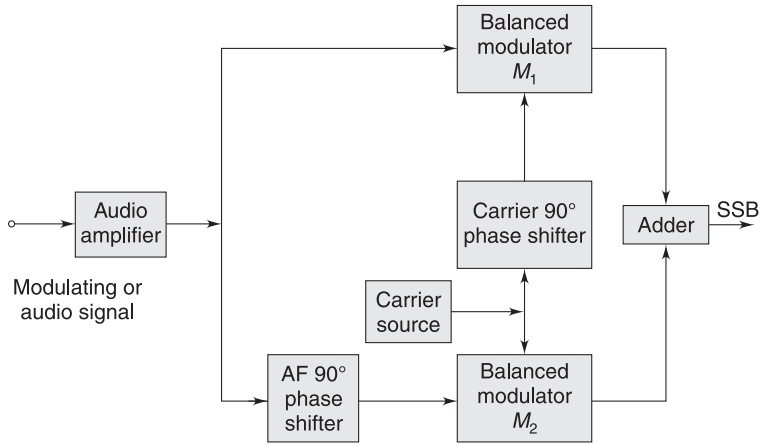


Fig. 11.17 *Phase-shift method for SSB SC generation*

90°. The two lower sidebands are thus out of phase, and when combined together in the adder, they cancel each other. The upper sidebands are in phase at the adder and therefore, they add producing SSB in which the lower sideband has been cancelled.

Mathematical Analysis Let the expressions for modulating signal and carrier signal be sin $\omega_m t$ and sin $\omega_c t$ respectively.

Now the balanced modulator M_1 will receive sin $\omega_m t$ and sin $(\omega_c t + 90°)$ and modulator M_2 will receive $(\omega_m t + 90°)$ and sin $\omega_c t$.

We know that the output of balanced modulator M_1 will contain sum and difference frequencies.

Hence
$$v_1 = \cos [(\omega_c t + 90°) - \omega_m t] - \cos [(\omega_c t + 90°) + \omega_m t]$$
$$= \cos [(\omega_c t - \omega_m t + 90°) - \cos (\omega_c t + \omega_m t + 90°)]$$
LSB USB

Similarly, the output of balanced modulator M_2, will contain

$$V_2 = \cos [\omega_c t - (\omega_m t + 90°)] - \cos [\omega_c t + (\omega_m t + 90°)]$$
$$= \cos (\omega_c t - \omega_m t - 90°) - \cos (\omega_c t + \omega_m t + 90°)$$

Thus, the output of adder will be

$$v_{out} = v_1 + v_2 = 2 \cos (\omega_c t - \omega_m t + 90°) \quad \ldots \text{Resulting in LSB}$$
$$= |v_1 - v_2| = 2 \cos (\omega_c t + \omega_m t + 90°) \quad \ldots \text{Resulting in USB}$$

The extent of SB suppression depends upon (i) balancing capacity of the modulator (ii) control accuracy of the 90° phase of the carrier waves and (iii) error in maintenance of 90° shift between the two modulating waves.

The phase shift method was introduced to overcome the bulk of filter system using *L-C* filters. This method has the advantages of (i) ease of switching from one SB to another (ii) ability of generating SSB at any frequency and (iii) use of low audio frequencies for modulation.

Drawbacks In case the phase shifter provides a phase change other than 90° at any audio frequency then that particular frequency will not be completely removed from the unwanted SB. Also, in case the two modulators are not identical, cancellation will not be complete.

11.13 FREQUENCY MODULATION

As described earlier frequency modulation is produced when the frequency of carrier is varied inaccordance with the modulating signal, while the amplitude of the carrier remains constant.

11.13.1 Characteristic of FM Wave

The amount of deviation in frequency is proportional to the amplitude of the signal, while the rate is determined by the frequency of modulation source. For example, suppose 100 mc. Carrier is being modulated by a sinusoidal signal and for given modulation system there is 2 kc., frequency shift per volt of signal magnitude. Then, if the modulating signal has an amplitude of 20 volt and frequency of 2 kc., the instantaneous frequency of modulated wave will vary between 99.96 mc., and 100.04 mc. at a rate of 2000 times per second.

Figure 11.18 illustrates the form of a frequency modulated wave when a sinusoidal modulating voltage is employed.

11.13.2 Analysis of FM Wave

The modulating signal can be written as

$$e_m = E_m \cos \omega_m t \tag{11.28}$$

where E_m is amplitude of signal (peak) and ω_m is frequency of signal voltage.

The general expression for an unmodulated carrier is given by

$$e_c = E_c \sin (\omega_c t + \theta) \tag{11.29}$$

where quantity $(\omega_c t + \theta) = \phi(t)$ is the total instantaneous 'phase' of function;

so
$$e_c = E_c \sin \phi(t) \tag{11.30}$$

The instantaneous angular frequency is

$$\omega_i = \frac{\partial \phi(t)}{\partial t} = \frac{\partial}{\partial t} (\omega t + \theta) = \omega_c.$$

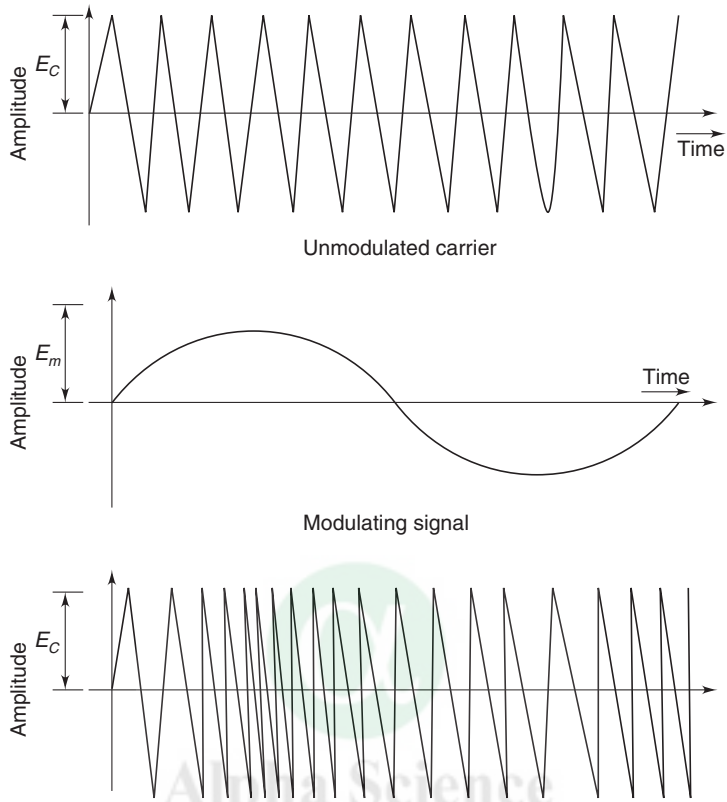


Fig. 11.18 *A frequency modulated wave*

By definition, the instantaneous angular frequency of frequency modulated wave can be written as

$$\omega_I = \omega_c + K_f E_m \cos \omega_m t \tag{11.31}$$

where proportionally factor K_f determines the maximum variation in frequency for a given signal strength E_m. Since

$$\omega_i = \frac{\partial \phi(t)}{\partial t},$$

we have

$$\phi(t) = \int_0^t \omega_i \, dt = \int_0^t (\omega_c + K_f E_m \cos \omega_m t) \, dt$$

$$= \omega_c^+ + K_f \frac{E_m}{\omega_m} \sin \omega_m t + \theta_0 \tag{11.32}$$

The initial phase θ_0 is neglected as it plays no part in modulation process.

Thus FM wave can be written as

$$e = E_c \left(\sin \omega_c + K_f \frac{E_m}{\omega_m} \sin \omega_m t \right) \tag{11.33}$$

The instantaneous frequency of fm wave is

$$f = \frac{\omega}{2\pi} = f_c + K_f \frac{E_m}{2\pi} \cos \omega_m t.$$

Therefore, $$f_{\max} = f_c + K_f \frac{E_m}{2\pi}$$

and $$f_{\min} = f_c - K_f \frac{E_m}{2\pi}.$$

$\therefore$ Frequency deviation

$$\Delta f = f_{\max} - f_c = f_c - f_{\min} = \frac{K_f E_m}{2\pi} \tag{11.34}$$

The instantaneous frequency can then be expressed as

$$f = f_c + \Delta f \cos \omega_m t.$$

The modulation index =

$$m_f = \frac{\text{frequency deviation}}{\text{modulating frequency}} = \frac{\Delta f}{f_m} = \frac{K_f E_m}{\omega_m} \tag{11.35}$$

And therefore the expression for fm wave becomes

$$e = E_c \sin (\omega_c t + m_f \sin \omega_m t) \tag{11.36}$$

In fm broadcasting system, Δf has a maximum permissible value of 75 kc/s and f_m may vary from about 33 c/s to 15 kc/s. Consequently, m_f can have values both greater than unity or less than unity whereas in amplitude modulation m_a should not exceed unit for distortionless transmission. Variation in transmitted frequency is shown in Fig. 11.19.

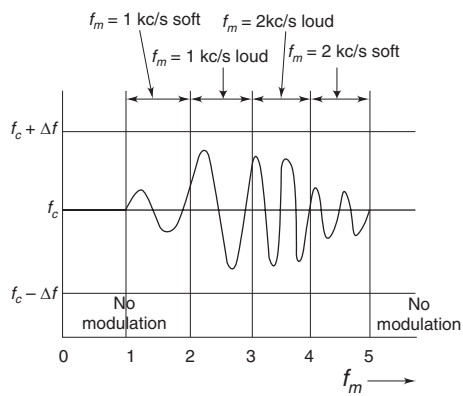


Fig. 11.19 *Variation of the transmitted frequency for an FM wave*

11.13.3 Power Relation in FM Wave

In amplitude modulated wave, intelligence is contained in side frequencies. The amplitude and power level of the side frequencies increase with percent modulation and consequently, an increase in modulation level also increases the power level of modulated wave. On the other hand, in frequency modulated wave, as the amplitude is not affected by modulation, the *total power* in the modulated wave does not change, regardless of the degree of modulation. It is also obvious from the operation of reactance tube, where it has been shown that the effect of modulating signal is to change the capacitance in oscillator tank circuit so as to cause a shift Δf in frequency of the oscillator, the modulating signal does not change the power output of the oscillator.

11.13.4 Frequency Spectrum of FM Wave

The expression of FM wave can written as

$$e = E_c \left[\sin \omega_c t \,.\, \cos (m_f \sin \omega_m t) + \cos \omega_c t \sin (m_f \sin \omega_m t)\right] \quad (11.37)$$

Evaluating above expression with the help of Bessel's expression,

$$\begin{aligned} e &= E_c \sin \omega_c t \left[J_0 (m_f) + 2I_2 (m_f) \cos 2\omega_m t + 2J_4 (m_f) \cos 4\, \omega_m t \ldots\right] \\ &\quad + E_c \cos \omega_c t \left[2\, J_1 (m_f) \sin \omega_m t + 2J_3 (m_f) \sin 3\omega_m t + \ldots\right] \\ &= J_0 (m_f)\, E_c \sin \omega_c t + J_1 (m_f)\, E_c \left[\sin (\omega_c + \omega_m)\, t - \sin (\omega_c - \omega_m)\, t\right] \\ &\quad + J_2 (m_f)\, E_c \left[\sin (\omega_c + 2\omega_m)\, t - \sin (\omega_c - 2\omega_m)\, t\right] \\ &\quad + J_3 (m_f)\, E_c \left[\sin (\omega_c + 3\omega_m)\, t - \sin (\omega_c - 3\omega_m)\, t\right] \end{aligned} \quad (11.38)$$

It is clear from above Exq. (11.38) that the spectrum of fm wave consists of carrier and an infinite number of side bands. The side frequency components extend above and below the carrier by an amount ω_m, $2\omega_m$, $3\omega_m$ etc. The side frequency pair differing from carrier frequency by ω_m is known as the first order side frequencies and that differing by $2\omega_m$ as second order side frequencies etc.

Table 11.1 *Bessel functions of first kind*

x	n or order								
m_f	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8
0.00	1.00								
0.5	0.94	0.24	0.03						
1.0	0.77	0.44	0.11	0.02					
1.5	0.51	0.56	0.23	0.06	0.01				
2.0	0.22	0.58	0.35	0.13	0.03				

Contd..

Contd..

2.5	– 0.05	0.50	0.45	0.22	0.07	0.02			
3	– 0.26	0.34	0.49	0.31	0.13	0.04	0.01		
4	– 0.40	– 0.07	0.36	0.43	0.28	0.13	0.05	0.02	
5	– 0.18	– 0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02

The amplitudes of the different frequency components of FM wave determined by the values of various order Bessel functions, depend upon the value of m_f. A graph of several of these functions is drawn in Fig. 11.20.

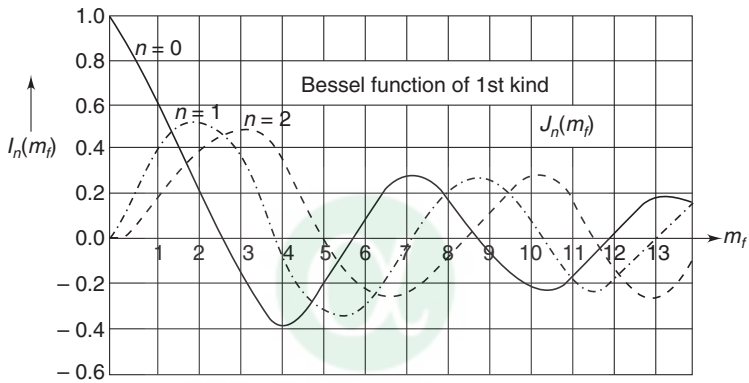


Fig. 11.20

It is to be noted from the graph that $J_0\ (m_f)$ at $m_f = 2.4$ has a root. This mean that carrier will vanish when frequency deviation $\Delta f = 2.4$. Bessel functions of the first kind for m_f values up to 5 are given in table.

11.13.5 Bandwidth of FM Waves

It is defined as the width of frequency spectrum that contains all components having an amplitude equal to or greater than one percent of the amplitude of unmodulated wave (carrier). Suppose, a given wave has n significant side frequency pairs then bandwidth is equal to

$$^{*}BW = 2nf_m \tag{11.39}$$

The relation follows from the fact that each side frequency pair extends above and below the carrier by an *equal* amount and the spacing in the spectrum

* For nth side frequency pair, one component of the pair will have frequency $(\omega_c + n\omega_m)$ and another $(\omega_c - n\omega_m)$ so that total width of frequency spectrum for this pair will be

$$= (\omega_c + n\omega_m) - (\omega_c - n\omega_m) = 2n\omega_m.$$

between components of successive order is equal to modulation frequency f_m. Further, from Eq. (11.35),

$$f_m = \frac{\Delta f}{m_f},$$

Therefore, $$BW = 2n \frac{\Delta f}{m_f}$$

or $$\frac{BW}{2\Delta f} = \frac{n}{m_f} \tag{11.40}$$

which predicts that ratio of bandwidth to the total frequency deviation depends upon the number of significant side-frequency pairs(n) obtained with a given modulation index (m_f). For values $m_f > 10$ it is seen that the ratio $\frac{n}{m_f}$ tends to approach unity and the bandwidth is almost equal to or slightly larger than $2\Delta f$. For narrow band frequency modulation, modulation index is kept at a low value by employing small frequency deviations. Moreover, the number of significant side frequency pairs, n, increases as the modulation index, m_f, of the wave is increased. Therefore, m_f is kept low for narrow band transmission.

Suppose frequency of modulating signal f_m is doubled, then, though Δf will remain unaffected, but m_f will be halved (Eq. 11.35).

Reduction in the value of m_f will, in its turn, reduce the value of n, the number of significant side frequency pairs. If the reduction in the value of n and m_f is by like amounts, bandwidth will remain constant even after doubling the modulating frequency.

Example 11.2 *Find the bandwidth needed for a frequency modulated signal if the frequency deviation is 60 kc/s and the modulating frequency is 5 kc/s.*

Solution Frequency deviation

$$\Delta f = 60 \text{ kc/s}.$$

Modulating frequency

$$f_m = 5 \text{ kc/s}.$$

Then modulation index will be

$$m_f = \frac{\Delta f}{f_m} = \frac{60}{5} = 12$$

From Bessel table corresponding to $m_f = 12$, it can be found that there are sixteen significant side frequency pairs (i.e. the highest J coefficient for $m_f = 12$ is J_{16}). That is, $n = 16$, so that

$$BW = 2nf_m = 2 \times 16 \times 5 \text{ kc/s.} = 160 \text{ kc/s.}$$

11.14 COMPARISON OF FM AND AM

The frequency modulation (FM) has the following advantages over the amplitude modulation (AM):

1. FM receivers may be fitted with amplitude limiters to remove the amplitude variations caused by noise. This makes FM reception a good deal more immune to noise than AM reception.
2. It is possible to reduce noise still further by increasing the frequency deviation. This is a feature which AM does not have because it is not possible to exceed 100 percent modulation without causing severe distortion.
3. Standard frequency allocations provide a guard band between commercial FM stations. Due to this, there is a less adjacent-channel interference than in AM.
4. FM broadcasts operate in the upper VHF and UHF frequency ranges at which there happens to be less noise than in the MF and HF ranges occupied by AM broadcasts.
5. The amplitude of FM wave is constant. It is thus independent of the modulation depth whereas in AM, modulation depth governs the transmitted power. This permits the use of low level modulation in FM transmitter and use of efficient class *C* amplifiers in all stages following the modulator. Further since all amplifiers handle constant power, the average power handled equals the peak power. In AM transmitter, the maximum power is four times the average power. Finally, in FM, all the transmitted power is useful whereas in AM, most of the power is carrier power which does not contain any information.

However, the FM have some drawbacks over AM which are given below:

1. FM transmitting and receiving equipments particularly used for modulation and demodulation are more complex and more costly.
2. A much wider channel typically 200 kHz is required in FM as against only 10 kHz in AM broadcast. This forms serious limitation of FM.
3. In FM, the area of reception is small as it is limited to only one line of sight.

11.15 PHASE MODULATION

Definition In the process of phase modulation, the instantaneous phase of the carrier is varied by an amount that is proportional to the instantaneous amplitude of the signal and at a rate that is proportional to modulating frequency. The amplitude of carrier remains unaltered.

Analysis The modulating signal has the form

$$e_m = E_m \sin \omega_m t \tag{11.41}$$

The instantaneous phase of the wave is given by the expression

$$\theta = \theta_0 + K_p E_m \sin \omega_m t \tag{11.42}$$

where proportionality factor K_p determines the maximum variation in phase for a given signal strength.

By definition, the instantaneous phase for phase modulated wave is given by

$$\phi(t) = \omega_c t + \theta_0 + K_p E_m \sin \omega_m t \tag{11.43}$$

Therefore the expression for phase modulated wave can be written as

$$e = E_c \sin (\omega_c t + K_p E_m \sin \omega_m t) \tag{11.44}$$

θ_0 representing the constant phase, which plays no part in the modulation process, is taken zero. The above expression can be written as

$$e = E_c \sin (\omega_c t + \Delta\phi \sin \omega_m t) \tag{11.45}$$

where $$\Delta\phi = K_p E_m \tag{11.46}$$

and is called phase deviation. If the instantaneous phase of a phase modulated wave is given by equation (3), then the instantaneous frequency of the phase modulated wave will be

$$\omega = \frac{d\phi}{dt} = \frac{d}{dt} (\omega_c t + \theta_0 + \Delta\phi \sin \omega_m t)$$

$$= \omega_c + \Delta\phi \, \omega_m \cos \omega_m t$$

or $$f = f_c + \Delta\phi \, f_m \cos \omega_m t = f_c + \Delta f \cos \omega_m t \tag{11.47}$$

where $$\Delta f = \Delta\phi \, . \, f_m \tag{11.48}$$

and is called maximum frequency shift for a phase modulated wave. Obviously variation of the rate of change ϕ with time will produce a change in Δf and consequently in the instantaneous frequency 'f' of the phase modulated wave. This means phase modulation, like frequency modulation, is also a frequency variation process. From Eq. (11.48), we can also conclude that if we double either the amplitude E_m or the frequency f_m of the signal then maximum frequency shift Δf will be doubled. This is in contrast to frequency modulation process in which

$$\Delta f = K_f \frac{E_m}{2\pi} \qquad \text{(Eq. 11.34, Section 11.13.2)} \tag{11.49}$$

and thus depends only on signal amplitude and not on signal frequency.

We have shown above that the expression for phase modulated wave is

$$\begin{aligned} e &= E_c \sin (\omega_c t + \Delta\phi \sin \omega_m t) \\ &= E_c J_0 (\Delta\phi) \sin \omega_c t + E_c J_1 (\Delta\phi) [\sin (\omega_c + \omega_m) t \\ &\quad - \sin (\omega_c - \omega_m) t] + E_c J_2 (\Delta\phi) [\sin (\omega_c + 2\omega_m) t \\ &\quad - \sin (\omega_c - 2\omega_m) t] + \ldots \end{aligned}$$

For the particular case $\Delta\phi = 0.5$, the values of $J_n (\Delta\phi)$ can be determined from the curves of Fig. 19. These are

$$J_0 (0.5) = 0.938;\ J_1 (0.5) = 0.242;$$

$$J_2 = (0.5) = 0.030$$

and $$J_n (0.5) \cong 0 \text{ for } n > 2.$$

The equation of the phase modulated wave then becomes

$$\begin{aligned} e &= 0.938 E_c \sin \omega_c t + 0.242 E_c [\sin (\omega_c + \omega_m) t \\ &\quad - \sin (\omega_c - \omega_m) t] + 0.030 E_c [\sin (\omega_c + 2\omega_m) t \\ &\quad - \sin (\omega_c - 2\omega_m) t], \\ &\approx 0.938 E_c \sin \omega_c t + 0.242 E_c [\sin (\omega_c + \omega_m) t \\ &\quad - \sin (\omega_c - \omega_m) t], \end{aligned}$$

where the second side band components which have amplitude of $0.03 E_c$ have been neglected. Approximation is more valid when $\Delta\phi < 0.5$. Equation can be further put in the form

$$e = 0.938 E_c \left[\sin \omega_c t + \frac{\Delta\phi}{2} \sin (\omega_c + \omega_m) t - \frac{\Delta\phi}{2} \sin (\omega_c - \omega_m) t\right]$$

The corresponding equation for AM wave is

$$e = E_c \left[\sin \omega_c t + \frac{m_a}{2} \cos (\omega_c + \omega_m) t + \frac{m_a}{2} \cos (\omega_c - \omega_m) t\right]$$

It is obvious, on comparing the two equations, that both are identical if $\Delta\phi = m_a$ with only essential difference of the relative phase of the carrier and the sidebands. The difference can be eliminated if the sidebands of amplitude modulated wave are shifted in phase by 90° with respect to the carrier.

Example 11.3 *A 7.5 kc/s modulating signal has sufficient amplitude to cause a phase shift of 50°. Find the amount of indirect FM produced.*

$$\Delta f = \Delta\phi,$$

$$f_m = \frac{50\pi}{180} \times 7.5 = 6.55 \text{ kc/s}$$

Comparison with Frequency Modulated Wave

If we compare Eq. (11.36), Section 11.13.2 and Eq. (11.45), then it becomes obvious that $\Delta\phi$ determines the characteristics of phase modulated wave in the same manner as the factor m_f determines characteristics of frequency modulated wave. If $\Delta\phi$ be made numerically equal to the factor m_f, then the two resulting waves, i.e., one phase modulated wave due to phase modulation process and another frequency modulated wave due to frequency modulation process will be identical.

Since $\Delta\phi$ plays the role of m_f in the case of phase modulated wave, the expression for bandwidth can be written as

$$\frac{BW}{2\Delta f} = \frac{n}{\Delta\phi} \tag{11.50}$$

If the frequency of the signal f_m is doubled, it will not affect $\Delta\phi$ (see Eq. 11.46) and therefore, unlike to frequency modulation case, there will be no effect on n, the number of significant side frequency pairs. But the value of Δf is doubled so that

$$\frac{BW}{2(2\Delta f)} = \frac{n}{\Delta\phi}$$

or
$$\frac{BW}{2\Delta f} = 2 \cdot \frac{n}{\Delta\phi} \tag{11.51}$$

or in other words bandwidth covered by more widely spaced components is doubled. This is in contrast to the case of frequency modulation where bandwidth remains constant ever after doubling the signal frequency f_m.

11.15.2 Production of FM Wave by Phase Modulation

The undesirable variation of bandwidth with modulating frequency, as stated above, can be corrected if E_m is made to vary inversely as the modulating signal frequency f_m. For this purpose, an integrating circuit can be used. If the input to this circuit is $e_i = E_m \cos \omega_m t$, then output will be

$$e_0 = \frac{1}{RC} \int E_m \cos \omega_m t \, dt = \frac{1}{RC} \cdot \frac{E_m \sin \omega_m t}{\omega_m}$$

or
$$|e_0| = \frac{E_m}{RC\,\omega_m}$$

so that for constant E_m, the amplitude of output wave, $|e_0|$, varies inversely as ω_m. If the output $|e_0|$ is applied to the modulator, then frequency deviation Δf will not vary with f_m as shown below. For this, output voltage, E_m, should be replaced by E_m/CRf_m. That is,

$$\Delta\phi = K_P \frac{E_m}{RC\,\omega_m},$$

so that
$$\Delta f = \Delta\phi\, f_m = K_P \frac{E_m}{RC\, f_m} \times f_m$$
$$= K_P \frac{E_m}{RC} \tag{11.52}$$

This mean by doubling f_m, Δf as well as $\Delta\phi$ will remain unaffected. Consequently, bandwidth remains constant which happens in the case of frequency modulated wave. In other words, we can state that in this process in which, prior to applying the modulating voltage to modulator, an integrating circuit is employed to vary E_m inversely as f_m, a frequency modulated wave is produced by the process of the phase modulation.

11.16 PULSE MODULATION SYSTEMS

We have discussed already the modulation of carrier frequency by sinusoidal signals. A signal waveform can be produced by taking samples of this wave at regular intervals. The sampling of information signal is done by the use of pulse on varying one of their basic characteristics such as amplitude, width (duration), or position in accordance with this modulating signal (Fig. 11.21). These sampled pulses, bearing the characteristics of modulating signal, are then modulated upon the radio frequency carrier and thus short bursts of the modulated carrier are transmitted.

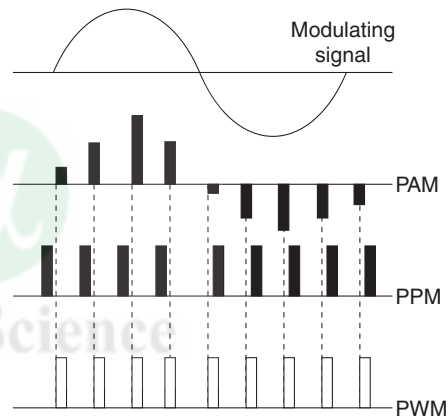


Fig. 11.21 *Various pulse modulation method*

We shall discuss pulse modulation methods in brief. Advantages of this system are:

(i) In such systems, amplitude limiting techniques may be employed to minimize noise effects.

(ii) The transmission can be made independent of fading and other difficulties.

(iii) Multiple signals may be transmitted on a single carrier.

(iv) Energy is radiated only in short bursts and therefore the consumption of power is low.

11.16.1 Pulse Amplitude Modulations (PAM)

In this process, relative amplitudes of successive pulses serve as a measure of signal amplitudes at the corresponding sampling instants. Width and position

of the pulse remain constant (fix). Modulating signal is sampled at the basic rate, usually $\frac{1}{2f_m}$ where f_m is the maximum value of frequency present in the modulating signal.

As shown in Fig. 11.22, these amplitude modulated pulses modulate the radio frequency carrier. The amplitude modulated radio frequency output is then transmitted in the form of short bursts. *The duration of each burst is equal to the duration of pulses in Fig. 11.22(a). Instantaneous amplitude of each radio frequency burst is proportional to the audio frequency signal shown in Fig. 11.22(c).*

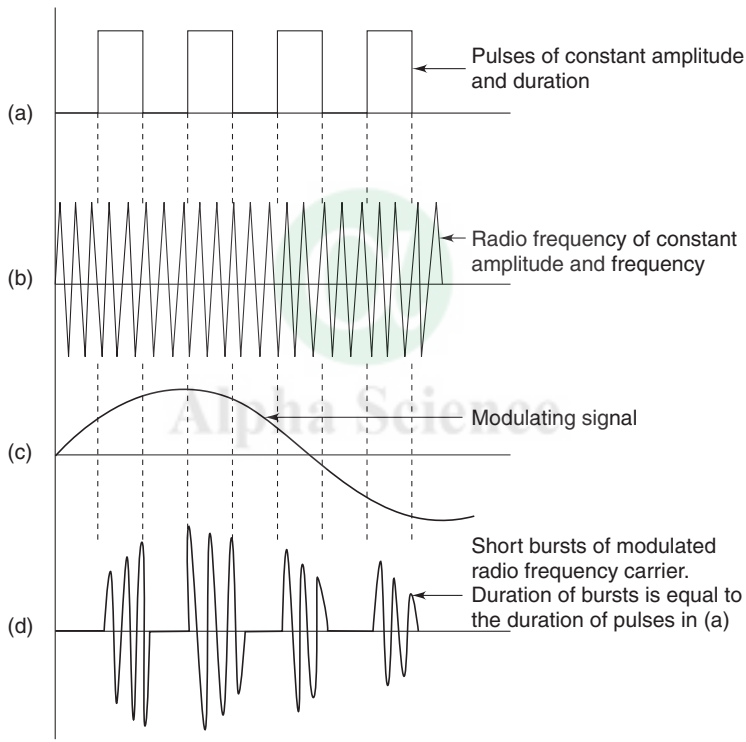


Fig. 11.22 *Pulse amplitude modulation*

The main advantages of pulse amplitude modulation are:

(a) greater peak power, compared to previous amplitude modulation, is obtained, and

(b) signal to noise ratio of the system is improved.

11.16.2 Pulse Duration or Width Modulation (PWM)

The width of the pulse is varied (around fixed value) in accordance with the modulating signal whereas amplitude and position remain fixed.

An emitter coupled multivibrator generating pulse-width modulation is shown in Fig. 11.23. it may be noted that

(i) the *starting time* of pulses from a monostable multivibrator is controlled by applying trigger pulses, and

(ii) the signal, to be sampled, controls the *duration* of these pulses from the monostable multivibrator, and

(iii) the gatewidth of this multivibrator depends on the voltage to which the capacitor, C, is charged and therefore if this voltage is varied in accordance with signal voltage, rectangular pulses of varying width will be generated at output.

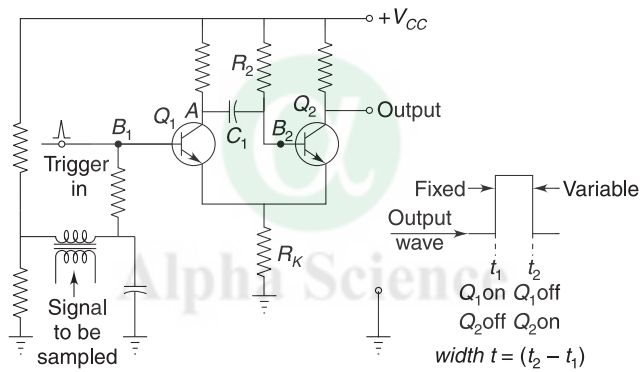


Fig. 11.23 *Emitter coupled monostable multivibrator*

Operation

For this type of modulator, stable state is with Q_1 OFF and Q_2 ON. When trigger pulse (a positive going pulse of large amplitude) is applied at the base of Q_1, then it starts conducting. As it draws the collector current, voltage at A falls. Since any charge on condenser C_1 cannot change instantaneously, the falling voltage at A is directly applied to the base of Q_2. This reduces the plate current of Q_2, making the emitters less positive with respect to ground. Its result is that Q_1 conducts more, making now B_2 more negative. Action continues so that Q_2 is finally cut-off while Q_1 remain in conduction. This is called regenerative action.

With Q_2 OFF, capacitor C_1 begins to charge towards $+V_{CC}$ through R_2. *The rate of charging depends upon the supply voltage and time constant R_2 C_1*. Thus when voltage across C_1 builds up to a level so as to make B_2 sufficiently positive, Q_2 starts conducting and Q_1 is cut-off by regenerative action. The time when Q_2 starts conducting depends upon:

(i) the time constant, $R_2\ C_1$.

(ii) the voltage across common-emitter resistor, R_k because Q_2 conducts only when the voltage at base B_2 becomes slightly more positive than voltage across R_k.

The voltage across R_k depends upon the current flowing through, which at the time is collector current of Q_1 (which is then ON). The collector current of Q_1, in turn, depends on the base bias voltage, at B_1. This base bias voltage, in turn, is governed by the instantaneous changes in applied modulating signal. Thus finally we arrive at the *instant* t_2, at which Q_2 conducts, and, *is determined by the signal to be sampled*. As instant t_1 is fixed, width $(t_2 - t_1)$ of the output pulse varies in accordance with the modulating signal.

11.16.3 Pulse Position Modulation (PPM)

In this process, the spacing along the time axis between the pulse do not remain uniform, but is varies in accordance with the modulating signal (position of the pulses is shifted) whereas the amplitude and width remain fixed.

Pulse position modulation can be achieved by differentiating width modulated pulses and then rectifying the differentiated output. After rectifying only those pulses which have differentiated trailing edge of width modulated pulses will be obtained as shown in Fig. 11.24.

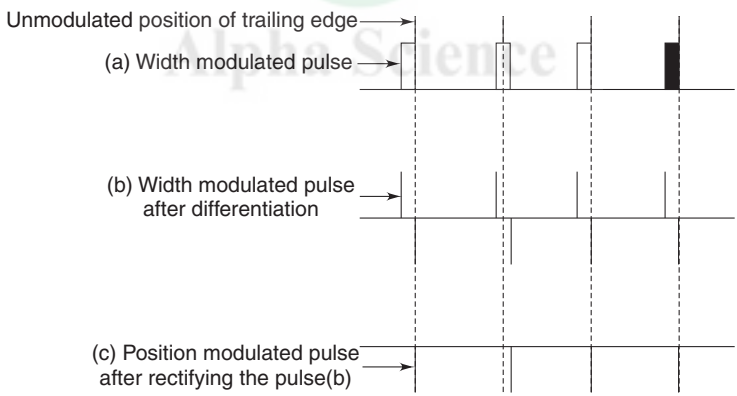


Fig. 11.24 *PPM from PWM*

11.16.4 Pulse Code Modulation (PCM)

In pulse code modulation, signal is first sampled, then sampled amplitude is selected in terms of nearest standard amplitude, called *quantization*. This selected standard amplitude is then indicated by a code of pulses. This finally signal is transmitted as code of pulses. We note that there is *no* modulation, only coding is involved. Thus an analog signal is converted into binary digital form.

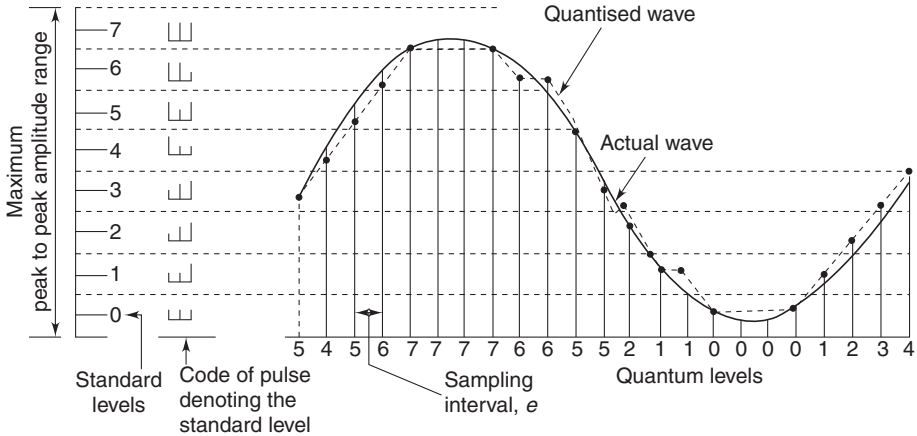


Fig. 11.25 *PCM sampling and quantizing*

As shown in Fig. 11.25, the signal is sampled at an interval *t*. This interval is kept small so that a sufficient number of samples can be taken. Though the signal is sampled at regular intervals, but due to sufficient number of samples taken at frequency intervals, nothing useful of the signal to be transmitted is lost.

The actual amplitude of the sampled signal is then not chosen as such for transmission, but it (sampled amplitude) is quantized into a nearest standard amplitude. For example, actual amplitude *b* is chosen as a nearest standard amplitude 4; actual amplitude *a* as a nearest standard amplitude 5. This is done to avoid the use of a prohibitively large number of code pulses to indicate the sampled amplitudes of the signal. In this process of quantizing the information, some distortion (as is obvious from dotted and solid curves) is introduced, but this error is minimized the number of standard amplitudes.

These selected standard amplitude are not transmitted as such but are indicated by a code of pulses. As shown in the figure, each of the standard amplitudes from 0 to 7, can be specified by a different combination of pulses in a three pulse code in which a pulse may be either present ot absent, or in other words code mainly concerns with the presence or absence of the pulse and such a code corresponds to binary digital code in which the presence of a pulse is indicated by 1 and absence by 0*. Thus the code for the eight standard amplitudes in three pulse code is given in the table.

* Remember *zero* state of the binary is that state in which output tube conducts while input tube is cut-off. This is initial state and no trigger pulse is applied to the first tube. When any pulse of appropriate polarity and amplitude is applied to the tube second, a reversal of state occurs and output tube is cut off while input tube conducts and this is the *one* state of the binary.

	2^2	2^1	2^0
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

Obviously, seven digit code is capable of transmitting $2^7 = 128$ standard amplitudes.

Thus, in the PCM system, groups of pulses or codes are transmitted which represent binary numbers corresponding to modulating voltage (called analog signal) levels.

Let us now explain, in brief, sampling, quantizing and coding with the help of diagrams. In Fig. 11.26, sampling of a continuously varying signal, $p(t)$, called

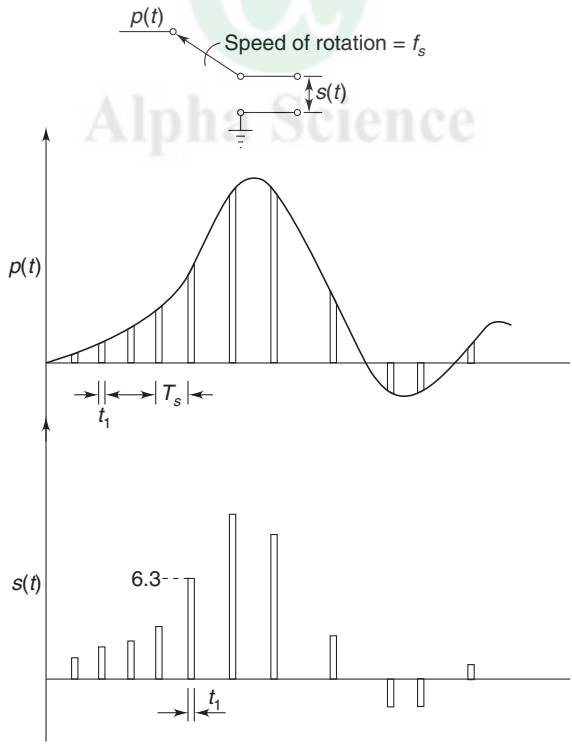


Fig. 11.26 *Sampling of the signal*

analog signal, is shown conceptually, $p(t)$ is sampled periodically at a rate of f_c samples per second. Suppose switch remains on $p(t)$ line for t_1 seconds, while rotating at a desired rate of $f_S = 1/T_s$ times per second ($t_1 << T_s$). The switch output $S(t)$ is then sampled version of $p(t)$. Both $p(t)$ and $s(t)$ are shown in figure. f_s is called *sampling rate* and T_s *is the sampling interval.*

Suppose sampled amplitude of one of the pulse is 6.3 then how it is quantized and binary coded is shown in Fig. 11.27.

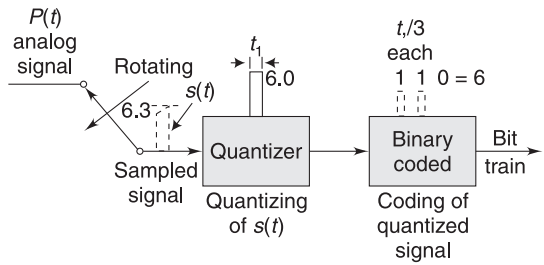


Fig. 11.27 *PCM transmission system section*

Quantisation Error The difference between the analog signal levels and the quantized levels is termed as quantization noise. This noise can be reduced by using a large number of quantizing levels at the expense of increased cost and bandwidth.

Bandwidth of PCM System The bandwidth of PCM system depends upon the highest frequency in the analog signal and the number of bits in the PCM signal. Suppose each quantized level is converted into n bits and let f_s represent the sampling frequency, then signaling rate is

$$r = nf_s \text{ bits/sec.}$$

If the highest frequency in the analog signal spectrum be W then $f_s = 2W$ and the signaling rate,

$$r = n\ 2W \text{ bits/sec.}$$

For a speech signal of 3.2 kHz, the usual sampling rate 8 kHz and if the system has 128 quantising levels, the number of bits required is $2^n = 128$ or $n = 7$. The final PCM signal will therefore have a bit rate of $7 \times 8 = 56$ k.bits/sec. It means system will require a *large bandwidth.*

11.17 COMPARISON BETWEEN FM AND PM

FM and PM are quite similar except by the manner in which the modulation index is defined.

1. For FM, the phase angle is proportional to modulating wave whereas for PM, the time derivative of phase angle is proportionally to modulating wave.
2. In FM, the frequency deviation is proportional to the amplitude of the modulating signal whereas in PM, the phase deviation is proportional to the amplitude of the modulating signal and, therefore, independent of its frequency.
3. In case of FM, the modulation index is inversely proportional to the modulating frequency whereas in PM, the modulation index is proportional to the modulating voltage.

It means that under identical conditions FM and PM are indistinguishable for a single modulating frequency. When the modulating frequency is changed, the PM modulation index will remain constant but the FM modulation index will decrease as modulating frequency is increased and vice versa.

11.18 PM SIGNAL GENERATION

Phase modulation (PM) and frequency modulation (FM) are closely related to each other because in both the cases there is a variation in the total phase angle. In PM, the phase angle varies linearity with baseband or modulating signal whereas in case of FM, the phase angle varies linearity with the integral of baseband or modulating signal. This means that FM wave may be obtained by using PM. Conversely, PM may be obtained using FM.

To get FM by using PM, we first integrate the modulating signal and then apply to the phase modulator. This process is illustrated with the help of a block diagram shown in Fig. 11.28.

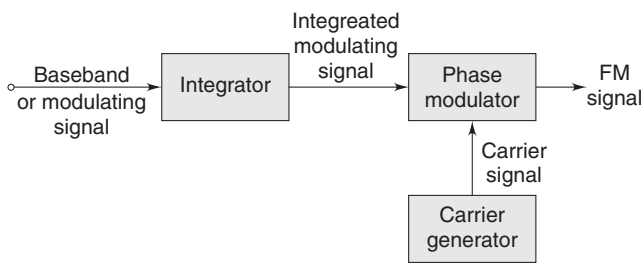


Fig. 11.28 *Generation of FM using phase modulator*

Similarly, PM wave may be generated by using frequency modulator by first differentiating modulating or baseband signal and then applying to the frequency modulator. The process is illustrated with the help of block diagram shown in Fig. 11.29.

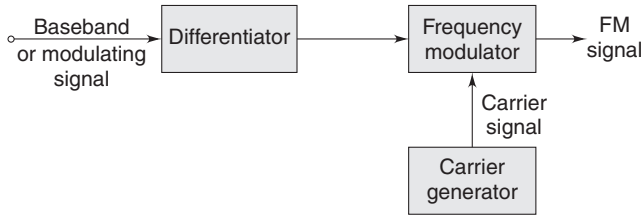


Fig. 11.29 *Generation of PM using frequency modulator*

11.19 DEMODULATION OR DETECTION

Demodulation or detection is a process of recovering the original modulating signal (intelligence) from the modulated carrier wave i.e., the demodulation is a process reverse of the process of modulation.

11.19.1 Necessity of Demodulation

The wireless signals transmitted from a transmitter consist of RF carrier waves and audio frequency signal waves. If the modulated wave is directly fed to the loudspeaker, no sound will be heard from the loudspeaker. This is because of the simple reason that the frequency of the carrier wave is very high and the loudspeaker diaphragm cannot respond to such high frequencies due to large inertia of their vibrating disc etc. Neither will such RF waves produce any effect on human ear as their frequencies are much beyond the audible frequencies (20 Hz to 20 kHz approximately). Hence it becomes essential to separate the audio frequency signal (intelligence) from the modulated carrier wave.

11.20 AM DETECTION

The process of modulation demands that the modulated wave has some definite average value and the carrier wave is separated out. Hence in demodulation of an AM wave two operations, *viz*., rectification of the modulated wave and elimination of RF component of the modulated wave are involved.

11.20.1 Essentials of AM Detection

The modulated wave has both positive and negative peaks of the same amplitude and, therefore, the average value of the wave is zero. Hence it becomes essential to reduce (or better, eliminate) one half of the modulated wave so that the average of the resultant wave is not zero. Either half of the modulated wave can be eliminated by the process of rectification. After rectification of the modulated wave, the RF portion of the remaining half of the wave is removed by means of a suitable filter so that a signal, that follows the envelope of the rectified signal i.e., the envelope of one half of the wave, is produced. The process is shown in Fig. 11.30.

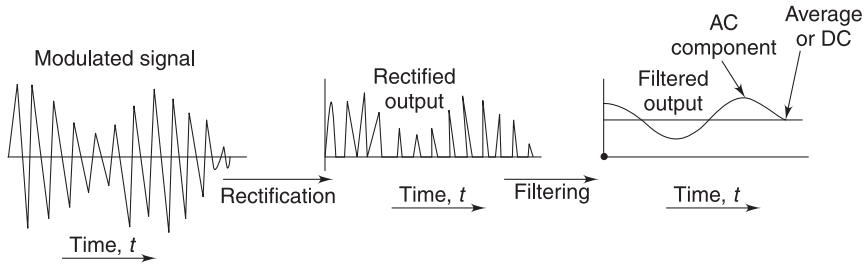


Fig. 11.30 *Amplitude detection process*

11.21 AM DEMODULATORS (OR DETECTORS)

The devices used for demodulation or detection are called **demodulators** or **detectors**. For amplitude modulation (AM), detectors or demodulators used are categorized as (i) square law detectors are (ii) envelope detectors.

AM signal with large carrier are detected by using the envelope detector. The envelope detector uses the circuit which extracts the envelope of the AM wave. In fact, the envelope of the AM signal can only be detected by means of a square law detector in which a device operating in the nonlinear region is used to detect the modulating signal.

11.21.1 Square Law Detector

The square law detector circuit is used for demodulating modulated signal of small amplitude (i.e. below 1 V) so that the operating region may be restricted to the nonlinear portion of the V-i characteristics of the device. Figure 11.31 shows the circuit of a square law detector. It may be observed that the circuit is very similar to the square law modulator. The only difference lies in the filter circuit. The square law modulator uses a bandpass filter whereas the square law detector makes use of a low-pass filter.

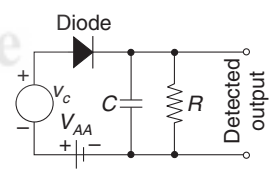


Fig. 11.31 *Basic circuit of a square law detector*

In the circuit, the dc supply voltage V_{AA} is used to get the fixed operating point in the non-linear region of the diode V-i characteristic. Since, the operation is limited to the nonlinear region of the diode characteristics, the lower half portion of the modulated waveform is compressed. This produces envelope applied distortion. Because of this, the average value of the diode current is no longer constant, rather it varies with time as indicated in Fig. 11.32.

The distorted output diode current is expressed by the non-linear V-i relationship (i.e. square law) as

$$i = av + bv^2 \tag{11.53}$$

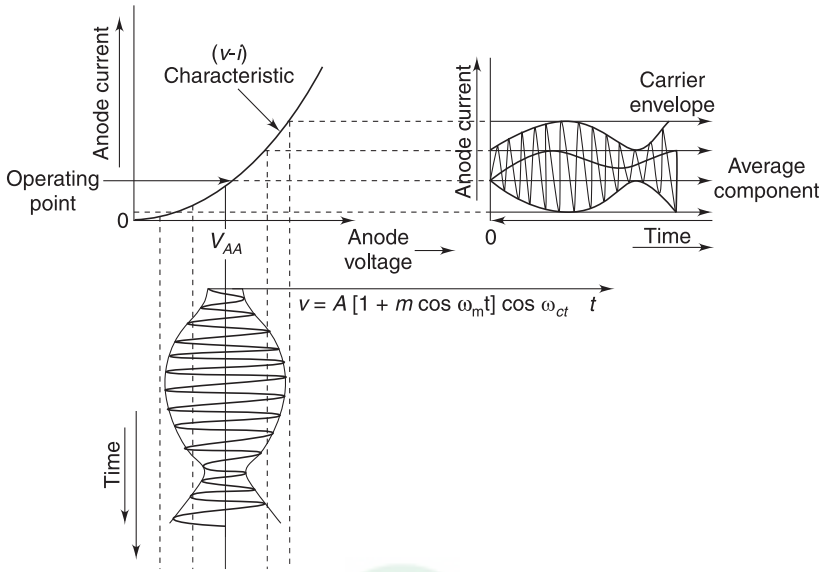


Fig. 11.32

where v is the input modulated voltage

AM wave may be expressed as

$$v = A\ (1 + m \cos \omega_m t) \cos \omega_c t \tag{11.54}$$

Substituting the value of v from Eq. (11.54) in Eq. (11.53), we have

$$i = a\ [A\ (1 + m \cos \omega_m t) \cos \omega_c t] + b\ [A\ (1 + m \cos \omega_m t) \cos \omega_c t]^2 \tag{11.55}$$

Now, if above equation is expanded, then we observe the presence of terms of frequencies like $2\omega_c$, $2(\omega_c \pm \omega_m)$, ω_m and $2\omega_m$ besides the input frequency terms.

Hence, this diode current i containing all these frequency terms is passed through a low-pass filter which allows to pass the frequencies below or up to modulating frequency ω_m and rejects the other higher frequency components. Thus, the modulating or baseband signal with frequency ω_m is recovered from the input modulated signal.

11.21.2 Linear Diode or Envelope Detector

Diode detection is called the *envelope detection* as it recovers the audio frequency signal envelope from the composite signal. Diode detector is also called the linear detector as its output is proportional to the voltage of the input signal.

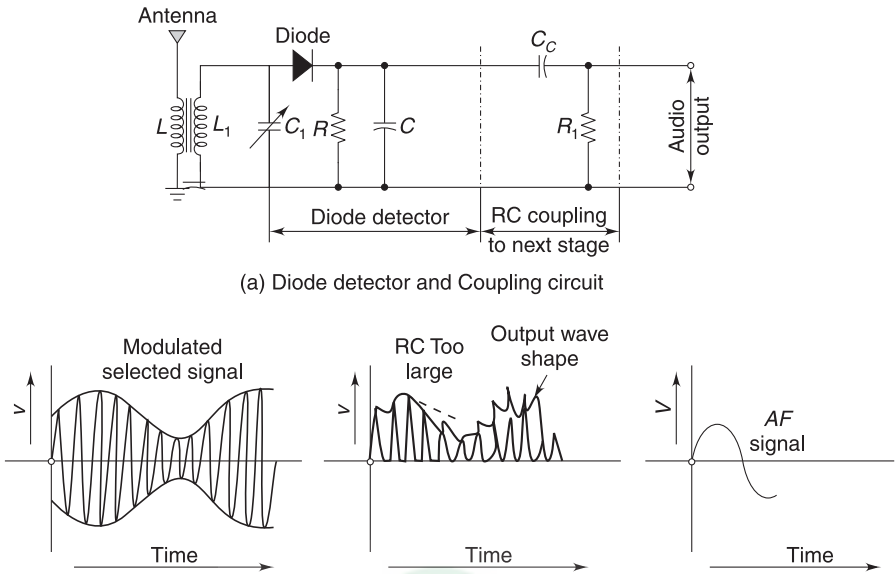


Fig. 11.33 *Linear diode detector*

Linear diode detector makes use of rectification property of a diode. Such detectors are widely used in commercial radio receivers. However, a linear diode detector needs a carrier voltage of 5V or more so as to provide satisfactory operation. With such a high value carrier voltage the cutin voltage of the diode may be neglected and the operation may be considered to be taking place over the linear portion of the dynamic voltage-current characteristic of the diode. The circuit of a basic diode detector is shown in Fig. 11.33. Here L_1-C_1 is a parallel resonant circuit. By varying C_1, the resonant frequency of this circuit can be varied and hence RF signal of any frequency can be tuned in. It is to be noted that the antenna receives various modulated signals of different frequencies from various stations and the desired frequency is to be tuned in. The modulated selected signal is now applied to the junction diode. The diode conducts when the carrier signal is positive and cuts off when it is negative. Consequently, each positive half cycle of the carrier wave appears across the parallel combination of R and C in the manner depicted in Fig. 11.33(b). The capacitor C charges to nearly the peak value of the positive half cycle of the carrier wave. A small voltage drop occurs across the diode. During the negative half cycle of the carrier the capacitor C discharges through the resistor R at a time constant RC. In choosing R and C care must be taken not to select too small a value for RC, otherwise the charge will be dissipated and voltage will be reduced to zero during the negative half

cycle. This condition can be avoided by choosing *RC* large compared to time it takes for the carrier to pass through one cycle. On the other hand, it is important that *RC* not be chosen excessively large, for then the capacitor will discharge so slowly that in the decreasing portion of the rectified output diagonal clipping occurs as indicated in Fig. 11.33(b). The best value of *RC* is that which causes the rate of discharge of the capacitor to follow the variations in the modulating signal. Analysis shows that the maximum permissible value of time constant *RC* for avoiding negative peak clipping is given by

$$\frac{1}{RC} \geq \omega_m \frac{m}{\sqrt{1 + m^2}} \tag{11.56}$$

The analysis, however, neglects many factors which affect the results appreciably. So the maximum permissible value based on empirical formula is given as

$$\frac{1}{RC} \geq \omega_m \, m \tag{11.57}$$

The output signal of the diode detector has somewhat jagged variation depicted in Fig. 11.33(b). It has a waveshape that very closely resembles to that of the modulating signal, with the exception that it contains a d.c. level. This, however, is readily removed by using capacitance-resistance coupling to the next stage, as shown in Fig. 11.33(a). The output of this coupling network is a signal that very closely matches the original modulating signal.

Mathematical Analysis From Section 11.3.1 Eq. (11.4)

$$v = V_c \, (1 + m \sin \omega_m t) \sin \omega_c t$$

The current during conduction period is

$$i = \frac{v}{R + r_d} = \frac{V_c \, (1 + m \sin \omega_m t) \sin \omega_c t}{R + r_d}$$

where r_d is the diode resistance for $0 \leq \omega_c t \leq \pi$

So
$$I_{av} = \frac{1}{2\pi} \int_0^{2\pi} id \, (\omega_c t)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \frac{V_c}{R + r_d} \, (1 + m \sin \omega_m t) \sin \omega_c t \, d \, (\omega_c t)$$

$\sin \omega_m t$ term is assumed to be constant because $\omega_m << \omega_c$.

$\therefore$
$$I_{av} = \frac{1}{2\pi} \frac{V_c}{R + r_d} \, (1 + m \sin \omega_m t) \, [\cos \omega_c t]_0^{\pi}$$

$$= \frac{V_c \, (1 + m \sin \omega_m t)}{\pi \, (R + r_d)} \quad \text{neglecting – ve sign}$$

$$V_{av} = I_{av} R$$

$$= \frac{V_c (1 + m \sin \omega_m t) R}{\pi (R + r_d)}$$

$$= \frac{V_c R}{\pi (R + r_d)} + \frac{R V_m \sin \omega_m t}{\pi (R + r_d)} \qquad (11.58)$$

$$\left(\because m = \frac{V_m}{V_c} \right)$$

In above Eq. (11.58) the first term represents a d.c. component and the second term represents the a.c. component, the value of which is proportional to the modulating voltage. By processing the second term, information can be extracted.

Detection efficiency,

$$\beta = \frac{\text{Average pd across the load}}{\text{Peak value of input signal voltage}}$$

$$= \frac{\dfrac{V_c (1 + m \sin \omega_m t) R}{\pi (R + r_d)}}{V_c (1 + m \sin \omega_m t)}$$

$$= \frac{R}{\pi (R + r_d)} = \frac{1}{\pi (1 + r_d/R)} = 31.8\% \quad \because r_d << R$$

Advantages of Diode Detectors

1. They can be operated as linear or power detectors.
2. They can handle comparatively large input signals.
3. Good linearity is achieved, as negligible distortion is introduced during rectification.
4. They can be used in simple automatic gain control circuits.

However, they suffer from the following drawbacks.

1. During conducting period the diode consumes some power, so Q of its tuned circuit along with its gain and selectivity is reduced.
2. The diode detectors are not capable of amplifying the rectified signal by themselves as is possible in case of transistor detector. However, signal can be amplified both before and after rectification so this drawback does not matter much.

Questions

11.1 What is modulation and why is it necessary?

11.2 Explain the need of modulation in a communication system. Write the types of modulation used for different communication systems and explain them in brief.

11.3 The equation of a carrier signal is given by $e_c = E_c \sin(\omega_c t + \phi)$ and that of modulating signal $e_m = E_m \cos \omega_m t$. Derive an equation for the modulated signal and modulation index for amplitude modulation.

11.4 What do you mean by modulation index? Derive an expression for that AM index. Also sketch waveforms for different values of index.

11.5 Define "amplitude modulation" and "modulation index". Write down the equation for (i) amplitude and (ii) the instantaneous voltage of the amplitude modulated wave. Sketch the graph of an amplitude modulated wave.

11.6 In amplitude modulation, derive a relation between total power of a modulated signal, carrier signal power and modulation index.

11.7

(i) Write an expression for AM wave $v(t)$

(ii) Draw the waveform for $v(t)$.

(iii) Show that modulation index,

$$m = \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}}.$$

11.8 Explain the square law modulator for the generation of AM signals.

11.9 Explain with a neat diagram, the working of a balanced modulator.

11.10 Define AM and FM. Use sketches to explain these definitions. Where are these employed?

11.11 Explain frequency modulation.

11.12 What is phase modulation? Obtain an expression for a phase modulated wave when the modulating signal is sinusoidal.

11.13 Explain the working of square law demodulator for detection of AM wave.

11.14 Explain the working of a linear diode detector.

11.15 Write short notes on:

(i) Need of modulation

(ii) Diode detector

(iii) Advantages of FM over AM

(iv) Types of modulation
(v) Types of amplitude modulation
(vi) Vestigial side band modulation system
(vii) Collector modulation
(viii) Square law diode modulator
(ix) Phase shift method for SSB generation
(x) Power relations in FM wave.
(xi) Phase modulation
(xii) Pulse amplitude modulation
(xiii) Pulse width modulation
(xiv) Pulse position modulation
(xv) Pulse code modulation

12 Noise

12.1 INTRODUCTION

Communication is the process whereby information is transferred from one point called the source to the other point called destination. *Signal* in communication is the electrical analogue of the message at the source. (The message at the origin is not electrical in nature. It is converted into electrical form using a transducer.

Signals are basically single valued functions of time which may be real or complex valued. Signals can be classified as *deterministic signals* and *random signals*. By *deterministic signals* we mean the class of signals that can be expressed explicitly as time functions (This means that the behaviour of the signal is known before hand). We say a signal is *random* if its value is not predictable in advance. In this sense, and from the receiving end viewpoint, all meaningful signals in communication are random, for if the signal were known before hand, there would be no point in transmitting it. Thus a known signal does not contain any information. In the present chapter, emphasis will be on unwanted *random signals* or *noise*. For example, we may consider a radio communication system. The received signal consists of an information – bearing signal component, a random interference component and receiver noise. The information bearing signal may be voice signal consisting of randomly spaced bursts of energy signals. The interference components are extraneous electromagnetic waves produced by other communication systems and atmosphere. A major source of noise arises from random motion of electrons in the components and devices at the front end of the receiver.

12.2 NOISE: CLASSIFICATION AND ORIGIN

Noise may be defined, in an electrical sense, as an unwanted form of electrical signal which tends to interfere with the proper and easy reception and reproduction of desired signals. Many disturbances of an electrical nature produce noise in receivers, modifying the signal in an unwanted manner. In radio receiver noise may produce hiss in loud speaker output, whereas in television receivers it may appear as *snow* or *confetti* (coloured snow) superimposed on the picture. In pulse

communication system, noise may produce unwanted pulses which may cancel the wanted ones and lead to serious error in the detected signal. Noise is thus seen as limiting the range of systems, for a given transmitted power. Noise also affects the sensitivity of receivers, by placing a limit on the weakest signal that can be amplified and reproduced properly.

There are numerous ways of classifying noise. It may be sub-divided according to type, source, effect or relation to receiver. The most convenient way is to divide noise in two broad groups e.g., noise whose sources are *external to the receiver* and noise created *within the receiver*. From this viewpoint noise can be considered to be of two basic types, *external noise* originating from sources outside the receiver and *internal noise* arising from the system itself.

12.2.1 External Noise

Various forms of noise created outside the receiver come under this category which can be further classified as *atmospheric noise* and *extra-terrestrial noise*.

(a) Atmospheric Noise

This noise arises from lightening discharges in thunderstorms and other electrical disturbances occurring in the atmosphere. This disturbances occur in the form of impulses normally used for broadcasting. Atmospheric noise thus consists of spurious radio signals with components distributed over a wide range of frequencies. These disturbances propagate in the same manner as ordinary radio waves of the same frequency. Thus, a receiver located at any point on the ground receives and reproduces these signals along with the desired signals. These atmospheric noise resulting from natural disturbances is referred to as *static*. Field strength of this noise signal is approximately, inversely proportional to frequency. As a result this noise intereferes more with radio receivers than television. Atmospheric noise becomes less severe at frequencies above 30 MHz because higher frequencies are limited to line of sight propagation (less than 80 km) and the effect of the noise components is very little in the VHF range and above.

(b) Extraterrestrial Noise

This noise arises from sun and other distant stars and can be classified as *solar noise* and *cosmic noise*.

(i) Solar Noise

It arises from *sun*. Under normal conditions (quiet condition), there is a constant noise radiation from the sun simply because it is a large body heated at a very high temperature (6000°C on the surface). It therefore, radiates over a very broad frequency spectrum which includes the frequencies we use for communications. However, the disturbances in the sun is variable and undergoes cycles at the

peak of which electrical disturbances erupt. These additional disturbances are several orders of magnitude greater than the noise generated during periods of quiet sun. The solar cycle repeats these period of great electrical disturbances approximately every 11 years. Further, these 11 year cycle peaks reach even a higher maximum peak every 100 years or so. These 100 year peaks appear to be increasing in intensity. Thus, the noise generated by sun changes periodically with the solar disturbances.

(ii) Cosmic Noise

This refers to noise coming from distant stars other than sun. These stars are also heated bodies with radiated signals spreading a wide range of frequency. The noise received from such stars is also called black-body noise and is distributed fairly uniformly over the entire sky. *Space noise* is observable in the range from about 8 MHz to about 1.43 GHz. This is the strongest component of noise in the range 20 – 120 MHz.

(iii) Industrial Noise or Man-made Noise

This noise is strongest in industrial areas. The frequency of man-made noise spans between 1 to 600 MHz. *Man-made noise* is found in urban, suburban and industrial areas. The intensity of the noise made by human easily outstrips that created by any other source, internal or external to the receiver. Under this heading, noise generated by sources such as automobile, and aircraft ignition, electric motors and switching gear, leakage from high voltage lines and other heavy electric machines are all included. Fluorescent lights are another powerful source of such noise and therefore, should not be used where sensitive receiver is installed. The noise is usually, produced by the arc discharge present in the above mentioned operations. Therefore, the noise produced is most intense in industrial and densely populated areas. The nature of this noise is so variable that it is difficult to analyse and model this noise on any basis other than statistical. The noise however, does obey the general principle that the received noise increases as the receiver bandwidth is increased.

12.2.2 Internal Noise

This noise is created by various components used in processing the received signal and is completely internal to the system. This can be reduced by proper design. The effect of this noise is most significant at the front end part of the receiver where the strength of the signal is usually weak. One unavoidable cause of this noise is the thermal motion of electrons in conducting media like wires, resistors etc. As long as communication systems are constructed with such network components as resistors, inductors, capacitors etc., this *thermal noise* will be present. The other source of internal noise arises from the use of electronic components in the receiver system. These electronic devices work with random

emission from cathode (in case of vacuum devices) or injection of electrons or holes (in the case of solid-state devices). This random emission or injection is responsible for the other type of internal noise called *shot noise*. Internal noise can thus be classified under two heads *shot noise* and *thermal noise*.

(a) Shot Noise

Shot noise arises in electronic devices because of discrete nature of current flow in the device. In vacuum devices this noise arises from random emission of electrons from the cathode while in solid-state devices it arises from the random diffusion of minority carriers and the random generation and recombination of electron hole pairs. When this noise is amplified and fed to loudspeaker it sounds like a shower of lead shots striking a metal plate and hence the name *shot noise*. The following analysis makes a quantitative estimation of the power spectral density of shot noise in a vacuum diode.

Consider a temperature limited diode as shown in Fig. 12.1. We assume that the cathode-plate potential difference is large enough to cause the electrons emitted thermionically by heated cathode to be pulled to the plate with such high velocities that space-charge effects are negligible. The plate current is then determined effectively by the rate at which electrons are emitted from the cathode. By considering the plate current as the sum of a succession of current pulses, with each pulse caused by the transit of one electron through the cathode plate space, we can find the mean square value of the randomly fluctuating component of the current as follows.

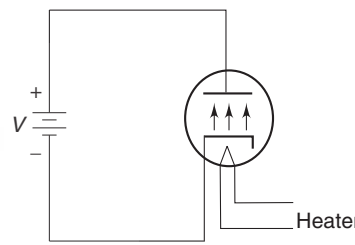


Fig. 12.1 *Vacuum diode*

Assume that a *single* electron is emitted from the cathode. The electron acquires velocity as it moves towards the plate and induces current i_n (t). The emitted electron experiences a force, F given by

$$F = qV/d \tag{12.1}$$

where, q is the electronic charge and V is the applied voltage and d is the separation between the electrodes of the diode. The acceleration of the electron is, therefore,

$$a = \frac{qV}{md} \tag{12.2}$$

where m is the mass of the electron.

Assuming the initial velocity to be zero, the velocity $v(t)$ of an electron at any time t is given by

$$v(t) = \frac{qV}{md}\, t \tag{12.3}$$

The kinetic energy of the electron at any time t is

$$E(t) = \frac{q^2V^2}{2md^2}\, t^2 \tag{12.4}$$

If the motion of this electron induces a charge Q on the plate, then the amount of work W that must be done to induce Q on the plate with potential V is

$$W = QV \tag{12.5}$$

Equating this work to the kinetic energy of the electron given by Eq. (12.4), we get

$$Q = \frac{q^2Vt^2}{2md^2} \tag{12.6}$$

The induced plate current is given by

$$i_c\,(t) = \frac{dQ}{dt} = \frac{q^2V}{md^2}\, t \tag{12.7}$$

That is,
$$i_c\,(t) = \frac{q}{d}\, v(t) \tag{12.8}$$

Thus the induced current is proportional to the velocity of the electron.

The transit time of the electron τ_a (time taken by the electron to reach the plate from cathode) can be calculated from Eq. (12.3) by integrating both sides with respect to t. Thus,

$$d = \int_0^{\tau_a} v(t)\, dt = \frac{qV}{md} \int_0^{\tau_a} t\, dt$$

That is,

$$d = \frac{qV}{2md}\, \tau_a^2 \tag{12.9}$$

and
$$\tau_a = \sqrt{\frac{2m}{qV}}\, d \tag{12.10}$$

Substituting the value of d from Eq. (12.9) into Eq. (12.7), we get

$$\begin{aligned} i_c\,(t) &= \frac{2q}{\tau_a^2}\, t, \quad \text{for } 0 < t < \tau_a \\ &= 0, \qquad \text{for } t > \tau_a \end{aligned} \tag{12.11}$$

Thus the electron induced current increases linearity from zero to $2q/\tau_a$ and goes to zero as soon as the electron reaches the plate at $t = \tau_a$. Each emitted electron from the cathode induces such triangular current pulses. The total plate current is thus, composed of a large number of these triangular current pulses

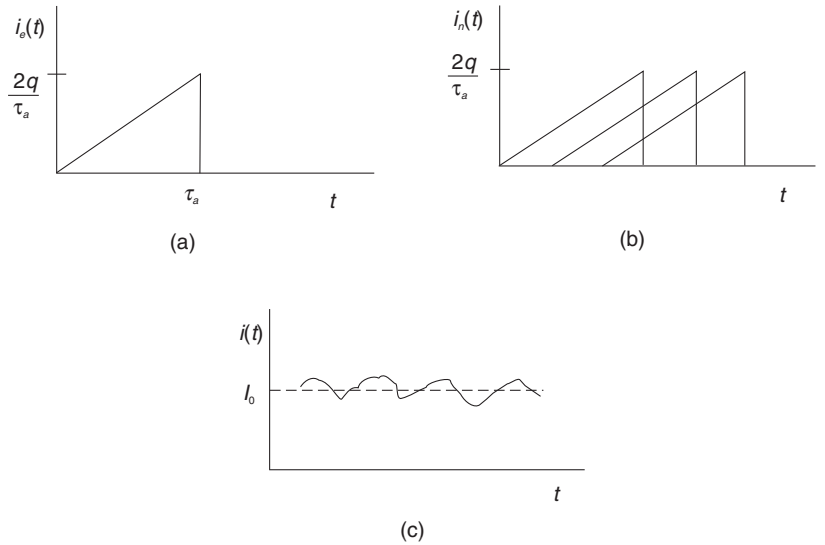


Fig. 12.2 *(a) Current pulse induced by a single electron. (b) Plate current comprising randomly distributed current pulses. (c) Shot noise current components*

distributed randomly as shown in Fig. 12.2. The diode current is the sum of all such current pulses. The area under each current pulse is

$$I_{os} = \frac{1}{2} \times \frac{2q}{\tau_a} \times \tau_a = q \text{ units} \tag{12.12}$$

Hence, the average value of the plate current is

$$\begin{aligned} I_o &= \overline{n}\, I_{os} \\ &= \overline{n} q \end{aligned} \tag{12.13}$$

where $\overline{n}$ is the average number of electrons emitted per second.

We shall now calculate the power density spectrum of the shot noise current. The shot noise current $i(t)$ consists of two components e.g., a constant current I_0 and a time varying component $i_n(t)$ (see Fig. 12.2(c)). Since $i_n(t)$ is random, it cannot be specified as a function of time. However, $i_n(t)$ represents a stationary random signal can be specified in terms of power density spectrum. As there one $\overline{n}$ pulses per second, it is reasonable to expect that the power density spectrum of $i_n(t)$ will be $\overline{n}$ times the energy density spectrum of $i_n(t)$. Thus, the power density spectrum $S_i(f)$ of $i_n(t)$ will be

$$S_i(f) = \overline{n}\, |I_e(f)|^2 \tag{12.14}$$

where $I_e(f)$ is the Fourier transform of $i_e(t)$

In order to find $I_e(f)$, we write $i_e(t)$ as

$$i_e(t) = \frac{2q}{\tau_a^2} [t\ u(t) - \tau_a\ u(t - \tau_a) - (t - \tau_a)\ u(t - \tau_a)] \tag{12.15}$$

Taking Fourier transform on both sides we get

$$I_e(f) = \frac{2q}{-(2\pi f)^2\ \tau_a^2} [1 - \exp(-j2\pi f\tau_a) - j2\pi f\tau_a \exp(-j2\pi f\tau_a))] \tag{12.16}$$

Substituting the value of $I_e(f)$ from Eq. (12.16) into Eq. (12.14) we get

$$S_i(f) = \frac{4I_0 q}{(2\pi f\tau_a)^4} [(2\pi f\tau_a)^2 + 2(1 - \cos(2\pi f\tau_a) - 2\pi f\tau_a \sin(2\pi f\tau_a)] \tag{12.17}$$

The average power density spectrum $S_i(f)$ can be plotted as a function of f. It is however, convenient to plot $S_i(f)$ in terms of $2\pi f\tau_a$ as shown in Fig. 12.3. The power density spectrum is nearly flat for $2\pi f\tau_a < 0.5$.

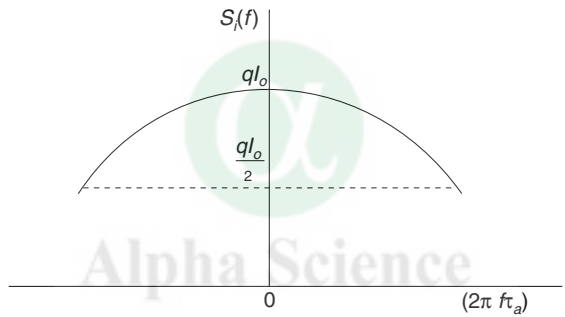


Fig. 12.3 *Power spectral density of shot noise versus frequency*

For a diode having $d = 1$ mm and $V = 10$ volts,

$$\tau_a = 10^{-9} \text{ sec.}$$

Thus, the power spectral density of the noise current component is essentially flat up to 80 MHz approximately. The power spectral density of the shot noise current may be considered to be constant equal to qI_o up to this frequency and is given by

$$S_i(f) = qI_o \tag{12.18}$$

In general, the power density spectrum of the shot noise current of an active component is given by

$$S_i(f) = KqI_o \tag{12.19}$$

where K is a constant and varies from device to device. For junction devices K also depends on the mode of biasing. In the case of bipolar junction transistor

$K = 2$. In active devices there are various other types of noise present. Some of these noise forms are discussed below.

(i) Partition Noise

Partition noise occurs whenever current has to divide between two or more paths, and results from the random fluctuations in the division. It is thus expected that a diode would be less noisy than a transistor (all other factors being equal). It is for this reason that the inputs of microwave receivers are often directly taken to diode mixers. The spectrum for partition noise is flat. Active three terminal components in which the control terminal draws less current are less noisy. In recent years, GaAs MESFETs (Metal-Semiconductor Field Effect Transistors) have been developed which draws almost zero gate bias current. These devices have low partition noise and therefore find application in low noise microwave amplification.

(ii) Low-Frequency or Flicker Noise

Below frequencies of a few kilohertz, a component of noise appears, the power spectral density of which increases as the frequency decreases. This is known as *flicker noise* (sometimes called *Ilf noise*). In vacuum tubes the main causes of flicker noise are slow changes which take place in the oxide structure of oxide coated cathodes and migration of impurity ions through the oxide.

In semiconductor devices flicker noise results from fluctuations in carrier density and is much more troublesome in semiconductor amplifying device than in vacuum tube counterparts at low frequencies. The fluctuation in the carrier density causes fluctuation in the conductivity of the material. This in turn, produces a fluctuating voltage drop when a direct current flows, which is the flicker-noise voltage. It follows that the mean square value of flicker noise voltage is proportional to the square of the direct current flowing.

(iii) High-Frequency or Transmit-Time Noise

In semiconductor devices, when the transit time of carriers crossing a junction is comparable with the time period of the signal, some of the carriers may diffuse back to the source or emitters. It can be shown that this gives rise to an input admittance in which the conductance component increases with frequency. This conductance has a noise current source associated with it in parallel. Since the conductance increases with frequency, the power spectral density also increases. A similar effect occurs in vacuum tubes when the transit time of electrons from cathode to control grid is comparable to the time period of the signal.

(iv) Generation-Recombination Noise

In semiconductor devices, some impurity centres will be ionized on a random basis, being energised thermally. Thus a random generation of carriers occurs

in semiconductor device. The carriers can also recombine with ionized impurity centres on a random basis, either directly or through trapping centres. The overall result is that the conductivity of the semiconductors have randomly fluctuating component which gives rise to a noise current when direct current flows through the semiconductance device. The power spectral density of this type of noise has yet to be established fully.

(b) Thermal Noise

This noise arises due to the random motion of free electrons in the conducting medium such as resistors. Each free electron inside a resistor is in motion due to its thermal energy. The path of individual electrom motion is random and zig-zag due to collosion with the lattice structure. The net effect of this random motion is that there is no preferred direction of the flow of a group of electrons and as a result the average current is zero. On the other hand, because of the random motion of the individual electrons, there is a finite value of the r.m.s. current. The power spectral density of the current due to free electrons is given by

$$S_i(f) = \frac{2kTG\alpha^2}{\alpha^2 + (2\pi f)^2} = \frac{2kTG}{1 + \left(2\pi\left(\frac{f}{\alpha}\right)\right)^2} \tag{12.20}$$

where k is the Boltzmann constant, T is the absolute temperature of the ambient, G is the conductance of the resistor and α is the average number of collisions per second of an electron.

The power spectral density of the thermal noise current is plotted as a function of $2\pi\left(\frac{f}{\alpha}\right)$ in Fig. 12.4. It is clearly seen that the spectrum of $S_i(f)$ may be assumed to be flat for $2\pi\left(\frac{f}{\alpha}\right) < 0.1$. The magnitude of α, the number of collisions per second is the order of 10^{14}. Hence, the spectrum is essentially flat up to frequencies in the range of 10^4 GHz. Thus, for all practical purposes, the power spectral density due to thermal noise of a resistor may be taken as

$$S_i(f) = 2kTG \tag{12.21}$$

Therefore, the power spectral density of noise is considered to be constant, and the contribution of any circuit is determined by the bandwidth of the circuit. Thermal noise may be considered to contain all frequency components in equal amount. For this reason thermal noise is often referred to as *white noise* (white signifying the presence of all colours or frequencies). Thermal noise has another name called *Johnson noise* after J.B. Johnson who made a detailed investigation of this noise in conductors.

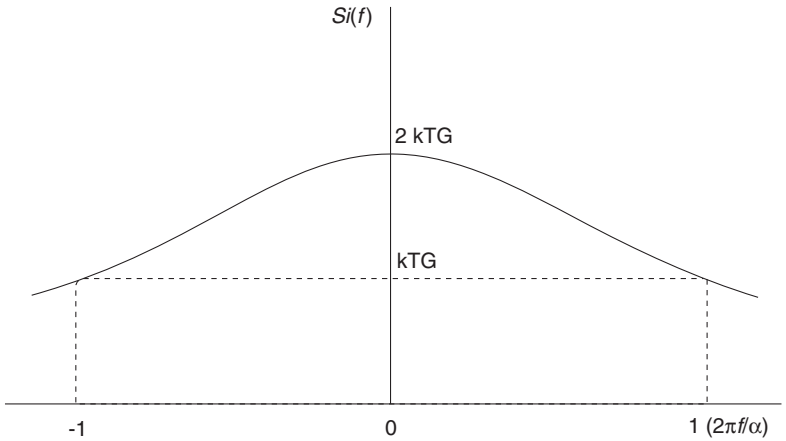


Fig. 12.4 *Power spectral density of thermal noise current versus frequency*

It is evident from previous discussion that a resistor R (Fig. 12.5) can be represented by a noiseless conductor $G = 1/R$ in parallel with a noise current source $i_n(t)$ with power spectral density $2kTG$ as shown in Fig. 12.5(b). Thevenin equivalent of this arrangement (Fig. 12.5b) is shown in Fig. 12.5(c) which is a voltage source equivalent circuit consisting of noise voltage source in series with a noiseless resistor R given by

$$v_n\,(t) = R\; i_n\,(t) \tag{12.22}$$

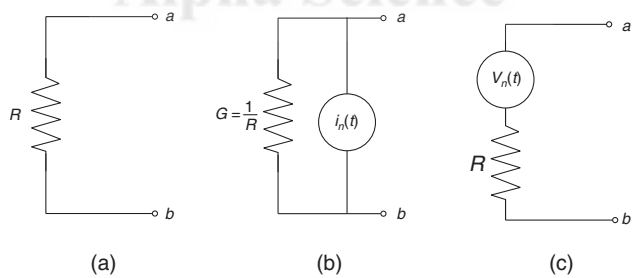


Fig. 12.5 *Noise equivalent circuit of a resistor*

Since the power density spectrum is a function of the square of the signal, the power spectral density, $S_v\,(f)$ of the noise voltage can be expressed in terms of power spectral density $S_i\,(f)$ of the noise current as

$$S_v\,(f) = R^2\; S_i\,(f) = 2kTR \tag{12.23}$$

From the interpretation of the power spectral density, the power carried by the frequency component Δf (Hz) centred at frequency f is

$$\Delta P = 2\; S_v\,(f)\; \Delta f = 4kTR\Delta f \tag{12.24}$$

Thus, if the noise voltage $v_n(t)$ is filtered by a narrow band filter of bandwidth Δf (Hz), the output voltage Δv_n of this filter will have a mean square value of

$$\overline{(\Delta v_n)^2} = 4kTR\Delta f \tag{12.25}$$

Since the major contribution of internal noise arises from thermal noise, we shall consider this noise in more details in subsequent sections.

(i) Resistors in Series

The noise voltage equivalent circuit of two resistors connected in series is shown in Fig. 12.6a. Since the two resistors in series can be replaced by a single resistor $R_s = R_1 + R_2$, the equivalent circuit can be drawn as shown in Fig. 12.6. Therefore, the mean squared noise voltage can be written as

$$\overline{v_n^2} = 4R_s\, kT\, \Delta f \tag{12.26}$$

where Δf is the bandwidth.

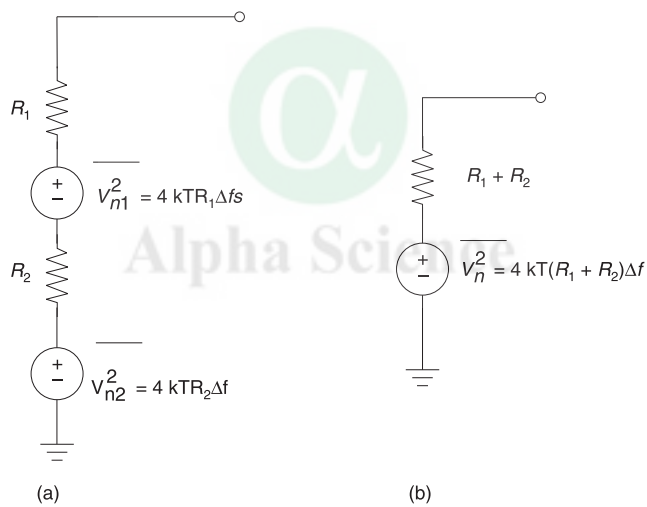


Fig. 12.6 *(a) Noise voltage circuit of two resistors connected in series. (b) Equivalent of two resistors connected in series*

Substituting, $R_s = R_1 + R_2$, in Eq. (12.26) we get

$$\begin{aligned}\overline{v_n^2} &= 4R_1\, kT\, \Delta f + 4R_2\, kT\, \Delta f \\ &= \overline{v_{n1}^2} + \overline{v_{n2}^2}\end{aligned} \tag{12.27}$$

This shows that the resultant noise voltage mean squared is equal to the sum of the mean square of the individual noise voltages, or alternatively,

$$\sqrt{\overline{v_n^2}} = \sqrt{\overline{v_{n1}^2} + \overline{v_{n2}^2}} \tag{12.28}$$

It should be noted that just adding the noise voltages $v_{n1} + v_{n2}$ would have given the wrong result. For a number of resistors R_1, R_2, … in series Eq. (12.27) can be extended as

$$\overline{v_n^2} = 4\,(R_1 + R_2 + R_3 + \ldots)\,kT\Delta f$$
$$\overline{v_n^2} = \overline{v_{n1}^2} + \overline{v_{n2}^2} + \overline{v_{n3}^2} + \ldots \tag{12.29}$$

(ii) Resistors in Parallel

For two resistors in parallel (Fig. 12.7a), the equivalent noise circuit is shown in Figs. 12.7(b) and (c). Here, it is simpler to work with conductance G and the noise current sources. Since two conductances in parallel can be replaced by a single conductance $G_p = G_1 + G_2$, the circuit can be equivalently represented as shown in Figs. 12.7(b) and (c). Thus,

$$\overline{i_n^2} = 4G_p\,kT\Delta f$$
$$= 4\,(G_1 + G_2)\,kT\Delta f$$
$$= \overline{i_{n1}^2} + \overline{i_{n2}^2} \tag{12.30}$$

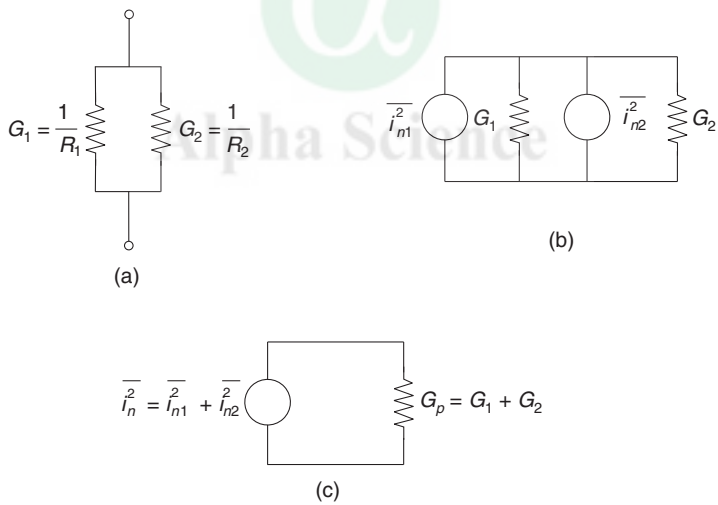


Fig. 12.7 *(a) Two resistors in parallel (b) and (c) Noise current equivalent circuit of two resistors in parallel*

The argument can be extended to any number of resistors in parallel to give

$$\overline{i_n^2} = 4\,(G_1 + G_2 + G_3 + \ldots)\,kT\Delta f \tag{12.31}$$

If it is desired to obtain the voltage-generator equivalent circuit for resistors in parallel, the equivalent parallel resistance R_p is to be determined where

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots \tag{12.32}$$

and

$$\overline{e_n^2} = 4R_p\, kT\Delta f \tag{12.33}$$

Example 12.1 *Two resistors, 22 kΩ and 47 kΩ, are at room temperature (300 K). Calculate, for a bandwidth of 100 kHz, the thermal noise voltage (a) for each resistor, (b) for the two resistors in series and (c) for the two resistors in parallel.*

Solution Here, $kT = 1.38 \times 10^{-23} \times 300$

$= 4.14 \times 10^{-21}$ J.

(a) For the 22 kΩ resistor

$$\overline{e_{n1}^2} = 4 \times (22 \times 10^3) \times 4.14 \times 10^{-21} \times 100 \times 10^3$$
$$= 36.4 \times 10^{-12}\ V^2$$

For the 47 kΩ resistor

$$\overline{e_{n2}^2} = 4 \times (47 \times 10^3) \times 4.14 \times 10^{-21} \times 100 \times 10^3$$
$$= 67.9 \times 10^{-12}\ V^2$$

(b) For the series combination

$$R_S = (22 + 47)\ \text{k}\Omega$$
$$= 69\ \text{k}\Omega$$

Therefore,

$$\sqrt{\overline{e_n^2}} = \left[\overline{e_{n1}^2} = \overline{e_{n2}^2}\right]^{1/2}$$
$$= [(36.4 + 67.9) \times 10^{-12}]^{1/2}$$
$$= 10.21\ \mu\text{V}$$

(c) For the parallel combinations of R_1 and R_2

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2}$$

That is,

$$R_p = \frac{22 \times 47}{22 + 47}$$
$$= 14.98\ \text{k}\Omega$$

Therefore,

$$\overline{e_n^2} = [4 \times 14.98 \times 10^3 \times 4.14 \times 10^{-21} \times 100 \times 10^3]^{1/2}$$
$$= 4.98\ \mu\text{V}$$

12.3 GENERALISATION OF THERMAL NOISE RELATIONSHIP

Consider a generalized circuit (two port network) consisting of linear, passive, bilateral circuit elements (such as resistors, inductors, capacitors). We shall extend the concept of thermal noise for single resistor or conductor to the case of this generalized circuit. A typical circuit is shown along with its Thevenin and Norton equivalent representations in Fig. 12.8. The admittance $Y_{ab}(f)$ across the terminal of the equivalent circuit in Fig. 12.8a can be written as

$$Y_{ab}(f) = G_{ab}(f) + jB_{ab}(f) \tag{12.34}$$

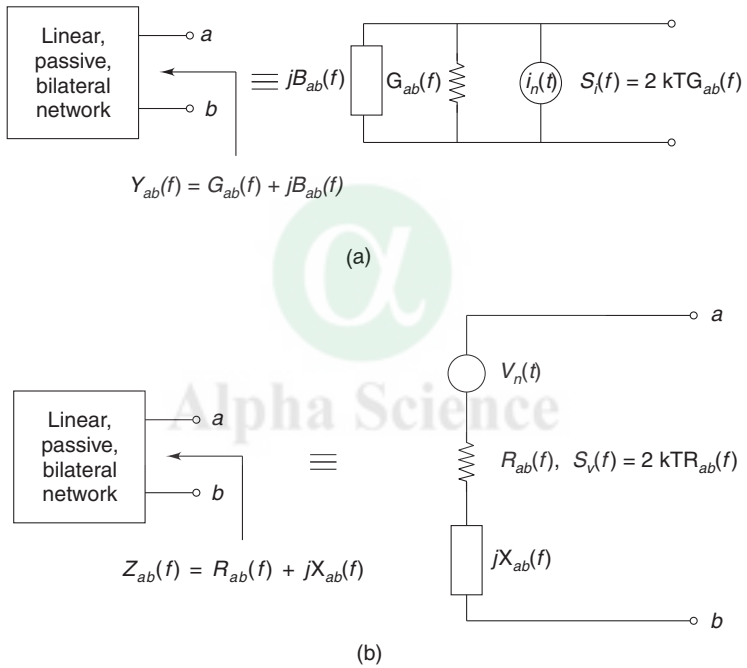


Fig. 12.8 *(a) RLC circuit with its Norton's equivalent representation. (b) RLC circuit with its Thevenin's equivalent representation*

Thus, the network can be represented by a noiseless admittance $Y_{ab}(f)$ in parallel with a noise current source of power density $S_i(f)$ given by

$$S_i(f) = 2kTG_{ab}(f) \tag{12.35}$$

Note that $G_{ab}(f)$, the equivalent conductance seen across the terminal *ab* varies with frequency in general. Hence, the equivalent noise current source has a power spectral density which is a function of frequency. An alternative form of noise equivalent circuit of the network can be obtained by using Thevenin Theorem (Fig. 12.8b).

The impedance $Z_{ab}(f)$ is given by

$$Z_{ab}(f) = R_{ab}(f) + jX_{ab}(f) \tag{12.36}$$

The equivalent circuit consists of a noiseless impedance $Z_{ab}(f)$ in series with a noise voltage source $v_n(t)$ of power spectral density $S_v(f)$ given by

$$S_v(f) = 2kT\, R_{ab}(f) \tag{12.37}$$

In general $R_{ab}(f)$, the real part of $Z_{ab}(f)$, is a function of frequency. Hence the power spectral density of the equivalent noise source is a function of frequency.

It should be noted that a major limitation of the theorem is that it applies to any linear bilateral passive circuit elements in which all of the noise sources are due to resistors only. Thus the theorem cannot be applied to circuits containing active components like vacuum tubes or transistor etc.

Example 12.2 *Determine the power spectral density of the thermal noise voltage across terminals ab of the circuit shown in Fig. 12.9. Also, find the power spectral density of the noise current.*

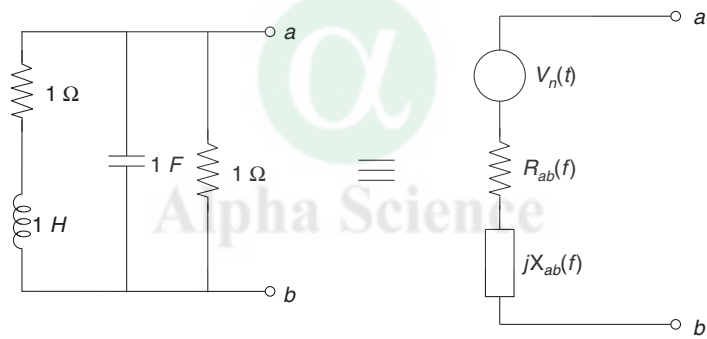


Fig. 12.9 *RLC circuit and its Thevenin's equivalent*

Solution The admittance $Y_{ab}(f)$ is given by

$$Y_{ab}(f) = 1 + j2\pi f + \frac{1}{1 + j2\pi f}$$

$$= \frac{2 - 4\pi^2 f^2 + 4j\pi f}{1 + j2\pi f} = \frac{2 + 4\pi^2 f^2 + j8\pi^3 f^3}{1 + 4\pi^2 f^2}$$

Hence the impedance $Z_{ab}(f)$ is given by

$$Z_{ab}(f) = \frac{1 + j2\pi f}{(2 - 4\pi^2 f^2) + j4\pi f}$$

$$= \frac{2 + 4\pi^2 f^2 - j8\pi^3 f^3}{(2 - 4\pi^2 f^2)^2 + (4\pi f)^2}$$

The power spectral density of the noise voltage $S_v(f)$ is given by

$$\begin{aligned} S_v(f) &= 2kT\, R_{ab}(f) \\ &= 2kT\, Re\,[Z_{ab}(f)] \\ &= \frac{2kT\,(2 + 4\pi^2 f^2)}{(2 - 4\pi^2 f^2) + (4\pi f)^2} \end{aligned}$$

The power spectral density of the noise current $S_i(f)$ is given by

$$\begin{aligned} S_i(f) &= 2kT\, R_e\,[Y_{ab}(f)] \\ &= 2\,kT\,\frac{2 + 4\pi^2 f^2}{1 + 4\pi^2 f^2} \end{aligned}$$

12.4 THERMAL NOISE CALCULATION FOR A SINGLE NOISE SOURCE

Consider a circuit consisting of noiseless elements as shown in Fig. 12.10. A random noise voltage $v_{ni}(t)$ is applied to the input terminals of the circuit. We wish to determine the root mean square value of the output noise voltage $v_{no}(t)$. Let $H(f)$ be the transfer function of the system. Let $S_{ni}(f)$ and $S_{no}(f)$ are the power spectral densities of the noise voltage signals $v_{ni}(t)$ and $v_{no}(t)$ respectively. Therefore,

$$S_{no}(f) = |H(f)|^2\, S_{ni}(f) \tag{12.38}$$

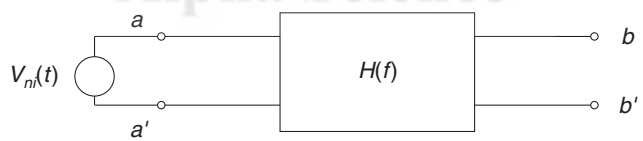


Fig. 12.10 *A single random noise source applied to a noiseless circuit*

The average power or the mean square value of the signal at the output is equal to the area under $S_{no}(f)$ versus frequency curve, that is

$$\begin{aligned} \overline{v_{no}^2(t)} &= \int_{-\infty}^{\infty} S_{no}(f)\, df \\ &= 2\int_{0}^{\infty} |H(f)|^2\, S_{ni}(f)\, df \end{aligned} \tag{12.39}$$

since the power density spectrum is an even function of frequency.

Thus, the r.m.s. value of $v_{no}(t)$ is

$$\sqrt{\overline{v_{no}^2(t)}} = \left[2\int_{0}^{\infty} |H(f)|^2\, S_{ni}(f)\, df\right]^{1/2} \tag{12.40}$$

12.5 WHITE NOISE AND ITS MATHEMATICAL REPRESENTATION

The noise analysis in communication system is based on an idealized form of noise process called *white noise*. The power spectral density of this noise is considered to be independent of frequency. The power spectral density of the white noise process $W(t)$ is represented as

$$S_W(f) = \frac{N_0}{2} \tag{12.41}$$

where the factor 1/2 has been included to indicate that half of the power associated with the positive frequency and remaining half with negative frequencies. This is illustrated in Fig. 12.11(a). The quantity N_0 has a dimension of W/Hz. The power spectral density of white noise shown in Fig. 12.11a shows that it has no dc power. In other words the mean or average value of white noise process is zero. The autocorrelation function of the power spectral density can be obtained by taking the inverse Fourier transform on both sides of the Eq. (12.41). The autocorrelation function for the white noise process can be written as

$$R_W(\tau) = \frac{N_0}{2}\delta(\tau) \tag{12.42}$$

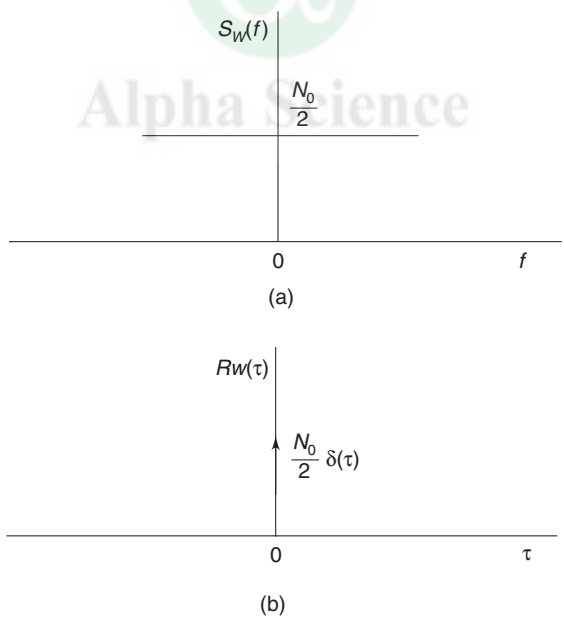


Fig. 12.11 *Graphical representation of white noise process (a) Power spectral density (b) Autocorrelation function*

Thus the autocorrelation function of the white noise consists of a delta function scaled by a factor $\frac{N_0}{2}$ and occurring $\tau = 0$. This is illustrated in Fig. 12.11b. Note that $R_W(\tau) = 0$ for $\tau \neq 0$. This means that no two different samples of white noise, no matter how close together they are in time have any correlation between them. In a strict sense white noise has infinite average power and is therefore physically realizable. However, white noise has convenient mathematical properties and therefore is useful for system analysis. The role of white noise process in noise analysis is similar to the one played by an impulse function or delta function in the linear system analysis. Like impulse function the effect of white noise is observed only after passing it through a system with finite bandwidth. It is therefore evident that as long as the bandwidth of the noise process at the input of the system is appreciably larger than the bandwidth of the system itself, we characterize the noise process as *white noise*.

12.6 MULTIPLE NOISE SOURCES: SUPERPOSITION OF POWER SPECTRAL DENSITY

In an electrical network there are a large number of noise sources arising from resistor and electronic components. All these sources are independent sources of random signals. It can be shown that for multiple noise sources applied to a linear system, the mean square value of the response is equal to the sum of the mean square values of the responses computed by assuming only one source at a time. This is also true for the power density spectra.

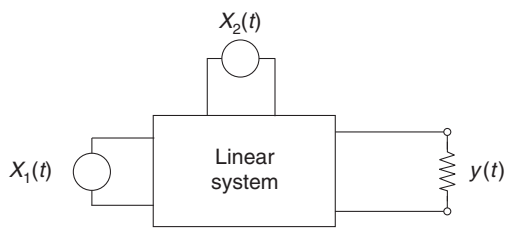


Fig. 12.12 *A linear system with two independent random noise signal input*

Consider a linear system with two independent random signal sources $x_1(t)$ and $x_2(t)$ at the input. Let $y(t)$ be the response of the system (see Fig. 12.12). Since the system is linear

$$y(t) = y_1(t) + y_2(t) \tag{12.43}$$

where $y_1(t)$ is the response of $x_1(t)$ alone (i.e., in the absence of $x_2(t)$) and $y_2(t)$ is the response of $x_2(t)$ alone (i.e. in the absence of $x_1(t)$)

From Eq. (12.43), we get

$$\overline{y^2(t)} = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} (y_1(t) + y_2(t))^2 \, dt$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} y_1^2(t) \, dt + \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} y_2^2(t) \, dt$$

$$+ \lim_{T \to \infty} \frac{2}{T} \int_{-T/2}^{T/2} y_1(t) \, y_2(t) \, dt$$

$$= \overline{y_1^2(t)} + \overline{y_2^2(t)} + 2 \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} y_1(t) \, y_2(t) \, dt \qquad (12.44)$$

The integral on the right hand side of Eq. (12.44) is the average value of the product $y_1(t)\, y_2(t)$. If $y_1(t)$ and $y_2(t)$ are independent random signals with zero mean, then the product $y_1(t)\, y_2(t)$ is equally likely to have positive and negative values. The average value of $y_1(t)\, y_2(t)$ is therefore zero. For a stable system, if the mean value of the input signal is zero, the mean value of the output signal will be also zero. Hence, the mean of $y_1(t)\, y_2(t)$ is zero. Thus, from Eq. (12.44) we get

$$\overline{y_2(t)} = \overline{y_1^2(t)} + \overline{y_2^2(t)} \qquad (12.45)$$

Expressing the mean square values i.e., average power of the signals in terms of power density spectra, we get from Eq. (12.45)

$$\int_{-\infty}^{\infty} S_y(f) \, df = \int_{-\infty}^{\infty} S_{y1}(f) \, df + \int_{-\infty}^{\infty} S_{y2}(f) \, df$$

$$= \int_{-\infty}^{\infty} (S_{y1}(f) + S_{y2}(f)) \, df \qquad (12.46)$$

where $S_y(f)$, $S_{y1}(f)$ and $S_{y2}(f)$ are the power spectral densities of $y(t)$, $y_1(t)$ and $y_2(t)$ respectively.

From Eq. (12.46), we may also write

$$S_y(f) = S_{y1}(f) + S_{y2}(f) \qquad (12.47)$$

Thus, we find that for s stable linear system, the power density spectra and the average power obey superposition principle for independent random noise signals.

12.7 EXCESS RESISTOR NOISE

The thermal-noise power density generated in a resistor does not vary with frequency. However, many resistors also generate a different type of noise which depends on frequency. This noise is called *excess noise*. The excess noise is inversely proportional to frequency and is called *l/f noise* and also pink noise.

The amount of excess noise generated in a resistor depends on its composition. For example, carbon resistors generate the largest amount of excess noise whereas the wire-wound resistors generate negligible amount of such noise. However, inductance associated with wire-wound resistors restricts their use in high frequency application. Metal film resistors are usually the best choice for high frequency communication circuits from noise point of view. The variation of the mean-square value of the excess resistor noise current with frequency has been depicted in Fig. 12.13.

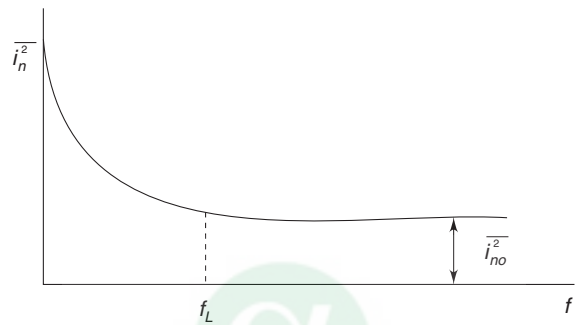


Fig. 12.13 *The plot of mean-square current due to excess resistor noise versus frequency*

12.8 NOISE IN ACTIVE CIRCUITS

In an active circuit there are various sources of noise. For example, in a transistor amplifier there is *thermal noise* generated in resistors plus the *shot-noise* generated by active components. In addition *flicker noise* (*l/f noise*) is also present. An equivalent circuit of the transistor amplifier identifying various sources of noise is shown in Fig. 12.14. Here, i_{n2} represents the *shot-noise* current density due to bias current in the output of the device and i_{n1} is the *shot-noise* current density due to the input bias current. The other noise source is due to the load resistor R_L.

The noise sources for commonly used transistors e.g., bipolar junction transistor (BJT), junction field-effect-transistor (JFET) and metal-oxide-semiconductor field-effect-transistor (MOSFET) are discussed below.

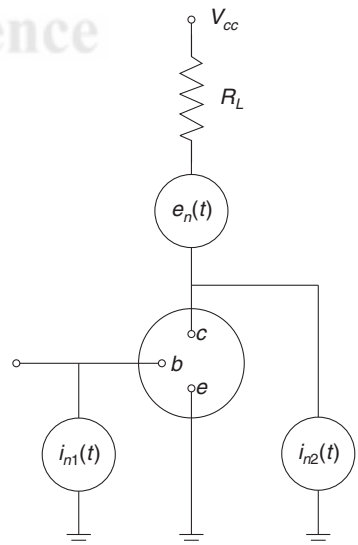


Fig. 12.14 *Noise-equivalent model of a bipolar junction transistor*

12.8.1 BJT Noise

The principal noise source in a bipolar junction transistor are the two shot noise sources and the thermal noise created by the base spreading resistance r'_b. These three noise sources are shown in Fig. 12.14 and are given by

$$\overline{e_n^2} = 4\ kTr'_b \quad V^2/\text{Hz} \tag{12.48}$$

$$\overline{i_{n1}^2} = 2\ qI_B \quad A^2/\text{Hz} \tag{12.49}$$

$$i_{n2}^2 = 2\ qI_C \quad A^2/\text{Hz} \tag{12.50}$$

where I_B and I_C are the base and collector current respectively. The noise current source $\overline{i_{n1}^2}$ is connected between the base and the emitter junctions and the other noise current source is connected between the collector and the emitter junctions.

12.8.2 FET Noise

The noise sources in field effect transistor (excluding 1/*f* noise) are given by

$$\overline{e_n^2} = \frac{2.8\ \text{kT}}{g_m} \quad V^2/\text{Hz} \tag{12.51}$$

and

$$\overline{i_n^2} = 2\ qI_g \quad A^2/\text{Hz} \tag{12.52}$$

where g_m is the trans-conductance and I_g is the gate leakage current of the field-effect-transistor. The noise sources of MOSFET (Metal-Oxide-Semiconductor-Field-Effect-Transistor) and JFET (Junction Field Effect Transistor) are the same but I_g is negligible for MOSFET. The shot noise increases with the frequency. At very high frequencies, the total noise current is

$$\overline{i_n^2} = 2\ qI_g + \frac{2.8\ \text{kT}}{g_m}(2\pi f C'_{gs})^2 \tag{12.53}$$

where C'_{gs} is approximately two third of the transistor gate-to-source capacitance.

12.9 NOISE FIGURE OF AN AMPLIFIER

From our previous discussion, it is evident that the signal is always contaminated with *noise*. Moreover, when a signal is processed by any system, additional noise is being added. The ratio of signal power to the noise power is a good indication of the purity of the signal. This ratio of signal to noise power is simply called *signal-to-noise ratio* (*SNR*) or *S*/*N*.

Consider the case of an amplifier. When a signal is amplified by an amplifier the noise generated by the amplifier is added to the original noise present in the signal. This causes deterioration in the *signal-to-noise ratio* of the output of the

amplifier compared to that at the input. The ratio of signal-to-noise power at the input to that at the output is an indication of the noisiness of the amplifier. In any amplifier, the noise present at the input (or noise generated in the source) is amplified and delivered to the load and the noise generated inside the amplifier is also delivered to the load. Thus the noise at the output is contributed both by the source and the amplifier.

We define the noise figure, $\overline{F}$ of an amplifier as this ratio of the total noise power in the load (at the output) to the noise power delivered to the load by the source alone. In terms of power density spectrum, it is the ratio of the total noise power density in the load (or at the output) to the noise power density delivered solely due to source. This is called *spectral noise figure*, F. Thus

$$F = \frac{S_{nto}(f)}{S_{nso}(f)} \tag{12.54}$$

where $S_{nto}(f)$ is the power spectral density of the total noise at the output of this amplifier and $S_{nso}(f)$ is the power spectral density of the noise at the output due to source only. As already mentioned, the total noise power spectral density present at the output of the amplifier consists of two components e.g., the power spectral density of the noise in the load solely due to amplifier, S_{nao} and that delivered to load due to source. Thus

$$S_{nto}(f) = S_{nso}(f) + S_{nao}(f) \tag{12.55}$$

From Eq. (12.54), the spectral noise figure can be written as

$$F = 1 + \frac{S_{nao}(f)}{S_{nso}(f)} \tag{12.56}$$

From Eq. (12.56), it is evident that the spectral noise figure is a measure of the noisiness of the amplifier.

Alternative Form

The spectral noise figure can be expressed in terms of the spectral signal-to-noise ratio at the input to that at the output. Let $S_{si}(f)$ and $S_{so}(f)$ represent the power spectral density of the desired signal at the input and the output respectively. thus

$$S_{so}(f) = |H(f)|^2 S_{si}(f) \tag{12.57}$$

where $H(f)$ is the transfer function of the amplifier.

Further, if $S_{nsi}(f)$ represent the power spectral density of the noise at the source (i.e. input) then

$$S_{nso}(f) = |H(f)|^2 S_{nsi}(f) \tag{12.58}$$

Thus, the spectral noise figure F can be written using Eq. (12.54) and (12.58) as

$$F = \frac{S_{nto}(f)}{|H(f)|^2 S_{nsi}(f)} \tag{12.59}$$

Substituting the value of $|H(f)|^2$ is terms of output and input signal power spectral density from Eq. (12.58) into Eq. (12.59), we get

$$F = \frac{\left(\frac{S_{si}(f)}{S_{nsi}(f)}\right)}{\left(\frac{S_{so}(f)}{S_{nto}(f)}\right)} \tag{12.60}$$

From Eq. (12.60) it is seen that the spectral noise figure is the ratio of the signal-to-noise power spectral density ratio at the input terminals to the signal-to-noise power spectral density ratio at the output terminals. Thus, the noise figure is a measure of the deterioration of the signal-to-noise power spectral density ratio in the process of amplification. By definition and also from equation (3), it is evident that the noise figure is always greater than unity. Hence, the signal-to-noise power spectrum ratio always deteriorates in the process of amplification. This deterioration is due to the noise introduced by the amplifier which also includes the noise introduced by the load impedance.

The noise figure we have discussed so far is the *spectral noise figure*. We may also define the noise figure in terms of total noise power at the output to the noise power contributed due to the source. Thus, the average noise figure $\overline{F}$ can be written as

$$\overline{F} = \frac{2\int_0^\infty S_{nto}(f)\,df}{2\int_0^\infty S_{nso}(f)\,df} = 1 + \frac{\int_0^\infty S_{nao}(f)\,df}{\int_0^\infty S_{nso}(f)\,df}$$

$$= 1 + \frac{N_{ao}}{N_{so}} \tag{12.61}$$

where, N_{ao} is the noise power at the output due solely to the amplifier and N_{so} is the noise power at the output due solely to the source noise.

From the definition of the noise figure we find that the noise figure of an amplifier is always greater than unity. For an ideal amplifier (noiseless one), it is unity. Thus the departure of the noise figure from unity is a measure of the *noisiness* of the amplifier. It must be noted that noise figure is not an absolute measure but a relative measure of the noisiness of the amplifier compared to source noise. It may appear from the expression (average or spectral) for noise figure that the noise figure can be made closer to unity by adding extra noise in the source. However, it is not a proper solution for improving the noise performance

of the amplifier. The reason is that when the noise in the source is increased, the amplifier only appears to be less noisy compared to the source. The overall signal-to-noise ratio at the output deteriorates badly in such cases making the output signal much more noisy. Thus the improvement of the noise figure of the amplifier is not possible by increasing the noise of the source alone. On the other hand the problem can be easily tackled by using a step-up transformer at the input which will increase both the input signal and the input noise. The increased noise at the source makes the amplifier look less noisy without deteriorating the signal-to-noise ratio at the input (since the signal is also increased by the same amount as the noise in the source by the step-up transformer). Thus the noise figure is reduced and the signal-to-noise ratio at the output terminals actually improves.

It may be pointed out here that in calculating the power density spectrum of the noise we may consider the output variable to be either voltage or current. The noise figure is independent of the output variable because voltage and current at the output are related by the same load impedance which means that in the ratio the load impedance cancels out. Noise figure is sometimes expressed in decibels. The average noise figure $\overline{F}$ in decibels is written as

$$\overline{F}_{\text{dB}} = 10 \log_{10} \overline{F} \tag{12.62}$$

since the minimum value of $\overline{F} = 1$, the noise figure of an ideal noiseless network is 0 dB.

12.10 AVAILABLE NOISE POWER SPECTRAL DENSITY

The power or power spectral density discussed so far refer to a *normalized* load of 1 ohm resistance. The normalized power of a signal $g(t)$ is defined as the power dissipated by a voltage source $g(t)$ across a 1 ohm resistor or by a current $g(t)$ flowing through 1 ohm resistor. If the voltage $g(t)$ is applied across a resistance of R ohms, the power dissipated will be obviously different. Denoting the normalized power by P_n, the power dissipated across the resistor R can be written as

$$P_R = \frac{P_n}{R} \tag{12.63}$$

The actual power dissipated by a signal $g(t)$ across a resistor of R ohms is $1/R$ times P_n, the *normalized power* of the signal $g(t)$.

The power spectral density represents the power dissipated for unit bandwidth of the frequency components of $g(t)$ across 1 ohms resistor. Therefore, the power density defined before is the normalized power density. Thus, the power spectral density dissipated across R will be

$$S_R(f) = \frac{S_n(f)}{R} \tag{12.64}$$

where $S_n(f)$ is the normalized power density of $g(t)$ denoted earlier by $S_g(f)$ and $S_R(f)$ is the actual power density dissipated across R.

12.11 AVAILABLE POWER DENSITY

Let us consider a voltage source $g(t)$ with a normalized power spectral density $S_g(f)$ applied from a source (having a source impedance Z_S) to a load Z_L (Fig. 12.15). Let the source impedance be given by

$$Z_S = R_S + jX_S \tag{12.65}$$

where R_S is the resistive component and X_S is the reactive component of the source impedance, Z_S.

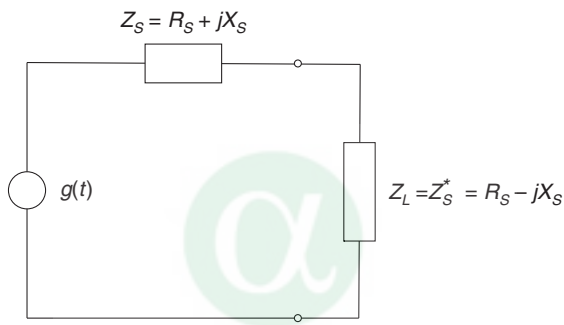


Fig. 12.15 *Illustrating available power spectral density*

From the maximum power transfer theorem, we know that the maximum power will be delivered to the load when the load impedance is the *complex conjugate* of the source impedance. This means that

$$\begin{aligned} Z_L &= Z^*_s \\ &= R_S - jX_S \end{aligned} \tag{12.66}$$

The matched condition is illustrated in Fig. 12.15. It can be easily seen that the signal $g(t)$ sees a resistance of $2R_S$ ohms and the actual power density dissipated by $g(t)$ ia

$$S'_g(f) = \frac{S_g(f)}{2R_S} \tag{12.67}$$

Out of the total power spectral density given by Eq. (12.67), one-half is dissipated across the source resistance and the remaining half is delivered to the load. Hence, the actual power density delivered to the load is

$$S_{av}(f) = \frac{S_g(f)}{4R_S} \tag{12.68}$$

$S_{av}(f)$ given by Eq. (12.68) represents the maximum power density that can be extracted from the signal $g(t)$. For this reason, the quantity $S_{av}(f)$ is called the *available power density*. Thus for a voltage source $g(t)$ with an internal source impedance $Z_S = R_S + jX_S$, the available power density is equal to $1/4R_S$ times the normalized power density of. In general, R_S is a function of frequency, and equation (4) can be written as

$$S_{av}(f) = \frac{S_g(f)}{4R_S(f)} \tag{12.69}$$

In a similar way, it can be easily verified that for a current source $i(t)$ having an internal admittance $Y_S(f) = G_S(f) + jB_S(f)$, $G_S(f)$ and $B_S(f)$ being the conductive and susceptive components of $Y_S(f)$, the available power density can be obtained as

$$S_{av}(f) = \frac{S_i(f)}{4G_S(f)} \tag{12.70}$$

where $S_i(f)$ is the normalized power spectral density of the current signal $i(t)$.

12.12 AVAILABLE NOISE POWER DENSITY OF AN R-L-C NETWORK

In this section we shall find the available power density due to thermal noise from a passive network containing RLC elements only. Such an RLC network can be represented by an equivalent impedance

$$Z_{ab}(f) = R_{ab}(f) + jX_{ab}(f) \tag{12.71}$$

This is shown in Fig. 12.16. The equivalent of the RLC network consists of a thermal noise voltage source of power density $2kTR_{ab}(f)$ in series with the impedance $Z_{ab}(f)$ gives by Eq. (1.71). In order to extract the maximum noise power out of this circuit we must use a load $Z_L(f)$ across the terminals ab, given by

$$\begin{aligned} Z_L(f) &= Z_{ab}^{*}(f) \\ &= R_{ab}(f) - jX_{ab}(f) \end{aligned} \tag{12.72}$$

The available power density of this source is given by (see Eq. (12.69))

$$S_{av}(f) = \frac{2kTR_{ab}(f)}{2R_{ab}(f)} = \frac{kT}{2} \tag{12.73}$$

Thus the available noise power (thermal) density from a passive RLC network is constant equal to $kT/2$. It has been assumed that all the resistors are at the same temperature T.

Using Eq. (12.73), the maximum noise power that can be extracted from an RLC network in a bandwidth, Δf is given by

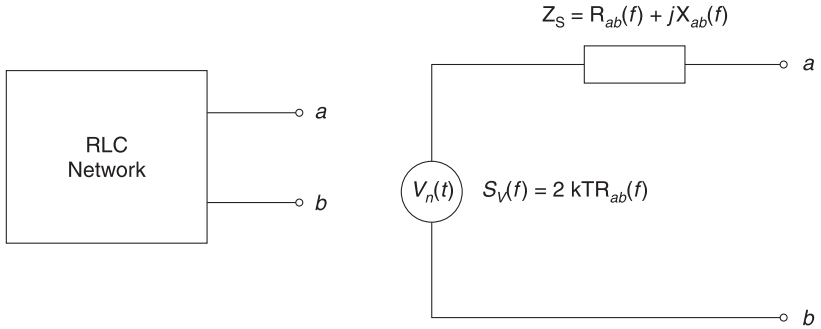


Fig. 12.16 *Passive network containing RLC elements*

$$\overline{v_{no}^2(t)} = 2\Delta f \frac{kT}{2} = kT\,\Delta f \tag{12.74}$$

where $v_{no}(t)$ is the output noise voltage delivered to the load under matched load condition.

12.13 EFFECTIVE NOISE TEMPERATURE

The available power density of any two terminal RLC network is $kT/2$ (see Eq. (12.73). This is not however, true for networks containing sources of noise other than thermal noise. Nevertheless, it is possible to express the available noise power density from any two terminal network which contains thermal and other form of noise in the form given by equation (3) by introducing the *effective noise temperature* T_N. If the available noise power spectral density from any two terminal RLC network is $S_{av}(f)$, then we may write

$$S_{av}(f) = \frac{kT_N}{2} \tag{12.75a}$$

that is,

$$T_N = \frac{2S_{av}(f)}{k} \tag{12.75b}$$

If $S_{av}(f)$ is a constant over the frequency range of interest, then the effective noise temperature is also constant given by Eq. (12.75b). If $S_{av}(f)$ is a function of frequency, the effective noise temperature T_N also becomes a function of frequency.

The available noise power from the two terminal network in the bandwidth Δf can be written as

$$\overline{v_{no}^2(t)} = kT_N\,\Delta f \tag{12.76}$$

For RLC network the effective noise temperature is same as that of ambient temperature, T. This is because in RLC networks the source of noise is only thermal which is given by Eq. (12.73).

Questions

12.1 Explain what do you understand by "Noise".

12.2 Define "Noise". Give its classification based on its origins.

12.3 What is shot noise? Derive an expression for its power spectral density as a function of frequency for a vacuum diode.

12.4 Write short notes on:

(i) Partition noise

(ii) Flicker noise

(iii) Transit-time noise

(iv) Thermal noise

12.5 For two resistors connected in series, draw noise voltage equivalent circuit and derive an expression for the resultant noise voltage.

12.6 For two resistors connected in parallel, draw the noise current equivalent circuit and derive an expression for the resultant near squared noise voltage.

12.7 What is white noise? Explain its graphical representation.

12.8 Show that for a stable linear system, the power density spectra and the average power obey superposition principle for independent random noise signals.

12.9 Explain noise in (i) a BJT, (ii) a FET

12.10 Explain spectral noise figure of an amplifier.

Index

Acceptor atoms 2.8
Acceptor-level 2.8
A.C. load line 5.37
Active network 1.12
Active region 4.15
Admittance 1.73
Alternator 8.1
Amplification factor 6.5, 10.17
Amplifier 6.26
Amplitude modulation 11.3
Anderson Bridge 1.90
Apparent power 1.56
Astable multivibrator 9.6
Atmospheric noise 12.2
Avalanche breakdown 2.46

Balanced modulation 11.25
Bandwidth 6.50
Base modulation 11.21
Base Transport Factor 4.6
Beta 4.16
Bilateral Circuit 1.11
Bipolar junction transistors 4.1
Bistable multivibrator 9.7
BJT Noise 12.21
Blocking capacitor 6.40
Branch 1.10
Bridge rectifier 3.12
Built in potential barrier 2.28

Capacitive reactance 1.62
Capacitor input filter 3.26
Carrier wave 11.2
Cascading 6.28
Circuit 1.10
Circuit element 1.10
Clampers 2.60
Class A 6.27
Class AB 6.27
Class B 6.27
Class C amplifier 6.27
Clippers 2.52
Collector efficiency 4.6
Collector modulation 11.18
Colpitt's oscillator 8.10
Conductance 1.1
Conduction band 2.3
Conductivity 1.2
Connection 1.7
Continuity equations 2.21
Cosmic noise 12.3
Current feedback biasing 5.16
Current source 1.11
Cut off region 4.9

D.C. load line 5.35
D.C. load line 5.1
Degree of amplitude modulation 11.7
Delta 1.7
Demodulation 11.2
Demodulation 11.47
Depletion MOSFET 10.38
Depletion region 2.28
Depletion region capacitance 2.31
De Sauty bridge 1.91
Detection 11.1, 11.47
Diffusion 2.10
Diffusion constant 2.16
Diffusion current 2.11
Dissipation factor 1.93
Distributed network 1.10
Donor atom 2.7

Doping 2.1
Double-diode clipper 2.59
Drain resistance 10.17
Drift 2.10
Drift current 2.10
DSBSC system 11.14
Duality 1.4
Dual networks 1.6
Dynamic characteristic 2.51

Early effect 4.10
Effective noise temperature 12.27
Electroluminescence 3.56
Emitter efficiency γ 4.5
Emitter modulation 11.22
Enhancement MOSFET 10.39
Envelope detector 11.49
Epitaxial growth 2.25
Esaki diode 3.42
Excess noise 12.19
External noise 12.2
Extrinsic semiconductor 2.7

Feedback 8.6
Feedback circuit 7.1
Fermi-Dirac energy distribution function 2.11
Fermi level 2.12
FET 4.1
FET Noise 12.21
Field Effect transistor 10.11
Field-effect transistor (FET) 10.8
Filter circuit 2.3
Flicker noise 12.8
Forbidden gap 6.36
Frequency modulation 11.29
Frequency response 3.8

Gain-bandwidth 6.50
Gain bandwidth product 7.10
Generation-Recombination noise 12.8

Half wave rectifier 3.2
Hartley oscillator 8.12
Holding current 10.45
Hot carrier effect 2.22
h parameters 6.3
Hybrid parameters 6.3

Ideal diode 2.47
Impedance 1.64
Impedance diagram 1.64
Internal noise 12.2
Intrinsic semiconductor 2.5

Johnson noise 12.9
j operator 1.58

Kirchhoff's current law 1.78
Kirchhoff's voltage law 1.17, 1.78

1/f noise 12.19
Light-emitting diode 3.56
Linear circuit 1.11
Linear element 1.12
Linear networks 1.12
Load line 5.35
Load line concept 2.50
Loop 1.12
Loop-gain 7.3
Lower cut-off frequency 6.46
L-section filter 3.24
Lumped networks 1.10

Man-made noise 12.3
Mass Action Law 2.8
Maximum Power Transfer Theorem 1.39
Maximum power transfer theorems 1.82
Maxwell's Inductance Bridge 1.86
Maxwell's L/C Bridge 1.88
Mesh 1.12
Metallurgical junction 2.26
Mobility 2.10
Modulated wave 11.2
Modulating wave 11.2
Modulation 11.1
Modulation index 11.5
Monostable multivibrator 9.7

MOSFET 10.37
Multivibrator 9.5

Negative feedback equation 7.3
Network 1.11
Node 1.12
Noise 12.1
Non-linear circuit 1.11
Non-linear resistors 2.1
Norton's theorem 1.32, 1.81
Nyquist criteria 7.23
Nyquist plot 7.22

Ohm's Law 1.1
Operating point 5.1
Oscillator 7.4, 8.1

Partition noise 12.8
Passive network 1.11
Peak inverse voltage 3.11
Percentage regulation 3.7
π-filter 3.26
phase angle 1.53
Phase modulation 11.3, 3.53
Phase reversal 5.30
Phase shift oscillator 8.14
Phasor 1.57
Photoconductivity 3.50
Photoconductor 3.51
Photodiode 3.54
Photoresistor 3.51
Photovoltaic cells 3.59
Pie connection 1.9
Pinch-off effect 10.14
Pinch-off voltage 10.11
Pinch-off" voltage 10.15
Polar notation of phasors 1.59
Positive feedback 8.6
Positive feedback equation 7.3
Potential divider bias 10.27
Potential source 1.11
Power factor 1.56
Pulse amplitude modulations 11.39
Pulse code modulation 11.4, 11.42
Pulse duration or width modulation 11.41
Pulse modulation 11.4
Pulse position modulation 11.42
Punch through 4.10

Q point 2.50
Quiescent point 5.1, 5.2, 10.23

Random signals 12.1
R-C coupled amplifier 6.31
Reactance 1.61
Rectifier equation 2.39
Regulation 3.7
Relaxation oscillator 10.6
Resistance 1.1
Resistivity 1.1
Ripple factor 3.6
rms value 1.55

Saturation region 4.9
Schering bridge 1.92
Self-biasing 10.26
Semiconductors 2.1
Shot noise 12.4
Signal-to-noise ratio 12.21
Silicon controlled rectifier 10.42
Single sideband generator 11.26
Sinusoidal oscillator 8.3
Solar noise 12.2
Source follower 10.33
Space-charge region 2.28
Square law detector 11.48
Square law diode modulator 11.23
Square-wave generator 8.3
SSBSC system 11.14
SSBTC system 11.14
Stability factor 5.8
Stabilization 5.6
Star 1.7
Step Graded junction 2.32
Superposition theorem 1.35
Susceptance 1.74

Switching circuit 9.1

T connection 1.9
Thermal noise 12.9
Thermal runaway 5.8
Thermal voltage 2.31
Thevenin's equivalent resistance 1.29
Thevenin's theorem 1.27, 1.80
Transconductance 6.5, 4.13
Transfer 2.51
Transistor 4.1
Transistor as a switch 9.1
Transition capacitance 3.48
Transition region 2.28
Transmission characteristic 2.51
Transmit-time noise 12.8
Transresistance 6.5
Triggering voltage 10.45
Tunnel diode 3.42
two-stage *R-C* coupled amplifier 6.39
T(*Y*) and π (delta) networks 1.20

Unijunction transistor 10.1
Unilateral circuit 1.11
Upper cut-off frequency 6.47

Valence band 2.3
Varactor 3.47
Voltage divider bias 5.17
Voltage regulation 3.37
Volt-equivalent of temperature 2.17
VSB modulation 11.15

Wheatstone bridge 1.50
White noise 12.9, 12.17
Wien bridge oscillator 8.17
Wien parallel bridge 1.94

Zener diodes 3.28
Zener effect 2.46