EC6503 TRANSMISSION LINES AND WAVEGUIDES
TWO MARKS QUESTION & ANSWERS

UNIT I - TRANSMISSION LINE THEORY

PART-A

1) What is Transmission Line?

A Transmission line is a mechanism of guiding electrical energy from one place to another. (or) Transmission line is an electrical line which is used to transmit electrical waves from one point to another.

Eg: i. Transfer of RF energy from transmitter to antenna. ii. Transmission lines can also be used as impedance transformers.

2) What are the types of Transmission Line?
   i) Open wire line
   ii) Cables
   iii) Micro strip lines
   iv) Coaxial Line
   v) Waveguides
   vi) Optical fibres

3) Write the applications of Transmission Line.
   1. Used to transfer energy from one circuit to another.
   2. Used as circuit element like inductor, capacitor and so on.
   3. Used as impedance matching device.
   4. Used of smith chart for calculating the l_{sc} and l_{oc}.
   5. Used as Frequency response of Resonant Circuit.

4) Define the Transmission Line Parameters. (N/D-18)

The parameters of a transmission line are:

i) Resistance (R): It is uniformly distributed along the length of the conductor and cross section area of the conductor. Its unit is ohm/unit length.

ii) Inductance (L): When a conductor carries a current, the conductor will be surrounded and linked by magnetic flux. The flux leakages per current give rise to the effect of inductance. Its unit is Henry/ unit length.

iii) Capacitance (C): The conductors are separated by insulating dielectric, so that capacitance will be distributed along the conductor length. Its unit is Farad/unit length.

iv) Conductance (G): The insulators of the open wire line may not be perfect and a leakage current will flow and leakage conductance will exists between conductors. Its unit is mho/unit length.

5) What are the secondary constants of a line? Why the line parameters are called distributed elements?

The secondary constants of a line are:

Characteristic Impedance (Z_0)

Propagation Constant (γ)

where $\gamma = \alpha + j\beta$ the constants α=attenuation constant, β= phase constant.

Since the line constants R, L, C, G are distributed through the entire length of the line, they are called as distributed elements. They are also called as primary constants.

6) Define Characteristic impedance (Z_0). (N/D-17, M/J-16)

Characteristic impedance is the impedance measured at the sending end of the line. **(or)**
The characteristic impedance or surge impedance (usually written Z_0) of a uniform transmission line is the ratio of the amplitudes of voltage and current of a single wave propagating along the line; that is, a wave travelling in one direction in the absence of reflections in the other direction. It is given by
$Z_0 = \sqrt{Z/Y}$, where $Z = R + j\omega L$ is the series impedance and $Y = G + j\omega C$ is the shunt admittance.

7) Define Propagation constant (γ).

Propagation constant is defined as the natural logarithm of the ratio of the sending end current or voltage to the receiving end current or voltage of the line. It gives the manner in the wave is propagated along a line and specifies the variation of voltage and current in the line as a function of distance. Propagation constant is a complex quantity and is expressed as $\gamma = \alpha + j\beta$
α = real part called the attenuation constant and β=imaginary part called the phase constant.

8) What is a finite line? Write down the significance of this line?

A finite line is a line having a finite length on the line. It is a line, which is terminated, in its characteristic impedance ($Z_R=Z_0$), so the input impedance of the finite line is equal to the characteristic impedance ($Z_s=Z_0$).

9) What is an infinite line?

An infinite line is a line in which the length of the transmission line is infinite. A finite line, which is terminated in its characteristic impedance, is termed as infinite line. So for an infinite line, the input impedance is equivalent to the characteristic impedance ($Z_s=Z_0=Z_R$).

10) What is wavelength of a line?

The distance the wave travels along the line while the phase angle is changing through 2π radians is called a wavelength (λ).
$$\lambda = 2\pi/\beta \text{ ; where } \beta \text{ - phase angle is defined as radian/km.}$$
$$\lambda = vt = v/f; v \text{ - velocity}$$

11) Define velocity of propagation?

The velocity of propagation is based on the observation of the change in phase along the line and is measured in m/s. (v).
$$v = \lambda t = 2\pi f/\beta$$
$$v = \omega/\beta \text{ ; } \omega = 2\pi f.$$

12) What are the types of line distortions?

The distortions occurring in the transmission line are called waveform distortion or line distortion.
Waveform distortion is of two types:
a) Frequency distortion
b) Phase or Delay Distortion.

13) How frequency distortion occurs in a line?

When a signal having many frequency components are transmitted along the line, all the frequencies will not have equal attenuation and hence the received end waveform will not be identical with the input waveform at the sending end because each frequency is having different attenuation. This type of distortion is called frequency distortion.

14) What is delay distortion?

When a signal having many frequency components are transmitted along the line, all the frequencies will not have the same time of transmission, some frequencies being delayed more than others. So the received end waveform will not be identical with the input waveform at the sending end because some frequency components will be delayed more than those of other frequencies. This type of distortion is called phase or delay distortion.

15) How to avoid the frequency distortion that occurs in the line?

In order to reduce frequency distortion occurring in the line, a) The attenuation constant α should be made independent of frequency. b) By using equalizers at the line terminals which minimize the frequency distortion.

16) What are Equalizers?

Equalizers are networks whose frequency and phase characteristics are adjusted to be inverse to those of the lines, which result in a uniform frequency response over the desired frequency band, and hence the attenuation is equal for all the frequencies.

17) How to avoid the frequency distortion that occurs in the line?

Phase distortion can be avoided by the use of coaxial cables. in such cables the internal inductance is low at high frequencies because of skin effect, the resistance is small because of large conductors, the capacitance and leakage are small because of the use of dielectric.

18) What is a distortion less line? What is the condition for a distortion less line? (A/M-18, N/D-18, N/D-17, N/D-16)

A line, which has neither frequency distortion nor phase distortion is called a distortion less line. The condition for a distortion less line is RC=LG. Also,
a) The attenuation constant should be made independent of frequency.
b) The phase constant should be made dependent of frequency.
c) The velocity of propagation is independent of frequency.

19) What is the drawback of using ordinary telephone cables?

In ordinary telephone cables, the wires are insulated with paper and twisted in pairs, therefore there will not be flux linkage between the wires, which results in negligible inductance, and conductance. If this is the case, there occurs frequency and phase distortion in the line.

20) How the telephone line can be made a distortion less line?

For the telephone cable to be distortion less line, the inductance value should be increased by placing lumped inductors along the line.

21) What is Loading?

Loading is the process of increasing the inductance value by placing lumped inductors at specific intervals along the line, which avoids the distortion.

22) What are the types of loading?

a) Continuous loading
b) Patch loading
c) Lumped loading

23) What is continuous loading?

Continuous loading is the process of increasing the inductance value by placing tapes of magnetic materials such as perm alloy or a iron core or a magnetic tape over the conductor of the line.

24) What is patch loading?
It is the process of using sections of continuously loaded cables separated by sections of unloaded cables which increases the inductance value.

25) What is lumped loading?
Lumped loading is the process of increasing the inductance value by placing lumped inductors at specific intervals along the line, which avoids the distortion.

26) What are the advantages and disadvantages of continuous loading?
> **Advantages:**
>> a) Attenuation is independent of frequency.
>> b) α is decreased by increasing L provided R is not increased.
>> c) Increase in L up to 100mH/unit length is possible.
> **Disadvantages:**
>> a) Expensive.
>> b) Achieves a small increase in L per unit length.
>> c) Existing methods cannot be modified by this method. Only replacement of line

is possible.

27) What are the advantages and disadvantages of Lumped loading?
> **Advantages:**
>> a) There is no limit to a value by which the inductance is increased.
>> b) Cost is low.
>> c) Existing lines can be modified by this method.
> **Disadvantages:**
>> a) For cables, Z_2 is essentially capacitive.
>> b) Cable capacitance and lumped inductances form a low pass filter.
>> c) For frequencies below cut-off, $f_c = 1/\pi\sqrt{LC}$ and the attenuation is reduced, but

above cut-off the attenuation rises as a result of filter action.

28) State Campbell's Equation.
$$\cosh N\ \gamma' = \cosh N\ \gamma + Z_c\ \sinh N\ \gamma\ /\ 2Z_0$$
> where, Z_c = is the loading coil impedance
>> γ = Propagation constant before Loading
>> γ' = Propagation constant after Loading
>> N = distance between the loading coils

29) When does the reflection occurs in a transmission line?
The load impedance is not terminated with characteristic impedance reflection takes place, (i.e) $Z_R \neq Z_0$, reflection occurs. Reflection is maximum when the line is open and short circuit.

30) Write the disadvantages of reflection.
> i) Reflected waves appears as echo on the sending end.
> ii) Efficiency is reduced.
> iii) The received energy is rejected by load and output reduces.
> iv) If the impedance of the generator is not Z_0, the reflected wave is reflected again at the

sending end, becoming a new incident wav. The energy is thus transmitted back on the line until dissipated in the line losses.

31) Define reflection coefficient.
Reflection Coefficient can be defined as the ratio of the amplitude of the reflected voltage to the incident voltage at the receiving end of the line Reflection Coefficient and is denoted by K.

K=Reflected voltage at load / Incident voltage at the load

$K = V_r / V_i$; $K = Z_R - Z_0 / Z_R + Z_0$

32) Define reflection factor.
The ratio which indicates the change in current in the load due to reflection at the mismatched junction is called reflection factor. It is denoted by k and defined by

$k = reflection\ factor = \left| 2\sqrt{Z_R Z_0} / Z_R + Z_0 \right|$

33) Define reflection loss. (A/M-18, M/J-16,
Reflection loss is defined as the number of nepers or decibels by which the current in the load under image matched conditions would exceed the current actually flowing in the load.
Reflection loss is inversely proportional to reflection factor k.

$Reflection\ loss,\ nepers = \ln \left| Z_R + Z_0 / 2\sqrt{Z_R Z_0} \right|$

$Reflection\ loss,\ decibels = 20\log \left| Z_R + Z_0 / 2\sqrt{Z_R Z_0} \right|$

34) Define the term insertion loss (N/D-18)
The insertion loss of a line or network is defined as the number of nepers or decibels by which the current in the load is changed by the insertion.
Insertion loss = Current flowing in the load without insertion of the network / Current flowing in the load with insertion of the network.

$Insertion\ loss\ in\ nepers = \ln \left| I_{R'} / I_R \right| = \ln 1/k_S + \ln 1/k_R + \ln 1/k_{SR} + \alpha l$

$Insertion\ loss\ in\ decibels = 20\ [\ \log 1/k_S + \log 1/k_R + \log 1/k_{SR} + 0.43\alpha L]$

35) What are the conditions for a perfect line?
For a perfect line, the resistance and the leakage conductance value were neglected. The conditions for a perfect line are R=G=0.

36) What is a smooth line?
A smooth line is one in which the load is terminated by its characteristic impedance and no reflections occur in such a line. It is also called as flat line.

37) Draw the equivalent circuit for transmission line.

EQ. Circuit of Transmission line
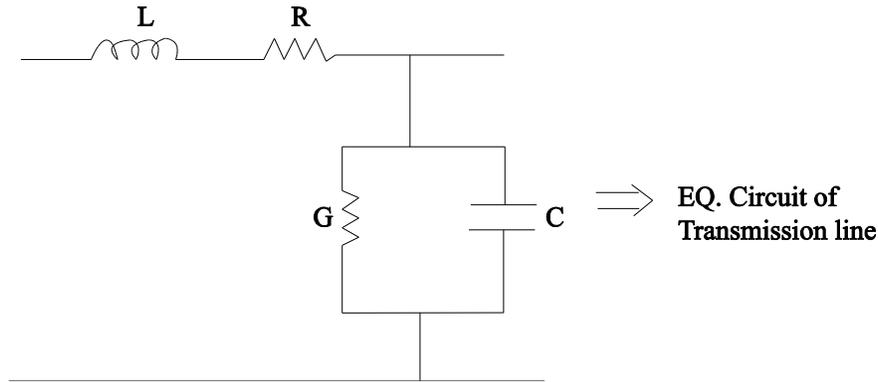
PART-B

1) Explain about differnt types of transmission lines. (A/M-15).

1. Open wire
2. Co-axial cables
3. Wave guides
4. Micro strip lines
5. OFC

Open wire:
These lines are the parallel conductors open through. Air and the conductors are separated by air as a dielectric. Eg: Telephone lines.

Cables:
The telephone cables consisting of 100 of conductors which are individually insulated with papers and these are twisted in pairs, (180° Phase Shift) and combine together and placed inside the protective lead or plastic sheet.

Co-axial Cable:
There are two conductors which are co-axially placed, one is hollow and other is placed inside the first conductor. The dielectric may be solid or gas. These lines are used for high voltage levels.

Wave Guides:
1. It is used for microwave frequency (i.e.) $3GHz - 30GHz(10^9)$
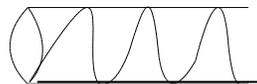
2. It follow conducting tubes having uniform cross reaction the energy is transmitted by inner walls of the tube by the phenomenon of total internal reflection.

2) Deduce the expressions for characteristic impedance and propagation constant of a line of cascade identical and symmentrical T sections of impedances. (N/D-11).

R, L, C, G – Primary Constants
Consider a transmission lines as a two conductor separated by dielectric.
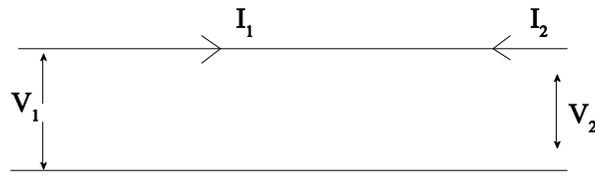
1. In a transmission line the terminal behaviour is not be same $V_1 \neq V_2$ $I_1 \neq I_2$

$V_2 - V_1 = \Delta V$
$I_2 - I_1 = \Delta I$
Where ΔV is the drop in series arm and ΔI is a drop in short arm.

2. Due to the drop, the voltage between the i/p and o/p there is a presence of resistor in the series arm.

3. Since the two wires separated by dielectric. There is capacitance and this capacitance not ideal and it has leakage inductance which should be ideally zero.

4. Initially the voltage and current on a line are in phase and later there is a phase shift which is due to the presence of inductance.
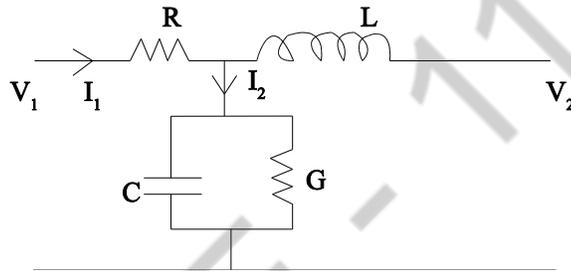
R, L, C, G are called primary constants of transmission line.
Z – Series impedance
Y – Shunt admittance

Let Z is equal to $Z = R + jwL$
$Y = G + jwC$

The characteristics of transmission line can be studied using the following parameter.

1) Characteristic impedence (Zo)
2) Propagation constant (γ) $\Big\}$ → Secondary constants

$$Z_o = \sqrt{Z/Y} = \sqrt{\frac{R + jwL}{G + jwC}} \text{ ohms}$$

$$\gamma = \sqrt{ZY} = \sqrt{(R + jwL)(G + jwC)}$$

Transmission line as a cascaded T section.

To study the behavior of TL it can be considered as a no. of identical T section connected in series and the last section is terminated as Z_o input impedance of the first section also Z_o.
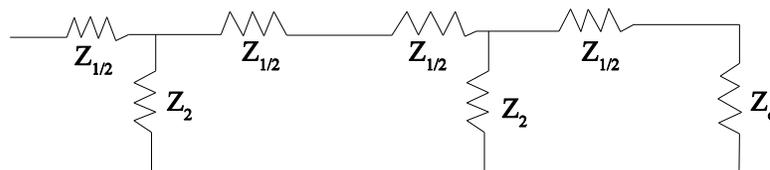
The two part n/w form of transmission line
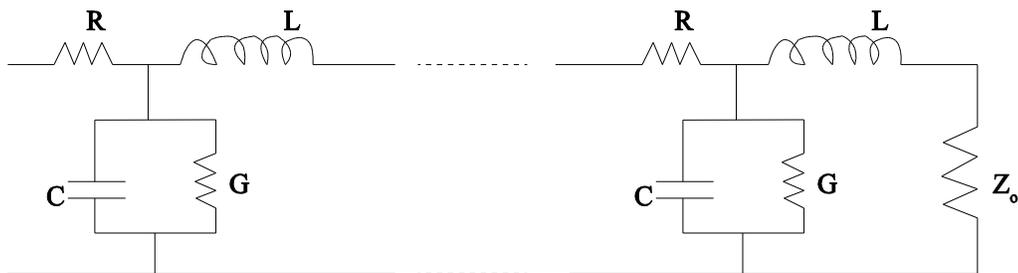
Note: Since the transmission line is introducing delay the signal is distorted. Since it is a wire it having some resistance and attenuation in the signal.

3) Discuss the general Solution of transmission line in detail. (N/D-17),(A/M-15),(N/D-11). (or) Derive the transmission line equation and hence obtain expression for voltage and current on a transmission line. (M/J-16),(M/J-14),(M/J-13),(M/J-12),(A/M-18).

GENERAL SOLUTION OF TRANSMISSION LINE:

i)   Consider equivalent circuit of transmission line.
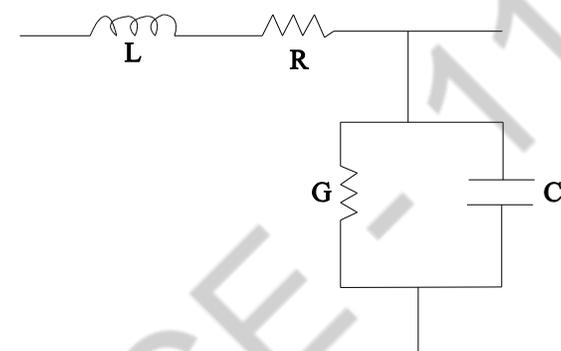
ii)   Let dx be the incremental length of line.

iii)   Series impedance $\begin{pmatrix} Z = R + jwL \\ Y = G + jwC \end{pmatrix}$

iv)   Shunt admittance

v)   Voltage drop across Z of T section is,

$$\left(\cancel{V} + dV \right) - \cancel{V} = I(R + jwL)dx$$

$(V + dV) \Rightarrow$ higher potential

$V \Rightarrow$ lower potential

$$dV = I(R + jwL)dx$$

$$\frac{dV}{dx} = I(R + jwL) \qquad\qquad \left[\because Z = R + jwL \right]$$

$$\frac{dV}{dx} = IZ \qquad\qquad 1$$

$\|^{by}$ $\left(\cancel{I} + dI \right) - \cancel{I} = V(G + jwc)dx$

$$dI = V(G + jwc)dx$$

$$\frac{dI}{dx} = VY \qquad\qquad 2$$

vi)   Differentiate eq 1 & 2 w.r.to   x

$$\frac{d^2V}{dx^2} = \frac{dI}{dx}Z$$

$$= (VY)Z \qquad \left[\because \frac{dI}{dx} = VY\right]$$

$$= VYZ$$

W.K.T $\quad \upsilon = \sqrt{YZ}$

$$\upsilon^2 = YZ$$

$$\therefore \quad \frac{d^2V}{dx^2} = \upsilon^2 V \qquad\qquad 3$$

$|||^{by}$ $\quad \dfrac{d^2I}{dx^2} = \dfrac{dV}{dx} Y = IZY \qquad \left[\because \frac{dV}{dx} = IZ\right]$

$$\frac{d^2I}{dx^2} = \upsilon^2 I \qquad\qquad 4$$

$$\frac{d^2V}{dx^2} = \upsilon^2 V; \quad \frac{d^2I}{dx^2} = \upsilon^2 I$$

This is a second order differential equation

The solution for the equation is

$V = A\,e^{\upsilon x} + B\,e^{-\upsilon x} \qquad\qquad$ 5a

$I = C\,e^{\upsilon x} + D\,e^{-\upsilon x} \qquad\qquad$ 5b

Where A, B, C, D are arbitrary constant

Diff equ 5a & 5b w.r.to x,

$$\frac{dV}{dx} = A\,\upsilon\,e^{\upsilon x} - B\,\upsilon\,e^{-\upsilon x}$$

$$ZI = \upsilon\,A\,e^{\upsilon x} - \upsilon\,B\,e^{-\upsilon x}$$

W.K.T $\qquad \upsilon = \sqrt{ZY}$

$$ZI = \sqrt{ZY}\,A\,e^{\sqrt{ZY}x} - \sqrt{ZY}\,B\,e^{-\sqrt{ZY}x}$$

$I = \sqrt{\dfrac{Y}{Z}}\,A\,e^{\sqrt{ZY}x} - \sqrt{\dfrac{Y}{Z}}\,B\,e^{-\sqrt{ZY}x} \qquad\qquad$ 6a

$|||^{by}$,

$$\frac{dI}{dx} = C\,\upsilon\,e^{\upsilon x} - D\,\upsilon\,e^{-\upsilon x}$$

$$YV = C\sqrt{ZY}\,e^{\sqrt{ZY}x} - D\sqrt{ZY}\,e^{-\sqrt{ZY}x}$$

$V = \sqrt{\dfrac{Z}{Y}}\,C\,e^{\sqrt{ZY}x} - \sqrt{\dfrac{Z}{Y}}\,D\,e^{-\sqrt{ZY}x} \qquad\qquad$ 6b

At the receiving end,

Assume x = 0, V = V$_R$, I = I$_R$

Sub in equ 5a, 5b, 6a, 6b

$V_R = A + B \qquad\qquad$ 7a

$I_R = C + D \qquad\qquad$ 7b

$I_R = \sqrt{\dfrac{Y}{Z}}\,A - \sqrt{\dfrac{Y}{Z}}\,B \qquad\qquad$ 8a

$V_R = \sqrt{\dfrac{Z}{Y}}\,C - \sqrt{\dfrac{Z}{Y}}\,D \qquad\qquad$ 8b

To find A & B, consider equ 7a & 8a

$V_R = A + B$

$$I_R = \sqrt{\frac{Y}{Z}} A - \sqrt{\frac{Y}{Z}} B$$

Multiply by $\sqrt{\frac{Z}{Y}}$ in above equ

$$I_R \sqrt{\frac{Z}{Y}} = A - B \qquad\qquad \text{I}$$

$$V_R = A + B \qquad\qquad \text{II}$$

$$I_R \sqrt{\frac{Z}{Y}} + V_R = 2A$$

$$A = \frac{I_R}{2} \sqrt{\frac{Z}{Y}} + \frac{V_R}{2}$$

$$A = \frac{V_R}{2Z_R} Zo + \frac{V_R}{2} \qquad\qquad \left[\because I_R = \frac{V_R}{Z_R}; \quad Zo = \sqrt{\frac{Z}{Y}} \right]$$

$$A = \frac{V_R}{2} \left(1 + \frac{Z_o}{Z_R} \right)$$

To find B, subtract II from I

$$I_R \sqrt{\frac{Z}{Y}} = A - B$$

$$(-) \ V_R = A + B$$

$$I_R \sqrt{\frac{Z}{Y}} - V_R = -2B$$

$$B = \frac{V_R}{2} - \frac{I_R}{2} \sqrt{\frac{Z}{Y}} \quad \Rightarrow B = \frac{V_R}{2} - \frac{V_R}{2Z_R} Z_O$$

$$B = \frac{V_R}{2} \left(1 - \frac{Z_o}{Z_R} \right)$$

$\|^{by}$,

$$C = \frac{I_R}{2} \left(1 + \frac{Z_R}{Z_O} \right)$$

$$D = \frac{I_R}{2} \left(1 - \frac{Z_R}{Z_O} \right)$$

Sub A, B, C, D in equ 5a & 5b

$$\left. \begin{array}{l} V = \dfrac{V_R}{2} \left(1 + \dfrac{Z_O}{Z_R} \right) e^{\upsilon x} + \dfrac{V_R}{2} \left(1 - \dfrac{Z_O}{Z_R} \right) e^{-\upsilon x} \\[3mm] I = \dfrac{I_R}{2} \left(1 + \dfrac{Z_R}{Z_O} \right) e^{\upsilon x} + \dfrac{I_R}{2} \left(1 - \dfrac{Z_R}{Z_O} \right) e^{-\upsilon x} \end{array} \right\}$$

To express V & I equation in terms of refection co-efficient 'K'

$$V = \frac{V_R}{2} \left(1 + \frac{Z_O}{Z_R} \right) \left[e^{\upsilon x} + \frac{\left(1 - \dfrac{Z_O}{Z_R} \right)}{\left(1 - \dfrac{Z_O}{Z_R} \right)} e^{-\upsilon x} \right]$$

$$V = \frac{V_R}{2} \left(\frac{Z_R + Z_O}{Z_R} \right) \left[e^{\upsilon x} + \frac{Z_R - Z_O}{Z_R + Z_O} e^{-\upsilon x} \right]$$

$$V = \frac{V_R}{2} \left(\frac{Z_R + Z_O}{Z_R} \right) \left[e^{\upsilon x} + K e^{-\upsilon x} \right]$$

$$\left[\text{where } K = \frac{Z_R - Z_O}{Z_R + Z_O}\right] \text{(Reflection Coefficient)}$$

$|||^{by}$,

$$I = \frac{I_R}{2}\left(1 + \frac{Z_R}{Z_O}\right)\left[e^{\upsilon x} + \frac{\left(1 - \dfrac{Z_R}{Z_O}\right)}{\left(1 - \dfrac{Z_R}{Z_O}\right)} e^{-\upsilon x}\right]$$

$$I = \frac{I_R}{2}\left(\frac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon x} + \frac{Z_O - Z_R}{Z_O + Z_R} e^{-\upsilon x}\right]$$

$$I = \frac{I_R}{2}\left(\frac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon x} - \frac{Z_R - Z_O}{Z_R + Z_O} e^{-\upsilon x}\right]$$

$$I = \frac{I_R}{2}\left(\frac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon x} - K e^{-\upsilon x}\right]$$

Thus one useful form of voltage & current equ are

$$V = \frac{V_R}{2}\left(\frac{Z_R + Z_O}{Z_R}\right)\left[e^{\upsilon x} + K e^{-\upsilon x}\right]$$

$$I = \frac{I_R}{2}\left(\frac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon x} - K e^{-\upsilon x}\right]$$

Another useful form of voltage & current equ can be derived by simplifying

$$V = \frac{V_R}{2}\left(1 + \frac{Z_O}{Z_R}\right) e^{\upsilon x} + \frac{V_R}{2}\left(1 - \frac{Z_O}{Z_R}\right) e^{-\upsilon x}$$

$$V = \frac{V_R}{2} e^{\upsilon x} \frac{V_R}{2} \frac{Z_O}{Z_R} e^{\upsilon x} + \frac{V_R}{2} e^{-\upsilon x} - \frac{V_R}{2} \frac{Z_O}{Z_R} e^{-\upsilon x}$$

$$= \frac{V_R}{2}\left(e^{\upsilon x} + e^{-\upsilon x}\right) + \frac{V_R}{2} \frac{Z_O}{Z_R}\left(e^{\upsilon x} - e^{-\upsilon x}\right)$$

$$V = V_R \cos h\upsilon x + I_R Z_O \sin h\upsilon x \qquad \left[\frac{V_R}{Z_R} = I_R\right]$$

$$V = V_R \cos h\sqrt{ZY}x + I_R Z_O \sin h\sqrt{ZY}x \qquad \left[\upsilon = \sqrt{ZY}\right]$$

$|||^{by}$,

$$I = \frac{I_R}{2}\left(1 + \frac{Z_R}{Z_O}\right) e^{\upsilon x} + \frac{I_R}{2}\left(1 - \frac{Z_R}{Z_O}\right) e^{-\upsilon x}$$

$$I = \frac{I_R}{2} e^{\upsilon x} + \frac{I_R}{2} \frac{Z_R}{Z_O} e^{\upsilon x} + \frac{I_R}{2} e^{-\upsilon x} - \frac{I_R}{2} \frac{Z_R}{Z_O} e^{-\upsilon x}$$

$$I = \frac{I_R}{2}\left(e^{\upsilon x} + e^{-\upsilon x}\right) + \frac{I_R}{2} \frac{Z_R}{Z_O}\left(e^{\upsilon x} - e^{-\upsilon x}\right)$$

$$I = I_R \cos h\upsilon x + \frac{V_R}{Z_O} \sin h\upsilon x$$

$$I = I_R \cos h\sqrt{ZY}x + \frac{V_R}{Z_O} \sin h\sqrt{ZY}x$$

Thus the voltage and current equation at any paint from the receiving end of transmission line are

$$V = V_R \cos h\sqrt{ZY}x + I_R Z_O \sin h\sqrt{ZY}x$$

$$I = I_R \cos h\sqrt{ZY}x + \frac{V_R}{Z_O} \sin h\sqrt{ZY}x$$

Voltage and current equation at a sending end of transmission line of length 'ℓ' is

$$V = V_R \cos h\sqrt{ZY}\ l + I_R\ Z_O \sin h\sqrt{ZY}\ l$$

$$I = I_R \cos h\sqrt{ZY}\ l + \frac{V_R}{Z_O} \sin h\sqrt{ZY}\ l$$

4) Prove that infinite line is equal to finite line terminated with characteristic impedance. (M/J-16), (N/D-17)

INFINITE LINE:
A line of finite length, terminated in a load equal to characteristic impedance. Appears to the sending end generator as an infinite line.

For $l \to \infty$, $\tanh \upsilon l = 1$

$$Z_S = Z_O \left[\frac{Z_R + Z_O \tanh \upsilon l}{Z_O + Z_R \tanh \upsilon l} \right] \Rightarrow Z_S = Z_O \left[\frac{Z_R + Z_O}{Z_O + Z_R} \right] \Rightarrow Z_S = Z_O$$

INPUT IMPEDANCE OF TRANSMISSION LINE:

Input impedance $Z_S = \dfrac{V_S}{I_S}$

Voltage and current eq. at the sending end of transmission line of length 'ℓ' is given by,

$$V_S = V_R \cosh \upsilon l + I_R\ Z_O \sinh \upsilon l$$

$$I_S = I_R \cosh \upsilon l + \frac{V_R}{Z_O} \sinh \upsilon l$$

$$Z_S = \frac{V_S}{I_S} = \frac{V_R \cosh \upsilon l + I_R Z_O \sinh \upsilon l}{I_R \cosh \upsilon l + \dfrac{V_R}{Z_O} \sinh \upsilon l}$$

$$= \frac{V_R \cosh \upsilon l + \dfrac{V_R}{Z_O} Z_O \sinh \upsilon l}{I_R \cosh \upsilon l + \dfrac{I_R Z_R}{Z_O} \sinh \upsilon l}$$

$$= \frac{V_R \left[\cosh \upsilon l + \dfrac{Z_O}{Z_R} \sinh \upsilon l \right]}{I_R \left[\cosh \upsilon l + \dfrac{Z_R}{Z_O} \sinh \upsilon l \right]}$$

$$= \frac{\cancel{I_R} Z_R \left[\cosh \upsilon l + \dfrac{Z_O}{Z_R} \sinh \upsilon l \right]}{\cancel{I_R} \left[\cosh \upsilon l + \dfrac{Z_R}{Z_O} \sinh \upsilon l \right]}$$

$$= \frac{Z_R \cosh \upsilon l + Z_O \sinh \upsilon l}{\dfrac{Z_O \cosh \upsilon l + Z_R \sinh \upsilon l}{Z_O}}$$

$$Z_S = Z_O \left\{ \frac{Z_R \cosh \upsilon l + Z_O \sinh \upsilon l}{Z_O \cosh \upsilon l + Z_R \sinh \upsilon l} \right\}$$

$$Z_S = Z_O \left[\frac{Z_R + Z_O \tanh \upsilon l}{Z_O + Z_R \tanh \upsilon l} \right]$$

Voltage and current eq. can also be expressed in terms of sending and as follows.

$$V_S = \frac{V_R}{2} \left(1 + \frac{Z_O}{Z_R} \right) e^{\upsilon l} + \frac{V_R}{2} \left(1 - \frac{Z_O}{Z_R} \right) e^{-\upsilon l}$$

$$I_S = \frac{I_R}{2}\left(1 + \frac{Z_R}{Z_O}\right) e^{\upsilon l} + \frac{I_R}{2}\left(1 - \frac{Z_R}{Z_O}\right) e^{-\upsilon l}$$

$$V_S = \frac{V_R}{2}\left(1 + \frac{Z_O}{Z_R}\right)\left[e^{\upsilon l} + \frac{\left(\dfrac{Z_R - Z_O}{Z_R}\right)}{\left(\dfrac{Z_R + Z_O}{Z_R}\right)} e^{-\upsilon l}\right]$$

$$V_S = \frac{V_R}{2}\left(\frac{Z_R + Z_O}{Z_R}\right)\left[e^{\upsilon l} + \left(\frac{Z_R - Z_O}{Z_R + Z_O}\right) e^{-\upsilon l}\right] \qquad\qquad 1$$

$|||^{by}$,

$$I_S = \frac{I_R}{2}\left(\frac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon l} + \frac{\left(\dfrac{Z_O - Z_R}{Z_O}\right)}{\left(\dfrac{Z_R + Z_O}{Z_O}\right)} e^{-\upsilon l}\right]$$

$$I_S = \frac{I_R}{2}\left(\frac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon l} + \left(\frac{Z_O - Z_R}{Z_R + Z_O}\right) e^{-\upsilon l}\right]$$

$$I_S = \frac{I_R}{2}\left(\frac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon l} - \left(\frac{Z_R - Z_O}{Z_R + Z_O}\right) e^{-\upsilon l}\right] \qquad\qquad 2$$

Divide 1 by 2

$$Z_S = \frac{V_S}{I_S} = \frac{\dfrac{V_R}{2}\left(\dfrac{Z_R + Z_O}{Z_R}\right)\left[e^{\upsilon l} - \left(\dfrac{Z_R - Z_O}{Z_R + Z_O}\right) e^{-\upsilon l}\right]}{\dfrac{I_R}{2}\left(\dfrac{Z_R + Z_O}{Z_O}\right)\left[e^{\upsilon l} - \left(\dfrac{Z_R - Z_O}{Z_R + Z_O}\right) e^{-\upsilon l}\right]}$$

$$Z_S = \frac{\cancel{V_R} \cancel{Z_R}\left(\dfrac{1}{\cancel{Z_R}}\right)\left[e^{\upsilon l} - \left(\dfrac{Z_R - Z_O}{Z_R + Z_O}\right) e^{-\upsilon l}\right]}{\cancel{V_R}\left(\dfrac{1}{Z_O}\right)\left[e^{\upsilon l} - \left(\dfrac{Z_R - Z_O}{Z_R + Z_O}\right) e^{-\upsilon l}\right]}$$

$$Z_S = Z_O\left[\frac{e^{\upsilon l} + K e^{-\upsilon l}}{e^{\upsilon l} - K e^{-\upsilon l}}\right]$$

where $K = \dfrac{Z_R - Z_O}{Z_R + Z_O}$

$\downarrow$

(REFLECTION CO-EFFICIENT)

5) Derive the secondary constants using R, L, G, C and find the velocity of propagation. (N/D-18), (A/M-11).

Expression for α, β in Terms of Primary Constant:

Propagation Constant $\qquad \vartheta = \alpha + j\beta$

α = Attenuation Factor

β = Phase Constant

$$\sqrt{(R + jwL)(G + jwC)} = \alpha + j\beta$$

Squaring on both sides

$$(R + jwL)(G + jwC) = (\alpha + j\beta)^2$$

$$RG - w^2LC + jwLG + jwRC = \alpha^2 - \beta^2 + j2\alpha\beta$$

$$\left(RG - w^2LC\right) + jw\left(LG + RC\right) = \left(\alpha^2 - \beta^2\right) + j(2\alpha\beta)$$

Equate real and imaginary terms

$$\left(RG - w^2LC\right) = \left(\alpha^2 - \beta^2\right) \qquad\qquad 1$$

$$w\left(LG + RC\right) = 2\alpha\beta \qquad\qquad 2$$

Square 2 on both sides

$$w^2\left(LG + RC\right)^2 = 4\alpha^2\beta^2 \qquad\qquad 3$$

From 1, $\qquad \alpha^2 = \beta^2 + RG - w^2LC \qquad\qquad 4$

Sub 4 in 3

$$w^2\left(LG + RC\right)^2 = 4\beta^2\left(\beta^2 + RG - w^2LC\right)$$

$$w^2\left(LG + RC\right)^2 = 4\beta^4 + 4\beta^2RG - 4\beta^2w^2LC$$

$$\frac{w^2}{4}\left(LG + RC\right)^2 = \beta^4 + \beta^2\left(RG - w^2LC\right)$$

$$\left(\beta^2\right)^2 + \beta^2\left(RG - w^2LC\right) - \frac{w^2}{4}\left(LG + RC\right)^2 = 0$$

$$\beta^2 = \frac{-\left(RG - w^2LC\right) \pm \sqrt{\left(RG - w^2LC\right) + w^2\left(LG + RC\right)^2}}{2}$$

$$\beta^2 = \frac{\left(w^2LC - RG\right) + \sqrt{\left(RG - w^2LC\right) + w^2\left(LG + RC\right)^2}}{2}$$

W.K.T $\qquad \alpha^2 = \beta^2 + RG - w^2LC$

$$\alpha^2 = \frac{\left(w^2LC - RG\right) + \sqrt{\left(RG - w^2LC\right) + w^2\left(LG + RC\right)^2} + \left(RG - w^2LC\right)}{2}$$

$$\alpha^2 = \frac{\left(w^2LC - RG\right) + \sqrt{\left(RG - w^2LC\right) + w^2\left(LG + RC\right)^2} + 2RG - 2w^2LC}{2}$$

$$\alpha^2 = \frac{\left(RG - w^2LC\right) + \sqrt{\left(RG - w^2LC\right) + w^2\left(LG + RC\right)^2}}{2}$$

6) Discuss the types of waveform distortion introduced by a transmission line. (N/D-15), (N/D-13), (M/J-13), (M/J-14), (N/D-10), (N/D-08).

WAVEFORM DISTORTION:

When a signal is transmitted over the line, it will not have all frequency with equal attenuation and equal time delay.

∴ the received waveform will not be identical with the input waveform.

This variation is known as distortion

2 types of distortion

1. Frequency distortion
2. Phase distortion

FREQUENCY DISTORTION:

Attenuation constant α is given by,

$$\alpha = \sqrt{\frac{\left(RG - w^2LC\right) + \sqrt{\left(RG - w^2LC\right)^2 + w^2\left(LG + RC\right)^2}}{2}}$$

α is a function of frequency

A complex applied voltage containing many frequency will not have equal attenuation.

∴ received waveform will not be identical with the input waveform.

This is known as frequency distortion.

It can be reduced using equalizers.

Equalizers are network whose frequency & phase characteristics are adjusted to be inverse to that of line resulting in uniform frequency response over the desired band.

PHASE DISTORTION:

Phase constant k is,

$$\beta = \sqrt{\frac{\left(w^2LC - RG\right) + \sqrt{\left(RG - w^2LC\right)^2 + w^2\left(LG + RC\right)^2} + \left(RG - w^2LC\right)}{2}}$$

For an applied voltage, received waveform will not be identical with the input waveform, since some components will be delayed more than others.

This is known as delay or phase distortion.

Velocity of propagation $\vartheta = \dfrac{w}{\beta}$

If β varies, ϑ varies, results in delay distortion.

It can be overcome by the use of coaxial cable.

7) Derive the condition for distortion less operation of transmission line. (N/D-15), (N/D-16), (M/J-14), (N/D-10), (N/D-12), (M/J-13), (N/D-13).

DISTORTIONLESS LINE:

If a line is to have neither frequency nor delay distortion, then α and velocity of propagation ϑ should not be a function of frequency.

CONDITION FOR DISTORTIONLESS LINE:

W.K.T

$$\beta = \sqrt{\frac{\left(w^2LC - RG\right) + \sqrt{\left(RG - w^2LC\right)^2 + w^2\left(LG + RC\right)^2}}{2}} \qquad 1$$

To bring β, a direct function of frequency, the term under the second radical be reduced to $\left(RG + w^2LC\right)^2$

$$\left(RG - w^2LC\right)^2 + w^2\left(LG + RC\right)^2 = \left(RG + w^2LC\right)^2 \qquad 2$$

$$R^2\cancel{G^2} + w^4\cancel{L^2}C^2 - 2w^2\cancel{RG}\,LC + w^2L^2G^2 + w^2R^2C^2 + 2w^2\cancel{LG}\,RC = R^2\cancel{G^2} + w^4\cancel{L^2}C^2 + 2w^2RGLC$$

$$w^2L^2G^2 + w^2R^2C^2 - 2w^2RGLC = 0$$

$$L^2G^2 + R^2C^2 - 2\,RGLC = 0$$

$$\left(LG - RC\right)^2 = 0$$

$$LG = RC \qquad 3$$

∴ the condition that makes β, a direct function of frequency is LG = RC

Sub eq. 2 in 1

$$\beta = \sqrt{\frac{\left(w^2LC - RG\right) + \sqrt{\left(RG + w^2LC\right)^2}}{2}}$$

$$\beta = \sqrt{\frac{\left(w^2LC - RG\right) + \sqrt{\left(RG + w^2LC\right)}}{2}} = \sqrt{\frac{2w^2LC}{2}}$$

$$\beta = w\sqrt{LC} \qquad 4$$

Velocity of propagation $\quad \vartheta = \dfrac{w}{\beta}$

$$\vartheta = \dfrac{w}{w\sqrt{LC}}$$

$$\vartheta = \dfrac{1}{\sqrt{LC}} \qquad\qquad 5$$

Velocity of propagation (ϑ) is the same for all frequency. Thus eliminating delay distortion.
To make α independent of frequency,

$$\alpha = \sqrt{\dfrac{\left(RG - w^2LC\right)^2 + \sqrt{\left(RG - w^2LC\right)^2 + w^2\left(LG + RC\right)^2}}{2}}$$

The term under the second radical forced to be equal to $\left(RG + w^2LC\right)^2$

$$\alpha = \sqrt{\dfrac{\left(RG - w^2LC\right) + \sqrt{\left(RG + w^2LC\right)^2}}{2}}$$

$$\alpha = \sqrt{\dfrac{\left(RG - w^2LC\right) + \left(RG + w^2LC\right)}{2}} = \sqrt{\dfrac{2RG}{2}}$$

$\alpha = \sqrt{RG} \qquad\qquad 6$

From 3

$LG = RC$

$$\dfrac{L}{C} = \dfrac{R}{G} \qquad\qquad 7$$

To achieve eq. 7 requires large L since G is small.
If G intentionally increased, α is increased results in poor line efficiency.
Reduction in R raises the size and thus the cost of conductors, so that above condition poses practical problems.

8) Discuss in detail about indctance loading of telephone cables and derive the attenuation constant, phase constant and velocity of signal transmission (v) for uniformaly loaded cable. (M/J-07).

TELEPHONE CABLE:
In telephone cable, the wires are insulated with paper and twisted in pairs.
This results in negligible value of inductance L & conductance G.

$\therefore$ wL << R, $\qquad$ G << wC

$\qquad$ $Z = R + jwL \quad \Rightarrow \quad Z = R$

$\qquad$ $Y = G + jwC \quad \Rightarrow \quad Y = jwC$

PROPAGATION CONSTANT $\qquad \upsilon = \sqrt{ZY}$

$$\upsilon = \sqrt{jwRC}$$

$$\upsilon = \sqrt{wRC}\ \angle 45°$$

$$\upsilon = \sqrt{wRC}\left[\cos 45° + j\sin 45°\right]$$

$$\upsilon = \sqrt{wRC}\left[\dfrac{1}{\sqrt{2}} + j\dfrac{1}{\sqrt{2}}\right]$$

$$\upsilon = \alpha + j\beta = \sqrt{\dfrac{wRC}{2}} + j\sqrt{\dfrac{wRC}{2}}$$

$$\alpha = \beta = \sqrt{\dfrac{wRC}{2}} \qquad\qquad 1$$

$$\vartheta = \frac{w}{\beta} = \frac{w}{\sqrt{\dfrac{wRC}{2}}}$$

$$\vartheta = \sqrt{\frac{2w}{RC}} \qquad\qquad 2$$

From 1 & 2, α and ϑ are function of frequency, thus the telephone cable having frequency & phase distortion.

9) Explain in detail about conditions for loading and its types.

For the mine to be free of distortion, condition is

$$RC = LG$$

For a practical transmission line, $\dfrac{R}{G} \gg \dfrac{L}{C}$ hence the signal is distorted.

To satisfy the condition $\dfrac{R}{G} = \dfrac{L}{C}$, reduce $\dfrac{R}{G}$ or increase $\dfrac{L}{C}$.

To reduce $\dfrac{R}{G}$,

1. R can be increased by increasing the area of cross section A

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

   This increases the size or cast of line.

2. To $\uparrow^3 G$, it is necessary to use poor insulator.
   But when $G\uparrow^3$, leakage will increase results in poor line efficiency.

3. Hence method of reducing $\dfrac{R}{G}$ is ineffective.

   To increases $\dfrac{L}{C}$, either increase L or decrease C.

4. If C is reduced, separation between the lines will became more. The brackets which were carrying more number of wires will carry only less number of wires due to increased separation. More no. of brackets, towers, posts required. Thus the line becomes much costlier.
   Hence method of reducing C is ineffective.

To increase the inductance, lumped inductors were spaced at regular interval.
This is called loading the line.

METHODS OF LOADING:
    i)   Continuous Loading
    ii) Lumped Loading

CONTINUOUS LOADING:
In this method, to increases L, tapes of magnetic materials such as perm alloy or u-metal having high permeability are wound on each conductor.
The increase in L is given by,

$$L = \frac{\mu}{\dfrac{d}{nt}+1} \ mH$$

Where,

       μ = Permeability of surrounding material
      d = Diameter of copper conductor

n = No. of layers
t = Thickness per layer

PROPAGATION CONSTANT OF CONTINUOUS LOADED LINE:

Assume $G = 0$, $\quad wL \gg R$

$$Z = R + jwL$$

$$Y = G + jwC \quad \square \quad jwC$$

$$\upsilon = \sqrt{ZY} = \sqrt{(R + jwL)(jwC)}$$

$$= \sqrt{\sqrt{R^2 + w^2L^2}\left|\tan^{-1}\left(\frac{wL}{R}\right)\right.\ \sqrt{w^2C^2}\angle\frac{\pi}{2}}$$

$$= \sqrt{wC\sqrt{R^2 + w^2L^2}\ \left|\frac{\pi}{2} - \tan^{-1}\left(\frac{R}{wL}\right)\right.\ \angle\frac{\pi}{2}}$$

Note: $= \tan^{-1}\left(\frac{wL}{R}\right) = \frac{\pi}{2} - \tan^{-1}\left(\frac{R}{wL}\right)$

$$\upsilon = \sqrt{wC\sqrt{R^2 + w^2L^2}\ \left|\pi - \tan^{-1}\left(\frac{R}{wL}\right)\right.}$$

$$\upsilon = \sqrt{wC\sqrt{R^2 + w^2L^2}}\ \angle\frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)$$

$$\upsilon = \sqrt{wC\ wL\sqrt{\frac{R^2}{w^2L^2} + 1}}\ \angle\frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)$$

$$\upsilon = \sqrt{w^2LC}\ \angle\frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)$$

$$\upsilon = w\sqrt{LC}\ \angle\frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)$$

Let $\theta = \frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)$ $\quad\begin{bmatrix}\because \text{ after loading } wL \gg R, \\ \therefore \ \dfrac{R^2}{w^2L^2} \text{ neglected}\end{bmatrix}$

$$\cos\theta = \cos\left(\frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)\right) = \sin\left(\frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)\right)$$

Since $wL \gg R$, $\tan\theta\ \square\ \theta$, $\tan^{-1}(\theta)\ \square\ \theta$, $\sin\theta\ \square\ \theta$

$$\therefore \cos\theta = \sin\left(\frac{1}{2}\ \frac{R}{wL}\right) \Rightarrow \cos\theta = \frac{R}{2wL}$$

$\|\|^{by}$

$$\sin\theta = \sin\left(\frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left(\frac{R}{wL}\right)\right)\ \square\ \sin\left(\frac{\pi}{2}\right)\ \square\ 1$$

$$\upsilon = w\sqrt{LC}\ \angle\theta = w\sqrt{LC}\ \angle e^{j\theta} = w\sqrt{LC}\left[\cos\theta + j\sin\theta\right]$$

$$\upsilon = w\sqrt{LC}\ \left[\frac{R}{2wL} + j(1)\right]$$

$$\upsilon = \frac{R}{2}\sqrt{\frac{C}{L}}\ + jw\sqrt{LC}$$

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}}\ ;\ \beta = w\sqrt{LC};\ \ \vartheta = \frac{w}{\beta} = \frac{\cancel{w}}{\cancel{w}\sqrt{LC}} \Rightarrow \vartheta = \frac{1}{\sqrt{LC}}$$

$$Z_O = \sqrt{\frac{Z}{Y}}\ = \sqrt{\frac{R + jwL}{G + jwC}}\ = \sqrt{\frac{jwL}{jwC}} = \sqrt{\frac{L}{C}} \Rightarrow Z_O = \sqrt{\frac{L}{C}}$$

ADVANTAGES OF CONTINUOUS LOADING:

1. Attenuation α is independent of frequency.
2. α is decreased by increasing L, provided R is not disturbed.
3. Increase in L up to 100mH/unit length possible

DISADVANTAGES OF CONTINUOUS LOADING:

1. Very expensive
2. Achieve only a small increases in L per unit length.
3. Existing lines can't be modified by this method. Only replacement is possible.

Discuss in detail about Lumped loading. (N/D-17)

LUMPED LOADING:

1. In this type of loading, the inductors are introduced in lumps at uniform distance along the line.
2. Lumped inductors are in the form of coils called loading coils.
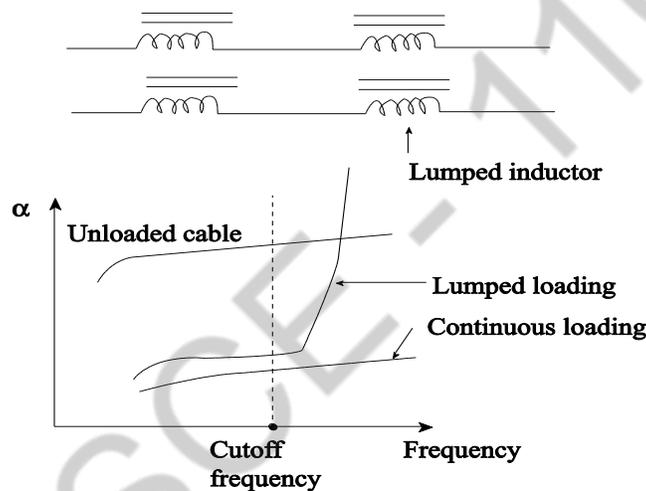
ATTENUATION FREQUENCY CHARACTERISTICS

From the figure, attenuation is independent of frequency for continuous loading, while for lumped loading α increases rapidly after the cut off frequency.

ADVANTAGE OF LUMPED LOADING:

1. Achieves a large increases in L
2. Cost involved is small
3. Existing lines can be modified.

DISADVANTAGES OF LUMPED LOADING:

1. For a cable, Z_2 is capacitive. The combination of inductance & capacitance forms a LPF.
2. For frequencies below cut off, α is reduced. Above cut off, attenuation rises as a result of filter action.

10) Derive the Campbell's equation. (N/D-10),(N/D-17).

CAMPBELL'S EQUATION:

Analysis of lumped loading can be obtained considering a symmetrical section of T line from the centre of one loading coil to the centre of next.

Let Z_C = LOADING COIL IMPEDANCE
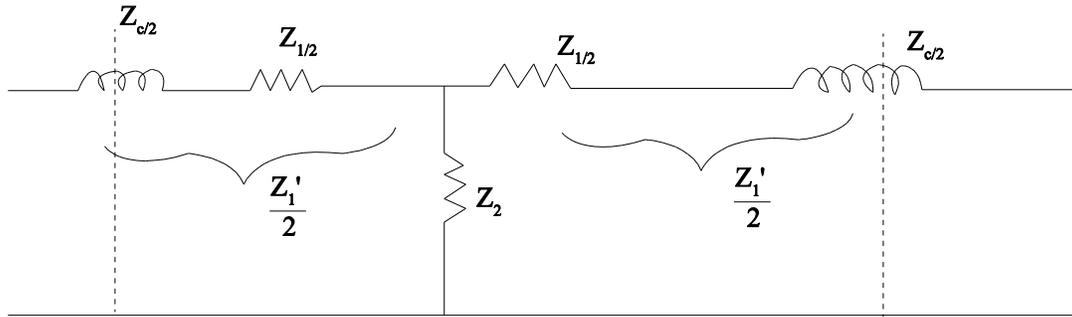
Before loading, length of the line is 'ι'

For a symmetrical T network,

$$\sinh \upsilon l = \frac{Z_O}{Z_2} \qquad 1$$

$$\cosh \upsilon l = 1 + \frac{Z_1}{2Z_2} \qquad 2$$

Series arm impedance $\dfrac{Z_1'}{2} = \dfrac{Z_C}{2} + \dfrac{Z_1}{2}$

Shunt are impedance $= Z_2$

From 1, $\quad Z_2 = \dfrac{Z_O}{\sinh \upsilon l}$

$$[\cosh \upsilon l - 1] Z_2 = \frac{Z_1}{2}$$

$$\frac{Z_1'}{2} = \frac{Z_C}{2} + [\cosh \upsilon l - 1] Z_2$$

$$\frac{Z_1'}{2} = \frac{Z_C}{2} + [\cosh \upsilon l - 1] \frac{Z_O}{\sinh \upsilon l}$$

After loading

$$\cosh \upsilon' l = 1 + \frac{Z_1'}{2Z_2}$$

$$= 1 + \frac{\dfrac{Z_1'}{2}}{Z_2}$$

$$= \frac{1 + \dfrac{Z_C}{2} + [\cosh \upsilon l - 1]\dfrac{Z_O}{\sinh \upsilon l}}{\dfrac{Z_O}{\sinh \upsilon l}}$$

$$\cosh \upsilon' l = 1 + \frac{\sinh \upsilon l}{Z_O}\left[\frac{Z_C}{2} + [\cosh \upsilon l - 1]\frac{Z_O}{\sinh \upsilon l} \right]$$

$$\cosh \upsilon' l = 1 + \frac{Z_C \sinh \upsilon l}{2Z_O} + \cosh \upsilon l - 1$$

$$\cosh \upsilon' l = \cosh \upsilon l + \frac{Z_C}{2Z_O} \sinh \upsilon l$$

This expression is known as Campbell's equation.

Using this, propagation constant of a loaded line can be computed.

11) Explain the significance of reflection coefficient and insertion loss. (A/M-11).

REFLECTION CO-EFFICIENT:

$$K = \frac{\text{REFLECTED VOLTAGE AT LOAD}}{\text{INCIDENT VOLTAGE AT LOAD}}$$

$$V = \frac{V_R}{2}\left[\left(1+\frac{Z_O}{Z_R}\right)e^{\upsilon l}+\left(1-\frac{Z_O}{Z_R}\right)e^{-\upsilon l}\right]$$

Where, $\quad \dfrac{V_R}{2}\left(1+\dfrac{Z_O}{Z_R}\right)\to$ Incident voltage

$$\frac{V_R}{2}\left(1-\frac{Z_O}{Z_R}\right)\to \text{Reflected voltage}$$

$$K = \frac{\dfrac{V_R}{2}\left(1-\dfrac{Z_O}{Z_R}\right)}{\dfrac{V_R}{2}\left(1+\dfrac{Z_O}{Z_R}\right)}=\frac{Z_R-Z_O}{Z_R+Z_O}\quad\Rightarrow\quad K=\frac{Z_R-Z_O}{Z_R+Z_O}$$

CASE i: ($Z_R = Z_O$)

$\quad$ K = 0, no reflection

CASE ii: LINE IS SHORT CIRCUITED ($Z_R = 0$)

$$K = -\frac{Z_O}{Z_O}=-1=1\angle 180°$$

REFLECTION IS MAXIMUM

CASE iii: LINE IS OPEN CIRCUITED ($Z_R = \infty$)

$$K = \frac{Z_R\left(1-\dfrac{Z_O}{Z_R}\right)}{Z_R\left(1+\dfrac{Z_O}{Z_R}\right)}=1\quad\Rightarrow\quad K=1\angle 0°$$

$\quad$ REFLECTION IS MAXIMUM

TRANSFER IMPEDANCE IN TERMS OF K:

$$Z_T = \frac{\text{SENDING END VOLTAGE}\left(V_S\right)}{\text{RECEIVING END CURRENT}\left(I_R\right)}$$

$$V_S = \frac{V_R}{2}\left[\left(1+\frac{Z_O}{Z_R}\right)e^{\upsilon l}+\left(1-\frac{Z_O}{Z_R}\right)e^{-\upsilon l}\right]$$

$$V_S = \frac{V_R}{2}\left(\frac{Z_R+Z_O}{Z_R}\right)\left[e^{\upsilon l}+\frac{\left(\dfrac{Z_R-Z_O}{Z_R}\right)}{\left(\dfrac{Z_R+Z_O}{Z_R}\right)}e^{-\upsilon l}\right]$$

$$V_S = \frac{V_R}{2}\left(\frac{Z_R+Z_O}{Z_R}\right)\left[e^{\upsilon l}+\left(\frac{Z_R-Z_O}{Z_R+Z_O}\right)e^{-\upsilon l}\right]$$

$$V_S = \frac{I_R \cancel{Z_R}}{2}\left(\frac{Z_R+Z_O}{\cancel{Z_R}}\right)\left[e^{\upsilon l}+\left(\frac{Z_R-Z_O}{Z_R+Z_O}\right)e^{-\upsilon l}\right]$$

$$\frac{V_S}{I_R} = \frac{1}{2}\left[\left(Z_R+Z_O\right)e^{\upsilon l}+\left(Z_R-Z_O\right)e^{-\upsilon l}\right]$$

$$\frac{V_S}{I_R} = \frac{1}{2}\left[Z_R\left(e^{\upsilon l}+e^{-\upsilon l}\right)+Z_O\left(e^{\upsilon l}-e^{-\upsilon l}\right)\right]$$

$$Z_T = \frac{V_S}{I_R} = Z_R \cosh \upsilon l + Z_O \sinh \upsilon l$$

$$Z_T = Z_R \cosh \upsilon l + Z_O \sinh \upsilon l$$

12) Derive the expressions for input impedance of open and short circuited lines. (N/D-16), (N/D-18), (M/J-12).

OPEN & SHORT CIRCUITED LINES:

INPUT IMPEDANCE $Z_S = Z_O \left[\dfrac{Z_R \cosh \upsilon l + Z_O \sinh \upsilon l}{Z_O \cosh \upsilon l + Z_R \sinh \upsilon l} \right]$

i) For short circuit line, $Z_R = 0$

$$Z_{SC} = Z_O \left[\frac{Z_O \sinh \upsilon l}{Z_O \cosh \upsilon l} \right] = Z_O \tanh \upsilon l$$

$$Z_{SC} = Z_O \tanh \upsilon l$$

ii) For open circuit line, $Z_R = \infty$

$$Z_{OC} = \frac{Z_O \cdot Z_R}{Z_R} \left[\frac{\cosh \upsilon l + \dfrac{Z_O}{Z_R} \sinh \upsilon l}{\dfrac{Z_O}{Z_R} \cosh \upsilon l + \sinh \upsilon l} \right]$$

$$Z_{OC} = Z_O \coth \upsilon l$$

$$Z_{OC} \cdot Z_{SC} = (Z_O \coth \upsilon l)(Z_O \tanh \upsilon l) = Z_O^2$$

$$Z_O = \sqrt{Z_{OC} \cdot Z_{SC}}$$

$$\frac{Z_{SC}}{Z_{OC}} = \frac{Z_O \tanh \upsilon l}{Z_O \coth \upsilon l}$$

$$(\tanh \upsilon l)^2 = \frac{Z_{SC}}{Z_{OC}}$$

$$\upsilon l = \tanh^{-1} \sqrt{\frac{Z_{SC}}{Z_{OC}}}$$

13) If Z=R+jwL and Y=G+jwC, show that the line parameters values fix the velocity of propagation for an ideal line. (N/D-11).

1. The distance travelled by the wave corresponding to a phase shift of 2π radian is called wavelength and represented by λ meters.

$$\lambda = \frac{2\pi}{\beta}$$

2. It is defined as the velocity with which for signal of single frequency propagates along the line at the particular frequency.

$$V = \lambda f = \frac{2\pi}{\beta} f = \frac{w}{\beta} \text{ metres}$$

α and β in terms of primary constants
   R, L, C, G

The propagation constant γ and Z_O are the secondary constants of the line. The γ in terms of complex form it is represented as

$$\gamma = \alpha + j\beta$$

$$\sqrt{(R+jwL)(G+jwC)} = \alpha + j\beta$$

Squaring on both sides

$$\alpha^2 + 2j\beta - \beta^2 = (R+jwL)(G+jwC)$$
$$= RG - w^2LC + jw(LG+RC)$$

Equating real and imaginary parts

$$\alpha^2 - \beta^2 = RG - w^2LC \Rightarrow \alpha^2 = \beta^2 + RG - w^2LC \ldots\ldots\ldots\ldots\ldots 1$$
$$2\alpha\beta = w(LG+RC)$$

Squaring on both sides

$$4\alpha^2\beta^2 = w^2(LG+RC)^2$$
$$\alpha^2\beta^2 = \frac{w^2}{4}(LG+RC)^2 \ldots\ldots\ldots\ldots\ldots 2$$

Sub α^2 value in 2

$$(\beta^2 + RG - w^2LG)\beta^2 = \frac{w^2}{4}(LG+RC)^2$$
$$\beta^4 + RG\beta^2 - w^2\beta^2 LG - \frac{w^2}{4}(LG+RC)^2 = 0$$
$$\beta^4 + \beta^2(RG - w^2LC) - \frac{w^2}{4}(LG+RC)^2 = 0$$

This is a Quadratic eq in β^2

$$= \frac{-(RG-w^2LC) \pm \sqrt{(RG-w^2LC)^2 + \cancel{4} \times \frac{w^2}{\cancel{4}}(LG+RC)^2}}{2}$$

$$\beta^2 = \frac{(w^2LC-RG) \pm \sqrt{(RG-w^2LC)^2 + w^2(LG+RC)^2}}{2}$$

$$\beta^2 = \frac{(w^2LC-RG) + \sqrt{(RG-w^2LC)^2 + w^2(LG+RC)^2}}{2}$$

W.K.T

$$\alpha^2 = \beta^2 + RG - w^2LC$$

$$\alpha^2 = \frac{(w^2LC-RG) + \sqrt{(RG-w^2LC)^2 + w^2(LG+RC)^2}}{2}$$

$$*RG - w^2LC$$

$$= \frac{(w^2LC-RG) + \sqrt{(RG-w^2LC)^2 + w^2(LG+RC)^2} + 2RG - 2w^2LC}{2}$$

$$\alpha^2 = \frac{RG - w^2LC + \sqrt{(RG-w^2LC)^2 + w^2(LG+RC)^2}}{2}$$

14) Derive about the Insertion loss.

Insertion Loss:
The insertion loss of a line or n/w is defined as the no. of nepers or decibels by which the current in the load is change by the insertion. The insertion of a 4 terminal n/w or a line b/w a generator and a load may improve or reduced the impedance match b/w the source and the load.

$$I_S = \frac{I_R}{2Z_O}(Z_R + Z_O)(e^{\gamma l} - Ke^{-\gamma l}) \ldots\ldots\ldots\ldots 1$$

From figure

$$I_S = \frac{E}{Z_g + Z_s}$$
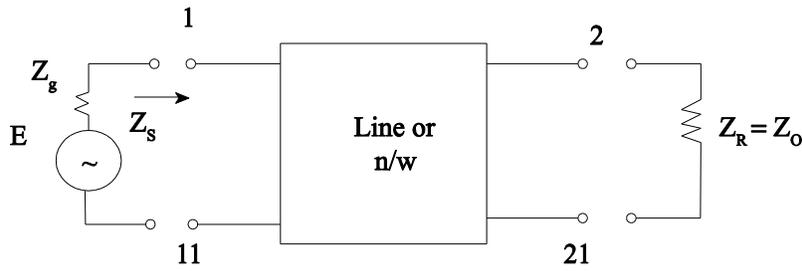
The I/P impedance $Z_S = Z_O \left(\dfrac{e^{\gamma l} + Ke^{-\gamma l}}{e^{\gamma l} - Ke^{-\gamma l}} \right)$

$$I_S = \frac{E}{Z_g + Z_O \left(\dfrac{e^{\gamma l} + Ke^{-\gamma l}}{e^{\gamma l} - Ke^{-\gamma l}} \right)}$$

$$= \frac{E}{\dfrac{Z_g \left(e^{\gamma l} - Ke^{-\gamma l} \right) + Z_O \left(e^{\gamma l} + Ke^{-\gamma l} \right)}{e^{\gamma l} - Ke^{-\gamma l}}}$$

$$= \frac{E \left(e^{\gamma l} - Ke^{-\gamma l} \right)}{Z_g \left(e^{\gamma l} - Ke^{-\gamma l} \right) + Z_O \left(e^{\gamma l} + Ke^{-\gamma l} \right)}$$

From equation

$$I_R = \frac{2Z_O I_S}{\left(Z_R + Z_O \right) e^{\gamma l} - Ke^{-\gamma l}}$$

$$I_R = \frac{2Z_O \ E \left(e^{\gamma l} - \cancel{K} e^{-\gamma l} \right)}{\left(Z_R + Z_O \right) e^{\gamma l} - \cancel{K} e^{-\gamma l} Z_g \left(e^{\gamma l} - Ke^{-\gamma l} \right) + Z_O \left(e^{\gamma l} + Ke^{-\gamma l} \right)}$$

Let I'_R is the current flowing in the load
To find

$$I'_R = \frac{E}{Z_R + Z_g}$$

$$I'_R \Big/ I_R = \frac{\dfrac{E}{Z_g + Z_R}}{\dfrac{2Z_O E}{\left(Z_R + Z_O \right) Z_g \left(e^{\gamma l} - Ke^{-\gamma l} \right) + Z_O \left(e^{\gamma l} + Ke^{-\gamma l} \right)}}$$

$$= \frac{\dfrac{\cancel{E}}{Z_g + Z_R}}{\dfrac{2Z_O \cancel{E}}{\left(Z_R + Z_O \right) Z_g \left(e^{\gamma l} - \dfrac{Z_R - Z_O}{Z_R + Z_O} e^{-\gamma l} \right) + Z_O \left(e^{\gamma l} + \dfrac{Z_R - Z_O}{Z_R + Z_O} e^{-\gamma l} \right)}}$$

$$= \frac{\dfrac{1}{Z_R + Z_g}}{\dfrac{2Z_O}{\left(Z_R + Z_O \right) \left[Z_g \left(\dfrac{Z_R + Z_O e^{\gamma l} - Z_R - Z_O e^{-\gamma l}}{Z_R + Z_O} \right) + Z_O \left(\dfrac{Z_R + Z_O e^{\gamma l} + Z_R - Z_O e^{-\gamma l}}{Z_R + Z_O} \right) \right]}}$$

$$= \cfrac{\cfrac{1}{Z_g + Z_R}}{\cfrac{Z_R + Z_O}{Z_R + Z_O}\left[Z_g\left(Z_R + Z_O e^{\gamma l} - Z_R - Z_O e^{-\gamma l}\right) + Z_O\left(Z_R + Z_O e^{\gamma l} + Z_R - Z_O e^{-\gamma l}\right)\right]}$$

$$= \frac{Z_g Z_R + \cancel{Z_g} Z_O e^{-\gamma l} - \cancel{Z_g} Z_R - Z_O \cancel{Z_O} e^{\gamma l} + Z_O Z_R - Z_O \cancel{Z_O} e^{\gamma l}}{2Z_O\left(Z_g + Z_R\right)}$$

$$= \frac{\left(Z_g + Z_O\right)\left(Z_O + Z_R\right)e^{\gamma l}\left(\alpha + j\beta\right) + \left(Z_O - Z_g\right)\left(Z_R - Z_O\right)e^{-\gamma l}\left(-\alpha + j\beta\right)}{2Z_O\left(Z_g + Z_R\right)}$$

Neglect $e^{-\alpha l} = 0$

$$= \frac{\left(Z_g + Z_O\right)\left(Z_O + Z_R\right)e^{\alpha l}e^{j\beta l}}{2Z_O\left(Z_g + Z_R\right)}$$

The physical significance of the eq can be studied by multiplying the numerator and denominator by $2\sqrt{Z_g Z_R}$

$$\frac{I_R{}'}{I_R} = \frac{\left(Z_R + Z_O\right)2\sqrt{Z_g Z_R}\left(Z_g + Z_O\right)e^{\alpha l}e^{j\beta l}}{4\sqrt{Z_g Z_R}\; Z_O\left(Z_R + Z_g\right)}$$

Talking the magnitude

$$= \frac{2\sqrt{Z_g Z_R}\left(Z_R + Z_O\right)\left(Z_g + Z_O\right)e^{\alpha l}}{4\sqrt{Z_g Z_R Z_O{}^2}\left(Z_R + Z_g\right)}$$

Rearranging the terms

$$\left|\frac{I_R{}'}{I_R}\right| = \frac{\left|Z_O + Z_g\right|}{2\sqrt{Z_g Z_O}}\;\frac{\left|Z_R + Z_O\right|}{2\sqrt{Z_R Z_O}}\;\frac{2\sqrt{Z_g Z_R}}{\left(Z_g + Z_R\right)}$$

$$K_S = \frac{2\sqrt{Z_o Z_g}}{\left|Z_g + Z_o\right|}, \quad K_R = \frac{2\sqrt{Z_R Z_o}}{\left|Z_R + Z_o\right|}$$

$$K_{SR} = \frac{2\sqrt{Z_g Z_R}}{\left|Z_g + Z_R\right|}$$

The current ratio

$$\left|\frac{I_R{}'}{I_R}\right| = \frac{K_{SR}e^{\alpha l}}{K_S \cdot K_R} = \frac{\dfrac{1}{K_S}\cdot\dfrac{1}{K_R}e^{\alpha l}}{\dfrac{1}{K_{SR}}}$$

In nepers

$$\ln\left|\frac{I_R{}'}{I_R}\right| = \ln\left(\frac{1}{K_S}\right) + \ln\left(\frac{1}{K_R}\right) - \ln\left(\frac{1}{K_{SR}}\right) + \alpha l$$

In decibels

$$= 20\log\left(\log\frac{1}{K_S} + \log\frac{1}{K_R} - \log\frac{1}{K_{SR}} + 0.43\alpha l\right)$$

Where K_R – reflection factor at load

K_S – reflection factor at source

K_{SR} – reflection factor for direct connection with generation and load

$e^{\alpha \ell}$ – loss in the line

UNIT II-HIGH FREQUENCY TRANSMISSION LINES
PART-A

1) State the assumptions for the analysis of the performance of the radio frequency line. (A/M-18)

1. Due to the skin effect, the currents are assumed to flow on the surface of the conductor. The internal inductance is zero.
2. The resistance R increases with square root of 'f' while inductance L increases with f. Hence $\omega L \gg R$.
3. The leakage conductance G is zero.

2) State the expressions for inductance L of a open wire line and coaxial line. (N/D-14)

For open wire line,
$L = 9.21 \times 10^{-7}(\mu/\mu r + 4 \ln b/a) = 10^{-7}(\mu r + 9.21 \log b/a)$ H/m

For coaxial line,
$L = 4.60 \times 10^{-7}[\log b/a]$ H/m

3) State the expressions for the capacitance C of a open wire line and coaxial line. (N/D-14)

For open wire line,
$C = (12.07)/(\ln b/a)\mu$ μf/m

For coaxial line,
$C = (55.5\varepsilon_r)/(\ln b/a)\mu$ μf/m

4) What is dissipation less line?

A line for which the effect of resistance R is completely neglected is called dissipation less line.

5) What is the nature and value of Z_0, α, β and ν for the dissipation less line? (N/D-11), (M/J-14) (N/D-17)

For the dissipation less line,
i) The Z_0 is purely resistive and given by, $Z_0 = R_0 = (L/C)^{\frac{1}{2}}$
ii) Attenuation constant, $\alpha = 0$
iii) Phase constant, $\beta = w(LC)^{\frac{1}{2}}$ radians/m
iv) Velocity of propagation, $\nu = 1/(LC)^{\frac{1}{2}}$ m/s

6) What are nodes and antinodes on a line? (N/D-17)

The points along the line where magnitude of voltage or current is zero are called nodes.

The points along the lines where magnitude of voltage or current first maximum are called antinodes or loops.
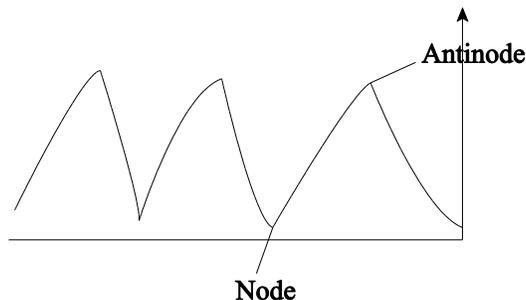
7) What is standing wave ratio? (M/J-13),

The ratio of the maximum to minimum magnitudes of voltage or current on a line having standing waves called standing waves ratio and it is denoted by S.

$$\text{VSWR} = S = \left| \frac{V_{max}}{V_{min}} \right| \quad \text{or} \quad S = \left| \frac{I_{max}}{I_{min}} \right|$$

8) What is the range of values of standing wave ratio and Reflection coefficient? (N/D-10)

The range of values of standing wave ratio is theoretically 1 to infinity.

      (i) $1 < \text{SWR} < \infty$

           Minimum value of SWR is 1.

           Maximum value of SWR is ∞.

      (ii) $0 < K < 1$

           Minimum value of reflection coefficient is 0.

           Maximum value of reflection coefficient is 1.

9) State the relation between standing wave ratio and reflection coefficient. (A/M-11), (N/D-12),(M/J-12)

$$S = \frac{1 + |K|}{1 - |K|} \text{ , } K = \text{Reflection coefficient.}$$

$$K = \frac{S - 1}{S + 1}, \text{ } S = \text{Standing wave ratio}$$

10) What are standing waves?

In a transmission line, if the load impedance is not equal to the characteristic impedance ($Z_R \neq Z_0$), the energy delivered to load is reflected back to source. The actual voltage at any point on the transmission line is the sum of the incident and reflected wave voltages at that point. The resultant total voltage wave appears to stand still on the line, oscillating in magnitude with time, having fixed positions of maxima and fixed minima. Such waves are called standing waves.

11) How will you make standing wave measurements on coaxial lines? (A/M-15)

For coaxial lines it is necessary to use a length of line in which a longitudinal slot, one half wavelength or more long has been cut. A wire probe is inserted into the air dielectric of the line as a pickup device, a vacuum tube voltmeter or other detector being connected between probe and sheath as an indicator. If the meter provides linear indications, S is readily determined. If the indicator is non linear, corrections must be applied to the readings obtained.

12) Give the input impedance of dissipation less line.

The input impedance of dissipation less line is given by,

$$Z_S = R_O \left(\frac{1 + |K| \underline{|\phi - 2\beta l}}{1 - |K| \underline{|\phi - 2\beta l}} \right)$$

13) Give the maximum and minimum input impedance of the dissipation less line.

Maximum input impedance, $Z_{Smax} = R_O \text{ } S$

Minimum input impedance, $Z_{Smin} = R_O / S$

14) Give the input impedance of open and short circuited lines. (N//D-10), (A/M-18)

The input impedance of open and short circuited lines are given by,

For a short circuit, $Z_{SC} = jR_O \tan \beta l = j \times S$

For a open circuit, $Z_{OC} = -jR_O C_O + \beta l = j X_S$

15) Why the point of voltage minimum is measured rather than voltage maximum?

The point of a voltage minimum is measured rather than a voltage maximum because it is usually possible to determine the exact point of minimum voltage with greater accuracy.

16) Why standing waves do exists in transmission line. (N/D-10)

When the transmission line is not terminated with its characteristic impedance ($Z_R \neq Z_0$), the energy delivered to the load is reflected back to source which give rise to the standing waves.

PART-B

1) Discuss the various parameters of open wire line and coaxial line at radio frequency. (N/D-15), (N/D-14), (N/D-18)

Due to skin effect the current is considered as following essentially on the surface of the conductor in a skin of very small depth. The internal inductance and internal flux are reduced nearly to zero, the inductance of open wire line becomes,

$$L = \frac{\mu_0}{2\pi} \ln\left(\frac{d}{a}\right) \quad H/m$$

a – radius of open wire line

d – distance b/w 2 open wire line

$\mu_0 = 4\pi \times 10^{-7} \quad H/m$

The value of capacitance of a open wire line is not affected by skin effect,

Capacitor $\quad C = \dfrac{\pi\varepsilon_0}{\ln\left(\frac{d}{a}\right)} \quad F/m$, where

$\varepsilon_0 = 8.854 \times 10^{-12}$ F/m.

If the current flows at high frequency over the surface of the conductor in a thin layer, there is an increase in resistance of the conductor. The skin depth of the surface layer of current is

$$\text{Skin depth} \ = \frac{1}{\sqrt{\pi f \mu}} \text{ metres}$$

Direct current resistance of open wire line is,

$$Rdc = \frac{K}{\pi a^2}$$

For alternating current resistance of open wire line is,

$$Rdc = \frac{K}{\pi a}$$

$$= K\sqrt{\frac{\pi f \mu \varepsilon}{\pi a}}$$

$$= \frac{K}{a}\sqrt{\frac{\mu 6}{\pi}} \ \sqrt{f} \quad ohm/m$$

This equation shows that the resistance increases with increasing frequency.

Parameters of coaxial cable

The parameters of the coaxial line are also modified by the presence of high frequency currents on the line. Because of skin effect, the current flows only on the surface of the conductor and it eliminates the flux linkages. The inductance of the capacitance of coaxial cable is,

$$L = \frac{\mu_0}{2\pi} \ln\left(\frac{b}{a}\right)$$

a – outer radius of inner conductor

b – inner radius of inner conductor

The capacitance of coaxial cable is not affected by high frequency current. The value of capacitance of coaxial cable is,

$$C = \frac{2\pi\varepsilon_0}{\ln\left(\frac{b}{a}\right)} \quad f/m \quad \varepsilon_1 = \varepsilon_0\varepsilon_r$$

Due to skin effect, the resistance of the coaxial cable is given by,

$$Rac = \frac{K}{\pi6}\left(\frac{1}{a} + \frac{1}{b}\right)$$

$$Rac = \frac{K\sqrt{\pi f\mu6}}{\pi}\left(\frac{1}{a} + \frac{1}{b}\right)$$

$$Rac = \frac{K\sqrt{\mu6}}{\pi}\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right)\frac{\Omega}{m}$$

Resistance increases with an increase in frequency.

2) Derive the line constant of a zero dissipation line. (M/J-16), (A/M-18)
The line constants for transmission line are

$$Z = R + j\omega L$$
$$Y = G + jwC$$

$$Z_o = \sqrt{\frac{R + jwL}{G + jwC}}$$

$$\gamma = \sqrt{(R + jwL)(G + jwC)}$$

For a transmission of energy at high frequency, wL>>R, we assume negligible losses or zero dissipation and G is also assumed to be zero. (R = G = 0)
The Characteristic impedance Z_0 is given by,

$$Z = j\omega L$$
$$Y = jwC$$
$$Z_o = \sqrt{\frac{L}{C}}$$

The propagation constant γ is given by,

$$\gamma = \sqrt{(R + jwL)(G + jwC)}$$
$$= \sqrt{(jwL)(jwC)}$$
$$\alpha + j\beta = jw\sqrt{LC}$$
$$\alpha = 0$$
$$\beta = w\sqrt{LC}$$

The velocity of propagation ν is given by,

$$V_P = \frac{w}{\beta} = \frac{1}{\sqrt{LC}}$$

$Z_o = R_o$ which is purely resistive

3) Discuss in detail about the voltages and currents on the dissipation less line. (N/D-17),()
The voltage at any point of distance 't' and the current from the receiving end of the transmission line is given by

$$V_S = \frac{V_R}{2Z_R}(Z_R + Z_O)(e^{\gamma l} + Ke^{-\gamma l})$$

$$I_S = \frac{I_R}{2Z_o}(Z_R + Z_O)(e^{\gamma l} - Ke^{-\gamma l})$$

$Z_o = R_o$

$\gamma = \alpha + j\beta$

$\gamma = j\beta$

$V_S = \dfrac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(e^{j\beta l} + Ke^{-j\beta l}\right)$

$I_S = \dfrac{I_R}{2Z_o}\left(Z_R + R_O\right)\left(e^{j\beta l} - Ke^{-j\beta l}\right)$

The voltage and current equation of Transmission Line can also be represented as

$V_S = V_R \cosh \gamma l + I_R Z_O \sinh \gamma l$

$I_S = I_R \cosh \gamma l + \dfrac{V_R}{Z_O} \sinh \gamma l$

$Z_o = R_o$

$\gamma = \alpha + j\beta$

$\gamma = j\beta$

$V_S = V_R \cosh j\beta l + I_R R_O \sinh j\beta l$

$I_S = I_R \cosh j\beta l + \dfrac{V_R}{R_O} \sinh j\beta l$

$V_S = V_R \cosh \beta l + jI_R R_O \sinh \beta l$

$I_S = I_R \cosh \beta l + j\dfrac{V_R}{R_O} \sinh \beta l$

$V_S = V_R \cos\left(\dfrac{2\pi}{\lambda}\right)l + jI_R R_O \sin\left(\dfrac{2\pi}{\lambda}\right)l$

$I_S = I_R \cos\left(\dfrac{2\pi}{\lambda}\right)l + j\dfrac{V_R}{R_O} \sin\left(\dfrac{2\pi}{\lambda}\right)l$

Case 1: when the line is open circuited

$I_R = 0$

$V_{SC} = V_R \cos\left(\dfrac{2\pi}{\lambda}\right)l$

$I_{OC} = j\dfrac{V_R}{R_O} \sin\left(\dfrac{2\pi}{\lambda}\right)l$ j is in phase quadrature $\underline{90°}$ phase shift

Case 2: when the line is short circuited

$V_R = 0$

$V_{SC} = j I_R R_O \sin\left(\dfrac{2\pi}{\lambda}\right)l$

$I_{SC} = I_R \cos\left(\dfrac{2\pi}{\lambda}\right)l$

From the above expression of voltage and current it is seen that the voltage and current are in phase quadrature. The voltage and current magnitude distribution for an open circuit line are in quadrature everywhere and thus no power is transmitted along the line when the line is short circuited the voltage and current are in quadrature and the waves are shifted by λ/4 from the open circuit case.
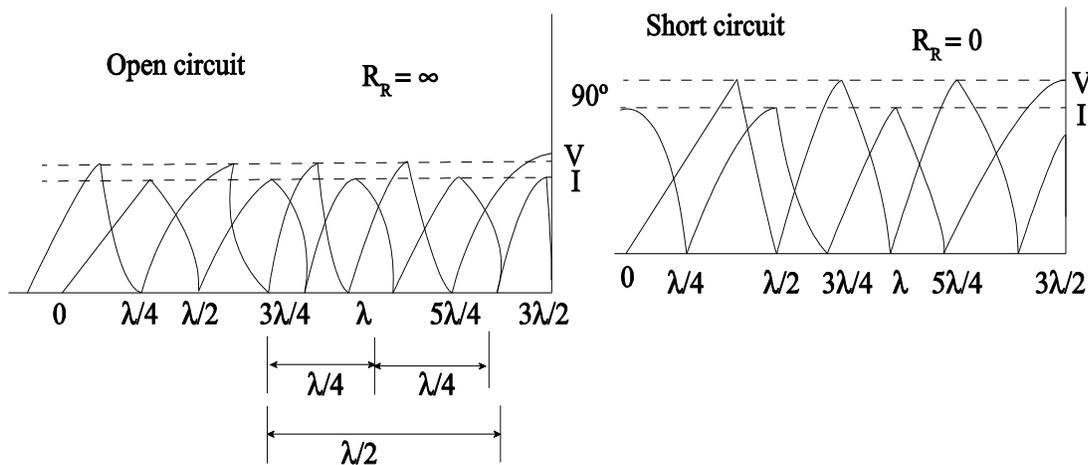
Distribution of voltage and current Under open circuit and short circuit

Case3: If line is matched $Z_R = Z_o$, the reflection coefficient and reflected wave becomes zero.

$$V_S = \frac{V_R}{2R_o}(Z_R + R_O)e^{\gamma l}$$

$$I_S = \frac{I_R}{2R_o}(Z_R + R_O)e^{\gamma l}$$

$$V_S = V_R e^{\gamma l}$$

$$I_S = I_R e^{\gamma l}$$

$$V_S = V_R e^{j\beta l}$$

$$I_S = I_R e^{j\beta l}$$
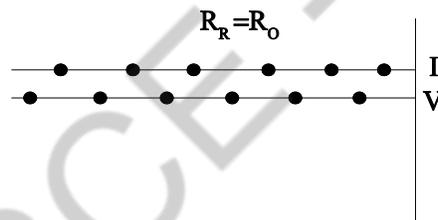
Voltage and current are in phase.

Voltage and current are same and repeated every $\lambda/2$.

At every $\lambda/4$ maximum to minimum process. From the above graph it is seen that at every $\lambda/4$ period the voltage and current characteristics changes to maximum to minimum and vice versa.

At every $\lambda/2$ the characteristics will be repeated.

Case 4: When $R_R = 3R_0$, K = 1/2, there is a finite value of voltage (or) current at all points on the line.

Since both voltage and current have values other than zero at the load, some power is being is transmitted. For resistive loads greater than R_0, voltage and current distributions resemble those of open circuited line.

5) Briefly explain on Standing waves. (N/D-16), (A/M-18)

In a transmission line, if the load impedance is not equal to the characteristic impedance ($Z_R \neq Z_0$), the energy delivered to load is reflected back to source. The actual voltage at any point on the transmission line is the sum of the incident and reflected wave voltages at that point. The resultant total voltage wave appears to stand still on the line, oscillating in magnitude with time, having fixed positions of maxima and fixed minima. Such waves are called standing waves.
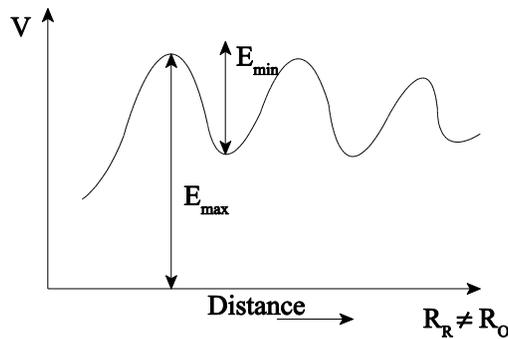
Standing waves on a dissipation less line terminated in a load not equal to R_0

When a line is terminated with a load other than R_o, then the voltage and magnitude of current and measure along the line is given to the following graphs.

If the line is either short circuited or open circuited at the receving and, we get nodes and antinodes in the voltage distributions as shown below,
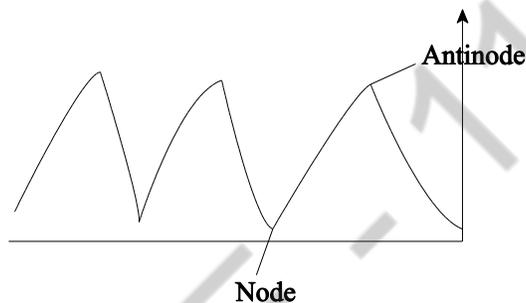
Standing waves on a line having open or short circuited terminations

i) Nodes are the points of zero voltage or current (E=I=0) in the standing wave systems.
ii) Antinodes or loops are points of maximum voltage or current.
iii) A line terminated with R_0 has no standing waves and thus no nodes or loops and is called as **smooth line.**
iv) For open circuit, the voltage nodes occur at distances $\lambda/4$, $3\lambda/4$, $5\lambda/4$ and so on from the open end of the line. Under the same conditions, the current nodes occur at a distance 0, $\lambda/2$, λ, $3\lambda/2$ and so on for open termination.
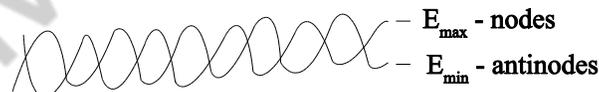
v) For short circuit, these nodal points shift by a distance of $\lambda/4$.
       Voltage nodes occur at 0, $\lambda/2$, λ and so on,
       Current nodes occur at $\lambda/4$, $3\lambda/4$, $5\lambda/4$ and so on.

From the graph it is observed that nodes are the paths of over current and antinodes are points of maximum voltage on current. A line terminated in Ro as no standing waves such a line called smooth line.

Standing Wave Ratio (SWR):
The ratio of maximum to minimum magnitudes of voltage or current on a line having standing waves called SWR denoted by 'S'.

$$VSWR = S = \left|\frac{V_{max}}{V_{min}}\right| \text{ or } S = \left|\frac{I_{max}}{I_{min}}\right| \text{ It is complicated.}$$
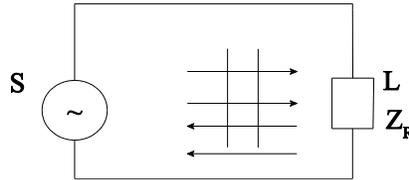
The maxima of voltage along the line occurs at point where the incident and reflected wave are in phase and add directly.

$$V = \frac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(e^{j\beta l} + K e^{-j\beta l}\right)$$

$$V_{max} = \frac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(1 + |K|\right)$$

|||^{rly}

$$V_{min} = \frac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(1 + |K|\right)$$

$$S = \left|\frac{V_{max}}{V_{min}}\right| = \frac{1 + |K|}{1 - |K|}$$

$$S = \frac{1 + |K|}{1 - |K|}$$

$$K = \frac{S - 1}{S + 1}$$ where 'S' is the voltage standing wave ratio.

It is measured by RF voltmeter
ISWR measured by RF ammeter.
The 'S' value ranges from   S = 1 to ∞
$$K = 0 \text{ to } 1$$

Relation b/w S and K

We know that $S = \left|\dfrac{V_{max}}{V_{min}}\right|$ and we know $V_{max} = V^+ + V^-$ $V_{min} = V^+ - V^-$ and the ratio $= \left|\dfrac{V^+ + V^-}{V^+ - V^-}\right|$

$$S = \left|\frac{V^+\left(1 + \frac{V^-}{V^+}\right)}{V^+\left(1 - \frac{V^-}{V^+}\right)}\right| = \frac{1 + |K|}{1 - |K|}$$

$$1 + \frac{V^-}{V^+} = 1 + K$$

Show that $K = \left|\dfrac{V_{max} - V_{min}}{V_{max} + V_{min}}\right|$

$$K = \frac{S - 1}{S + 1} = \frac{\dfrac{V_{max}}{V_{min}} - 1}{\dfrac{V_{max}}{V_{min}} + 1} = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$$

5) Derive an expression for the input impedance of the dissipation less lines. (N/D-13), (A/M-15), (M/J-14), (N/D-16), (M/J-12),

The input impedance of dissipation less line is given by,

$$Z_S = Z_o \cdot \frac{Z_R + Z_O \tanh \gamma l}{Z_O + Z_R \tanh \gamma l} \qquad \begin{aligned} Z_O &= R_O \\ \gamma &= j\beta \end{aligned}$$

$$= R_O \cdot \frac{Z_R + R_O \tanh j\beta l}{R_O + Z_R \tanh j\beta l}$$

$$Z_S = R_O \cdot \frac{Z_R + Z_O \, j\tanh\beta l}{Z_O + Z_R \, j\tanh\beta l}$$

The impedance is complex in general and is periodic with variation of βs, the period being π or s = λ/2. It can be also obtained as,

$$Z_S = Z_O \cdot \frac{Z_R \cosh\gamma l + Z_O \sinh\gamma l}{Z_O \cosh\gamma l + Z_R \sinh\gamma l}$$

$$= Z_O \cdot \frac{Z_R\left(\dfrac{e^{\gamma l} + e^{-\gamma l}}{2}\right) + Z_O\left(\dfrac{e^{\gamma l} - e^{-\gamma l}}{2}\right)}{Z_O\left(\dfrac{e^{\gamma l} + e^{-\gamma l}}{2}\right) + Z_R\left(\dfrac{e^{\gamma l} - e^{-\gamma l}}{2}\right)}$$

$$= Z_O \cdot \frac{\dfrac{Z_R\left(e^{\gamma l}\right) + Z_R\left(e^{-\gamma l}\right) + Z_O\left(e^{\gamma l}\right) - Z_O\left(e^{-\gamma l}\right)}{\not 2}}{\dfrac{Z_O e^{\gamma l} + Z_O e^{-\gamma l} + Z_R e^{\gamma l} - Z_R e^{-\gamma l}}{\not 2}}$$

$$= Z_O \, \frac{e^{\gamma l}\left(Z_R + Z_O\right) + e^{-\gamma l}\left(Z_R - Z_O\right)}{e^{\gamma l}\left(Z_O + Z_R\right) + e^{-\gamma l}\left(Z_R - Z_O\right)}$$

$$= Z_O \, \frac{e^{\gamma l}\left(Z_R + Z_O\right)\left[1 + \dfrac{e^{-\gamma l}\left(Z_R - Z_O\right)}{e^{\gamma l}\left(Z_R + Z_O\right)}\right]}{e^{\gamma l}\left(Z_R + Z_O\right)\left[1 + \dfrac{e^{-\gamma l}}{e^{\gamma l}}\left[\dfrac{Z_R - Z_O}{Z_R + Z_O}\right]\right]}$$

$$= Z_O \, \frac{\left(1 + K e^{-\gamma l} e^{-\gamma l}\right)}{\left(1 - K e^{-\gamma l} e^{-\gamma l}\right)}$$

$$Z_S = Z_O\left(\frac{1 + K e^{-2\gamma l}}{1 - K e^{-2\gamma l}}\right) \qquad \begin{aligned} \gamma &= j\beta \\ Z_O &= R_O \end{aligned}$$

$$Z_S = R_O\left(\frac{1 + K e^{-2j\beta l}}{1 - K e^{-2j\beta l}}\right)$$

Taking magnitude

$$Z_S = R_O\left(\frac{1 + |K|\underline{|\phi - 2\beta l}}{1 - |K|\underline{|\phi - 2\beta l}}\right)$$

i) The Input impedance is maximum $\phi - 2\beta l = 0$

$$\phi - 2\beta l = 0$$

$$l = {\phi}/{2\beta}$$

$$Z_{S(\max)} = R_O \frac{1 + |K|}{1 - |K|}$$

$$Z_{S\max} = R_O \; S$$

ii) Along the line, if we travel at the distance of λ/4 from the point where the impedance is maximum we get a point of $Z_{\min}$. $\left(\lambda = {2\pi}/{\beta}\right)$

$$l = {\phi}/{2\beta} + {\lambda}/{4} = {\phi}/{2\beta} + {2\pi}/{4\beta}$$

$$= \frac{2\phi\not\beta + \not 4\pi}{4\not\beta} = \frac{\not 2\left(\phi + \pi\right)}{\not 4\beta} = \frac{\phi + \pi}{2\beta}$$

$$l = \frac{\phi + \pi}{2\beta}$$

$$2\beta l = \phi + \pi$$

$$Z_{Smin} = R_O \frac{1 + |K| \underline{|\phi - 2\beta l}}{1 - |K| \underline{|\phi - 2\beta l}}$$

$$= R_O \frac{1 + |K| \underline{|\phi - (\phi + \pi)}}{1 - |K| \underline{|\phi - (\phi + \pi)}}$$

$$Z_{Smin} = R_O \frac{1 + |K| \underline{|-\pi}}{1 - |K| \underline{|-\pi}}$$

$$= R_O \frac{1 + |K|}{1 - |K|}$$

$$Z_{Smin} = \frac{R_O}{S}$$

6) Deduce the input impedance of an open circuit and short circuit dissipation less lines. (N/D-13), (M/J-12), (M/J-14), (M/J-16), (N/D-11), (A/M-18) (N/D-17)

The input impedance of dissipation less line is given by, We know that $Z_S = Z_o$

$$\frac{Z_R + R_O j \tan \beta l}{R_O + j Z_R \tan \beta l}$$

i) For a short circuit line:

$Z_R = 0$

$$Z_S = R_O \frac{R_O j \tan \beta l}{R_O}$$

$$Z_{SC} = j R_O \tan \beta l = j \times S$$
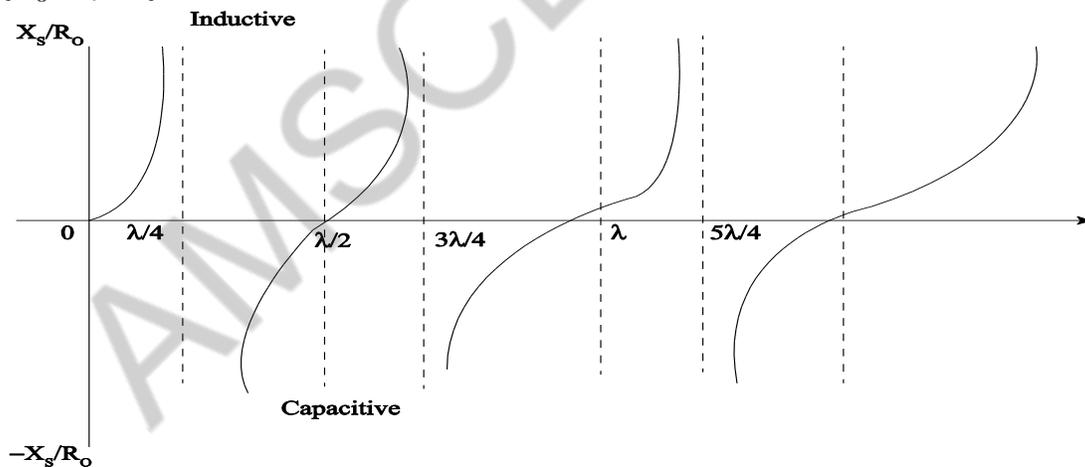
Short Circuit line

ii) For a open circuit line:

$$Z_{OC} = R_O \frac{Z_R + R_O j \tan \beta l}{R_O + j Z_R \tan \beta l} \qquad Z_R = \infty$$

$$= R_O \frac{\cancel{Z_R} \left[\dfrac{1 + R_O j \tan \beta l}{Z_R} \right]}{\cancel{Z_R} \left[\dfrac{R_O}{Z_R} + j \tan \beta l \right]}$$

$$= \frac{R_O}{j\tan\beta l} \qquad \frac{Z_{SC}}{R_O} = j\tan\beta l = j\frac{X_s}{R_O}$$

$$Z_{OC} = -jR_OC_O + \beta l = j\,X_s$$

Open Circuit line
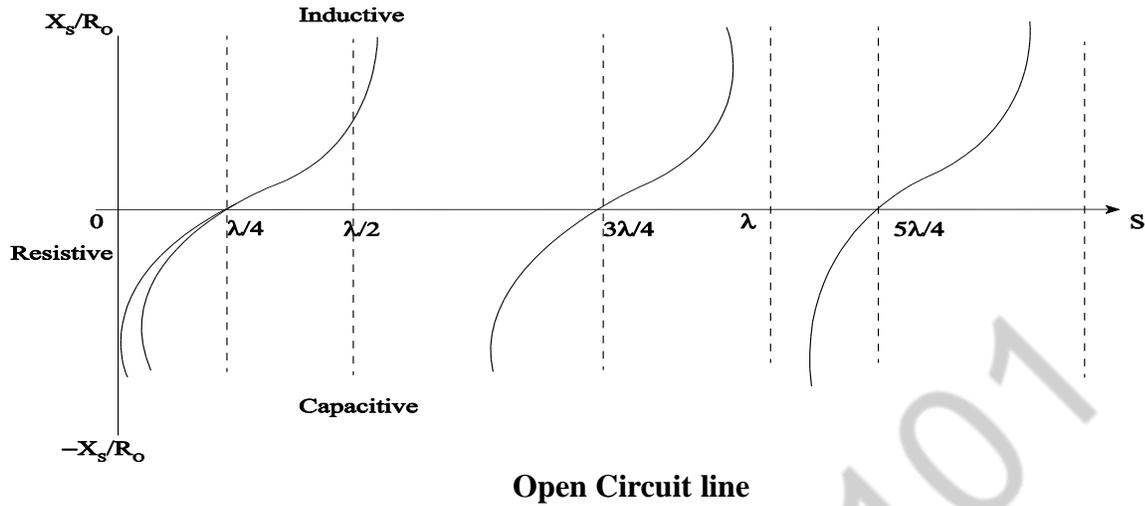
The variation of Z_{SC}/R_O is the length of line 't' can be plotted as follows,

Calculation of $\frac{X_s}{R_O} = \tan\beta l$

for $l = 0$ $\quad \frac{X_s}{R_O} = \tan\frac{2\pi}{\lambda}(0) = 0$

for $l = \lambda/4$ $\quad \frac{X_s}{R_O} = \tan\frac{2\pi}{\lambda}\left(\frac{\lambda}{4}\right)$

$$= \tan\frac{\pi}{2} - \infty$$

for $l = \lambda/2$ $\quad \frac{X_s}{R_O} = \tan\left(\frac{2\pi}{\lambda}\right)\left(\frac{\lambda}{2}\right)$

$$= \tan\pi = 0$$

for $l = 3\lambda/4$ $\quad \frac{X_s}{R_O} = \tan\left(\frac{2\pi}{\lambda}\right)\left(\frac{3\lambda}{4}\right)$

$$= \tan\frac{3\pi\lambda}{2\lambda} = \tan\frac{3\pi}{2} = \infty$$

for $l = \lambda$ $\quad \frac{X_s}{R_O} = \tan\left(\frac{2\pi}{\lambda}\right)(\lambda)$

$$= \tan 2\pi = 0$$

Calculation of $\frac{X_s}{R_O} = \cot\beta l$

for $l = 0$ $\quad \frac{X_s}{R_O} = \cot\frac{2\pi}{\lambda}(0) = -\infty$

for $l = \lambda/4$ $\quad \frac{X_s}{R_O} = \cot\frac{2\pi}{\lambda}\left(\frac{\lambda}{4}\right)$

$$= \cot\frac{\pi}{2} = 0$$

for $l = \lambda/2$ $\quad \frac{X_s}{R_O} = \cot\left(\frac{2\pi}{\lambda}\right)\left(\frac{\lambda}{2}\right)$

$$= \cot\pi = 0$$

for $l = \dfrac{3\lambda}{4}$ $\dfrac{X_s}{R_O} = \cot\left(\dfrac{2\pi}{\lambda}\right)\left(\dfrac{3\lambda}{4}\right)$

$$= \cot\dfrac{3\pi}{2} = 0$$

for $l = \lambda$ $\dfrac{X_s}{R_O} = \cot\left(\dfrac{2\pi}{\lambda}\right)(\lambda)$

$$= \cot 2\pi = 0$$

7) Derive the measurement of power and impedance on the line of negligible losses. (N/D-15), (M/J-14),(N/D-17), (N/D-11) (N/D-17)

We know that the voltage and the current on dissipation less line is given by

$$V = \dfrac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(1 + |K|\right) \underline{|\phi - 2\beta l}$$

$$V_{max} = \dfrac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(1 + |K|\right)$$

$$I = \dfrac{I_R}{2R_O}\left(Z_R + R_O\right)\left(1 + |K|\right) \underline{|\phi - 2\beta l}$$

$$I_{max} = \dfrac{I_R}{2R_O}\left(Z_R + R_O\right)\left(1 + |K|\right)$$

$$Z_{max} = \dfrac{V_{max}}{I_{max}} = \dfrac{\dfrac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(1 + |K|\right)}{\dfrac{I_R}{2R_O}\left(Z_R + R_O\right)\left(1 + |K|\right)}$$

$$Z_{max} = R_O$$

$||^{rly}$

$$V_{min} = \dfrac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(1 - |K|\right)$$

$$I_{min} = \dfrac{I_R}{2R_O}\left(Z_R + R_O\right)\left(1 - |K|\right)$$

$$Z_{min} = \dfrac{V_{min}}{I_{min}} = R_O$$

When the voltage is high, the current will be low at the same instant of time

$$\dfrac{V_{min}}{I_{min}} = \dfrac{\dfrac{V_R}{2Z_R}\left(Z_R + R_O\right)\left(1 + |K|\right)}{\dfrac{I_R}{2R_O}\left(Z_R + R_O\right)\left(1 - |K|\right)}$$

$$R_{max} = R_O\dfrac{1 + |K|}{1 - |K|} = R_O\ S$$

$$\dfrac{V_{min}}{I_{min}} = R_{min} = \dfrac{R_O}{S} = \dfrac{1 - |K|}{1 + |K|}$$

The effective power flowing into the resistance is given by

$$P = \dfrac{V^2_{max}}{R_{max}}, \quad P = \dfrac{V^2_{min}}{R_{min}}$$

$$P^2 = \dfrac{V^2_{max}V^2_{min}}{R_{max}R_{min}}$$

$$P = \dfrac{|V_{max} - V_{min}|}{R_O}$$

(or)

$$P = \left| I_{max} \cdot I_{min} \right| R_O$$

Measurement of load impedance we know that the input impedance.

$$Z_{Smin} = R_O \left[\frac{Z_R + jR_O \tan \beta l}{R_O + jZ_R \tan \beta l} \right]$$

We know that $R_{min} = R_O/S$

From V_{min} at a distance of 't' meters away we get the load impedance.

$$\frac{R_O}{S} = R_O \left[\frac{Z_R + jR_O \tan \beta l'}{R_O + jZ_R \tan \beta l'} \right]$$

$$R_O + jZ_R \tan \beta l' = S(Z_R + jR_O \tan \beta l')$$

$$R_O + jZ_R \tan \beta l' = SZ_R + SjR_O \tan \beta l'$$

$$R_O + jZ_R \tan \beta l' - SjR_O \tan \beta l' = SZ_R$$

$$Z_R = \frac{R_O [1 - j S \tan \beta l']}{S - j \tan \beta l'}$$

$$R_O [1 - j S \tan \beta l'] = SZ_R - jZ_R \tan \beta l'$$

$$R_O [1 - j S \tan \beta l'] = Z_R (S - j \tan \beta l')$$

Here V_{min} preferred than V_{max} since the position of V_{min} can be measured with greater accuracy.

IMPEDANCE MATCHING IN HIGH FREQUENCY LINES

3.1 INTRODUCTION

Transmission lines are used to transmit power from a source to a load. Maximum power is transmitted by the line to the load with minimum losses when no reflected wave is present or when the load impedance is equal to the characteristic impedance of the line, that is, when the load is properly matched. Usually this condition is not achieved.

In practice, however, it is not always possible to design a circuit whose input or output impedance is matched to the adjacent circuits. For example, there may be a circuit which is to be connected to a signal generator of $50\,\Omega$ output impedance but the input impedance of the circuit is not $50\,\Omega$. When this connection is made, not only the maximum power is not being transferred by the generator to the load, but the reflected wave might enter the generator and may alter its characteristics like frequency, etc. It is, therefore, essential to devise a technique which can avoid reflections from the circuits.

Impedance matching is one of the important aspects of high frequency circuit analysis. To avoid reflections and for maximum power transfer, the circuits have to be impedance matched. Transmission line sections can be used for the purpose of impedance matching. Due to low loss, the transmission line provides impedance matching with negligible loss of power. There are various impedance matching techniques using short length transmission lines ($\lambda/4$, $\lambda/8$) and using stubs (single stub, double stub) which are discussed in the following sections.

3.2 THE EIGHT WAVE LINE: ($\lambda/8$)

A transmission line of length $\dfrac{\lambda}{8}$ is called eight-wave $\left(\dfrac{\lambda}{8}\right)$ line.

The input impedance of a line of length $s = \dfrac{\lambda}{8}$ is:

$$Z_s = R_0 \left[\frac{Z_R + j R_0 \tan \dfrac{2\pi s}{\lambda}}{R_0 + j Z_R \tan \dfrac{2\pi s}{\lambda}} \right] \qquad \beta s = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{8} = \frac{\pi}{4}$$

$$= R_0 \left[\frac{Z_R + j R_0 \tan(\pi/4)}{R_0 + j Z_R \tan(\pi/4)} \right] \qquad \tan \frac{\pi}{4} = 1$$

$$R_s = R_0 \left[\frac{Z_R + j R_0}{R_0 + j Z_R} \right]. \qquad \ldots (3.1)$$

If such a line is terminated in pure resistance R_R

$$Z_s = R_0 \left[\frac{R_R + j R_0}{R_0 + j R_R} \right]. \qquad \ldots (3.2)$$

The numerator and denominator have identical magnitudes so that $|Z_s| = R_0$ ... (3.3)

Thus an eight wave line may be used to transform any resistance to an impedance with a magnitude equal to R_0 of the line or to obtain a magnitude match between a resistance of any value and a source of R_0 internal resistance.

3.3 THE QUARTER WAVE LINE, IMPEDANCE MATCHING

A transmission line of length $\frac{\lambda}{4}$ is called the quarter-wave line.

The generalised expression for the input impedance of the dissipationless line,

$$Z_s = R_0 \left[\frac{Z_R + j R_0 \tan(\beta \cdot s)}{R_0 + j Z_R \tan(\beta \cdot s)} \right]$$

Re-arranging the above equation,

$$Z_s = R_0 \left[\frac{\dfrac{Z_R}{\tan\left(\dfrac{2\pi s}{\lambda}\right)} + j R_0}{\dfrac{R_0}{\tan\left(\dfrac{2\pi s}{\lambda}\right)} + j Z_R} \right] \quad \text{where } \beta = \frac{2\pi}{\lambda} \qquad \ldots (3.4)$$

If the line is made quarter wave long or $s = \dfrac{\lambda}{4}$

$$\frac{2\pi s}{\lambda} = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{4} = \frac{\pi}{2} \quad , \qquad \tan\left(\frac{\pi}{2}\right) = \infty$$

$$Z_s = R_0 \left[\frac{j R_0}{j Z_R} \right] \quad , \qquad \boxed{Z_s = \frac{R_0^2}{Z_R}} \cdot \qquad \ldots (3.5)$$

(i.e.,) The input impedance of the line is equal to the square of R_0 of the line divided by the load impedance.

A quarter wave section of line may be thought of as a transformer to match a load of Z_R ohms to a source of Z_S ohms. Such a match can be obtained if the characteristic impedance R_0' of the matching quarter wave section of the line is properly chosen according to,

$$R_0' = \left| \sqrt{Z_S \cdot Z_R} \right|. \qquad \qquad ...(3.6)$$

R_0' of the matching section should be equal to the geometric mean of the source and load impedance.
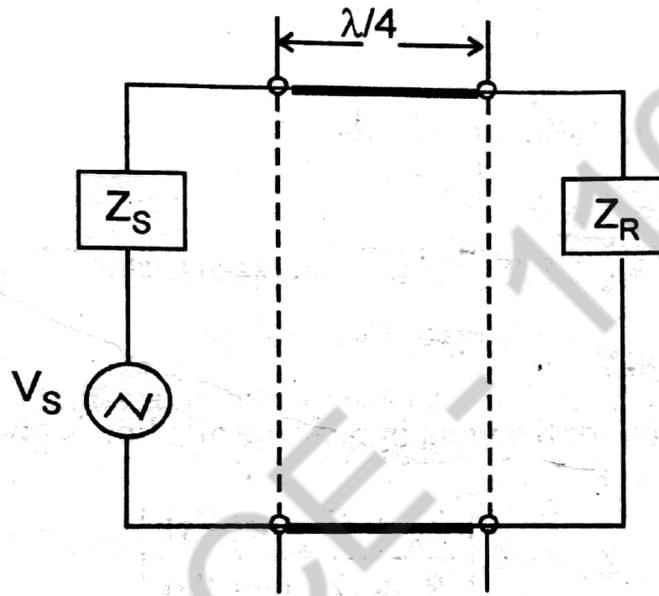
Fig. 3.1 : Quarter wave section as a impedance transformer

A quarter wave line may be considered as an impedance inverter in that it can transform a low impedance in to high impedance and vice versa.

An important application of the quarter wave matching section is to couple a transmission line to a resistive load such as an antenna. The quarterwave matching section then must be designed to have a characteristic impedance R_0' so chosen that the antenna resistance R_A is transferred to a value equal to the characteristic impedance R_0 of the transmission line. The characteristic impedance of the matching section is then given by,

$$R_0' = \sqrt{R_A \cdot R_0} \qquad \qquad ...(3.7)$$

The range of values of R_A and R_0 that can be matched satisfactorily is limited roughly to about 10 to 1.

A quarter wave transformer may also be used if the load is not pure resistance. It should then be connected between points corresponding to I_{max} or E_{min} at which places the transmission

line has resistive impedance given by $\dfrac{R_0}{S}$ or SR_0. For step down in impedance from the line value of R_0 the matching transformer characteristic impedance should be,

$$R_0' = \sqrt{R_0 \cdot \dfrac{R_0}{S}} = R_0\sqrt{\dfrac{1}{S}} \qquad \ldots (3.8)$$

Another application of the short circuited quarter wave line is an insulator to support an open wire line or the center conductor of a coaxial cable. Such lines are referred to as **copper insulators**.
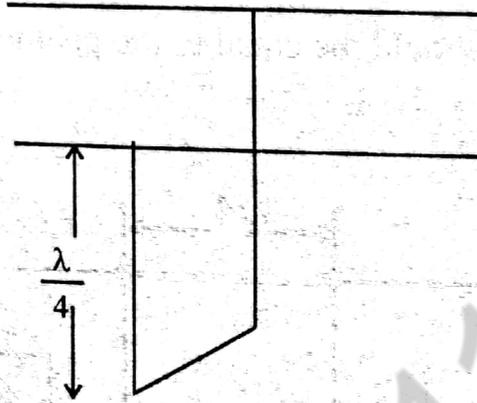
Fig. 3.2 : Quarterwave line as-insulator

3.4 THE HALF-WAVE LINE

The generalised expression for the input impedance of a line is given by,

$$Z_{in} = R_o\left[\dfrac{Z_R + jR_o\tan(\beta s)}{R_o + jZ_R\tan(\beta s)}\right]$$

But $\beta = \dfrac{2\pi}{\lambda}$

$$Z_{in} = R_o\left[\dfrac{Z_R + jR_o.\tan\left(\dfrac{2\pi}{\lambda}s\right)}{R_o + jZ_R.\tan\left(\dfrac{2\pi}{\lambda}s\right)}\right]$$

But for a half wave line, $s = \dfrac{\lambda}{2}$,

$$Z_{in} = R_o\left[\dfrac{Z_R + jR_o.\tan\left(\dfrac{2\pi}{\lambda}\cdot\dfrac{\lambda}{2}\right)}{R_o + jZ_R.\tan\left(\dfrac{2\pi}{\lambda}\cdot\dfrac{\lambda}{2}\right)}\right]$$

$$Z_{in} = R_o \left[\frac{Z_R + jR_o \tan(\pi)}{R_o + jZ_R \tan(\pi)} \right] \qquad \dots (3.9)$$

$$\therefore Z_{in} = R_o \left[\frac{Z_R}{R_o} \right]$$

$$\boxed{\therefore Z_{in} = Z_R} \qquad \dots (3.10)$$
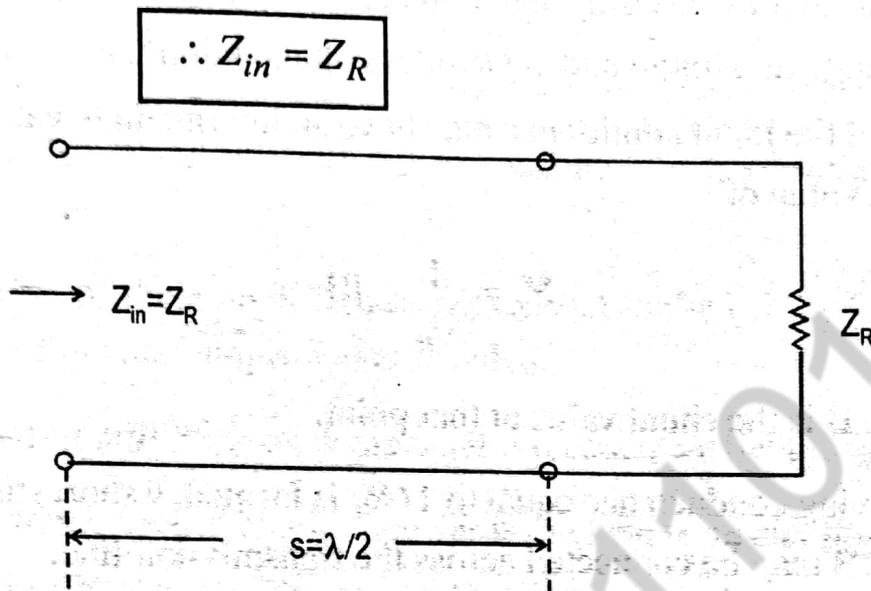
Fig. 3.3 : A half wave line

From it is clear that a half-wave line repeats its terminating impedance. In other words, the half wave line may be considered as one to one transformer.

3.5 IMPEDANCE MATCHING BY STUBS

For maximum power transfer, we know that the source and load impedances should match. However, in the case of long transmission lines, these impedances must be equal to the characteristic impedance Z_0 of the line. In many situations, the source (such as transmitter) and the load (such as an antenna) connected by the long transmisison line never match in impedance values. In such situations, maximum power does not get transferred from the transmitter to the antenna.

We have seen that stub lines can transform and match impedances. So, to match impedances in long transmission lines, stub lines of suitable lengths can be employed. There are two methods of stub-line matching: single-stub matching and double-stub matching.

3.6 SINGLE STUB IMPEDANCE MATCHING ON A LINE

→ For greatest efficiency and delivered power, a high frequency transmission line should be operated as a smooth line (or) with an R_0 termination.

→ However the usual loads such as an antenna do not in general have a resistance of value equal to R_0, so that in many cases it is necessary to introduce an impedance transforming section between line and the load to make the load appear to the line as a resistance of value R_0.

→ The quarter wave line or transformer is used as impedance matching devices.

→ Another means of achieving this is the use of an open or closed stub line of suitable length as a reactance shunted across the transmission line at a designated distance from the load, to tune the length of the line and the load to resonance with an anti-resonant resistance equal to R_0.

The theory of this method is easily stated in general terms. Since the input conductance of a line is $1/SR_0$ at a voltage maximum and S/R_0 at a voltage minimum, then at some intermediate point A, the real part of the Input admittance may have an intermediate value of $1/R_0$ or the input admittance at A has a value of

$$Y_s = \frac{1}{R_0} \pm jB.$$

... (3.11)

The susceptance B is the shunt value at that point.

→ After the point having conductance equal to $1/R_0$ is located, a short stub line having input susceptance of $\mp B$ may be connected across the transmission line.
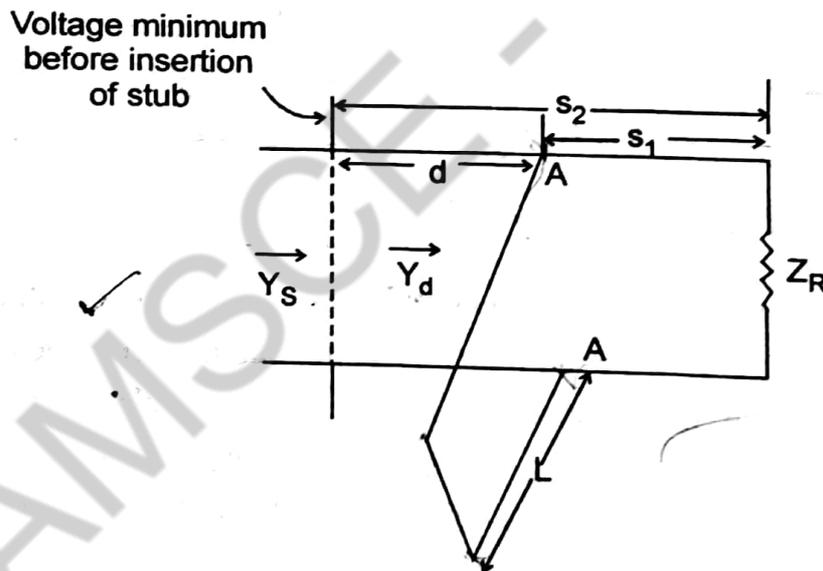
Fig. 3.4: Location of Single Stub for Impedance Matching

The input admittance at this point is given by,

$$Y_s = \frac{1}{R_0} \pm jB \mp jB = \frac{1}{R_0}.$$

... (3.12)

(or)

The input impedance of the transmission line at point A looking towards the load is,

$$Z_s = R_0.$$

... (3.13)

→ The line from the source to A is then terminated in R_0 and is a smooth line. From A to the load, reflection and standing waves occur; but since this distance can always be made less than a wavelength, the losses are not severe.

→ Since both the location and the length of the stub must be determined, two independent measurements must be made on the original line and load to secure sufficient data.

→ The most easily obtained measurements are the standing wave ratio S and the position of a voltage minimum, usually the minimum nearest to the load.

→ A voltage minimum is chosen rather than a maximum, since its position can be determined more accurately.

→ If the location of the stub is fixed with respect to the original voltage minimum, no knowledge of the load impedance is needed.

→ Because of the paralleling of elements, it is most convenient to work with admittances.

→ The input admittance Y_s, looking towards the load from any point on the line, may be written as:

$$Y_S = \frac{1}{Z_S} = \frac{1}{R_0}\left(\frac{1-|K|\angle\phi-2\beta s}{1+|K|\angle\phi-2\beta s}\right) \qquad \text{... (3.14)}$$

Writing $G_0 = \frac{1}{R_0}$ and changing to rectangular coordinates,

$$Y_S = G_o\left[\frac{1-|K|\cos(\phi-2\beta s)-j|K|\sin(\phi-2\beta s)}{1+|K|\cos(\phi-2\beta s)+j|K|\sin(\phi-2\beta s)}\right] \qquad \text{... (3.15)}$$

and upon rationalizing

$$Y_S = G_o\left[\frac{1-|K|^2-2j|K|\sin(\phi-2\beta s)}{1+|K|^2+2|K|\cos(\phi-2\beta s)}\right] = G_{\hat{s}} + jB_{S} \qquad \text{... (3.16)}$$

Expressing the shunt conductance as a dimensionless ratio $\dfrac{G_S}{G_0}$ or on a per unit basis,

$$\frac{G_S}{G_0} = \left[\frac{1-|K|^2}{1+|K|^2+2|K|\cos(\phi-2\beta s)}\right] \qquad \text{... (3.17)}$$

and the shunt susceptance on a per unit basis is,

$$\frac{B_S}{G_0} = \left[\frac{-2|K|\sin(\phi-2\beta s)}{1+|K|^2+2|K|\cos(\phi-2\beta s)}\right] \qquad \text{... (3.18)}$$
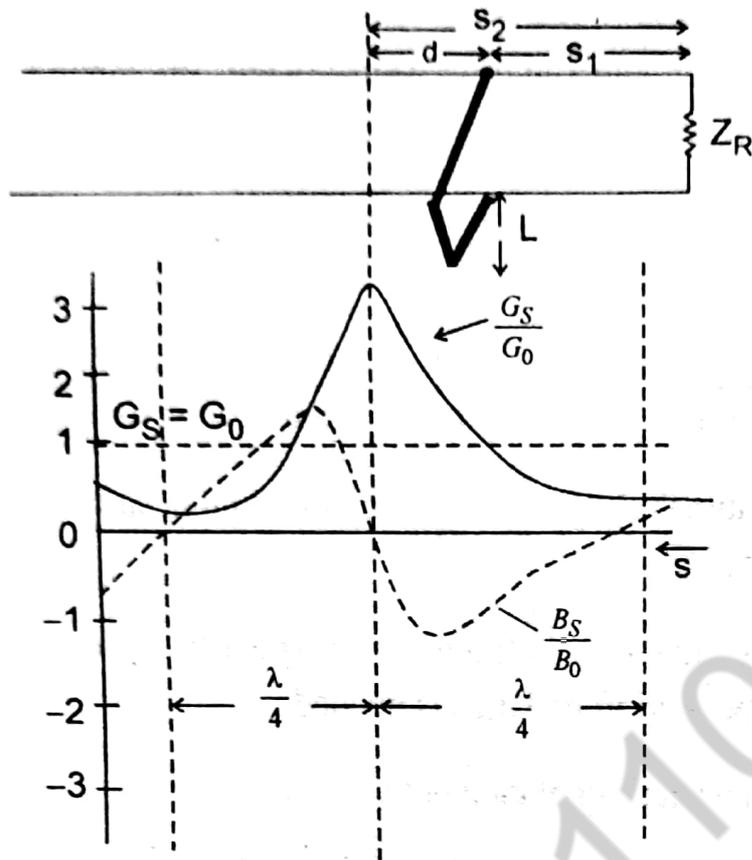
Fig. 3.5 : Admittance conditions on a line indicating proper location of the stub for |K| = 0.5

The value of $\dfrac{G_S}{G_0}$ has a maximum and this maximum occurs for the value s_2 at which the cosine term is -1 (or) $\phi - 2\beta s_2 = -\pi$,

$$\phi - 2\beta s_2 = -\pi$$

$$\boxed{s_2 = \dfrac{\phi + \pi}{2\beta}} \qquad \qquad \dots (3.19)$$

At a distance s_2 from the load,

$$\dfrac{G_S}{G_0} = \dfrac{1 - |K|^2}{1 + |K|^2 - 2|K|} = \dfrac{(1 + |K|)(1 - |K|)}{(1 - |K|)^2}$$

$$\dfrac{G_S}{G_0} = \dfrac{1 + |K|}{1 - |K|} = S \qquad \qquad \dots (3.20)$$

Since this equation states that $R_s = \dfrac{R_0}{S}$, the point of maximum $\dfrac{G_S}{G_0}$ is recognized as a point of minimum voltage, at a distance s_2 from the load.

At a distance s_1 from the load it can be seen that $G_s = G_0$. This is the point at which the stub is to be connected and the value of G_s / G_0 is unity at that point.

$$1 = \frac{1 - |K|^2}{\cancel{1} + |K|^2 + 2|K|\cos(\phi - 2\beta s_1)} \quad \dots (3.21)$$

$$\cancel{1} + |K|^2 + 2|K|\cos(\phi - 2\beta s_1) = \cancel{1} - |K|^2$$

$$\cos(\phi - 2\beta s_1) = -|K|$$

$$\cos^{-1}(-|K|) = \phi - 2\beta s_1$$

Since $\cos^{-1}(-K) = -\pi + \cos^{-1}|K|$

$$\phi - 2\beta s_1 = -\pi + \cos^{-1}|K|$$

$$d = \frac{\cos^{-1}\left(\frac{S-1}{S+1}\right) \lambda/4}{\pi}$$

$$\quad \dots (3.22)$$

$$s_1 = \frac{\phi + \pi - \cos^{-1}|K|}{2\beta} \quad \dots (3.23)$$

Hence the distance d from the voltage minimum to the point of stub connection is,

$$\boxed{d = s_2 - s_1} \quad \dots (3.24)$$

Substituting 3.19 and 3.23 we get,

$$d = \frac{\phi + \pi}{2\beta} - \left[\frac{\phi + \pi - \cos^{-1}|K|}{2\beta} \right]$$

$$d = \frac{\cos^{-1}|K|}{2\beta} = \frac{\cos^{-1}\left(\frac{S-1}{S+1}\right)}{2 \cdot \frac{2\pi}{\lambda}} = \frac{\cos^{-1}\left(\frac{S-1}{S+1}\right)\frac{\lambda}{4}}{\pi} \cdot$$

The stub should then be located at a distance d measured in either direction from a voltage minimum. Ordinarily the stub is placed on the load side on that minimum which is nearest to the load given by:

$$\boxed{d = \frac{\cos^{-1}\left(\frac{S-1}{S+1}\right)\frac{\lambda}{4}}{\pi}} \quad \dots (3.25)$$

The input susceptance of the line at the stub location nearest to the load can be obtained from equation (3.18) and equation (3.22).

$$B_s = G_0 \cdot \left[\frac{-2|K|\sin(-\pi + \cos^{-1}|K|)}{1 + |K|^2 + 2|K|\cos(-\pi + \cos^{-1}|K|)} \right] \quad \dots (3.26)$$

$$\cos\left(-\pi \pm \cos^{-1}|K|\right) = -|K|$$

$$\sin\left(-\pi \pm \cos^{-1}|K|\right) = \pm\sqrt{1-K^2}$$

$$B_s = G_0\left(\frac{2|K|\sqrt{1-k^2}}{1+|K|^2-2|K|^2}\right)$$

$$B_s = G_0\left(\frac{2|K|\sqrt{1-k^2}}{1-|K|^2}\right) \qquad \text{... (3.27)}$$

The susceptance of the stub required to cancel the line susceptance must be negative of B_s.

The susceptance of the short circuited stub is

$$B_{sc} = -G_0 \cot \beta L = \frac{-G_0}{\tan \beta L} \qquad \text{... (3.28)}$$

where L is the length of the short circuited stub.

If the stub and the line have equal G_o, then

$$B_{SC} = -B_S$$

$$\frac{G_0}{\tan \beta L} = G_0\left(\frac{2|K|\sqrt{1-k^2}}{1-|K|^2}\right) \qquad \text{... (3.29)}$$

$$\frac{G_0}{\tan \beta L} = G_0\left(\frac{2|K|}{\sqrt{1-|K|^2}}\right)$$

$$\frac{1}{\tan\left(\frac{2\pi}{\lambda}\right)L} = \frac{2|K|}{\sqrt{1-|K|^2}}$$

$$\boxed{L = = \frac{\lambda}{2\pi}\tan^{-1}\frac{\sqrt{1-|K|^2}}{2|K|}} \qquad \text{... (3.30)}$$

By use of the standing wave ratio existing before connection of the stub, this equation may be conveniently expressed as,

$$L = \frac{\lambda}{2\pi} \tan^{-1} \frac{\sqrt{S}}{S-1} \qquad \qquad \text{... (3.31)}$$

This is the length of the short circuited stub to be placed d meters towards the load from a point at which a voltage minimum existed before attachment of the stub.

- The susceptance of the line at *d* is then canceled and the line appears to be terminated in a resistance of value R_o at that point; it will be a smooth line between the generator and the point of connection of the stub.

- It is also possible to place the stub '*d*' meters towards the source from the voltage minimum. The sign of the reatance is reversed on that side with respect to the sign for the location nearer the load. The stub length L' then should be

$$L' = \frac{\lambda}{2} - L \quad \text{for a short ciruited stub.} \qquad \text{... (3.32)}$$

Advantages of Short Circuited Stub:

i) Short circuited stub is preferred over an open circuited stub because of greater ease in construction.

ii) Inability of open circuit stub to maintain high enough insulation resistance at the open circuit point to ensure that the stub is really open circuited.

iii) A shorted circuit stub has a lower loss of energy due to radiation, since the short circuit can be definitely established with a large metal plate, effectively stopping all field propagation.

3.7 THE CIRCLE DIAGRAM FOR THE DISTORTIONLESS LINE

Design of dissipationless line can be simplified significantly by drawing circle diagram.

(a) Constant S Circle:

The input impedance equation for a dissipationless line may be written on a per unit basis as,

$$Z_s = R_0 \left[\frac{1 + |K| \angle \phi - 2\beta s}{1 - |K| \angle \phi - 2\beta s} \right]$$

$$\frac{Z_S}{R_0} = \frac{1 + |K| \angle \phi - 2\beta s}{1 - |K| \angle \phi - 2\beta s} \qquad \qquad \text{... (3.33)}$$

Since $\dfrac{Z_S}{R_0}$ is complex, it is possible to write

$$\frac{Z_S}{R_0} = r_a + j x_a \qquad\qquad \text{... (3.34)}$$

$r_a \rightarrow$ resistance, $x_a \rightarrow$ reactance.

where r_a and x_a are the values of resistance or reactance per unit of R_0. Then

$$\frac{r_a + j x_a}{1} = \frac{1 + |K|\angle\phi - 2\beta s}{1 - |K|\angle\phi - 2\beta s} \qquad\qquad \text{... (3.35)}$$

The most easily measured quantity is the standing wave ratio S, and $|K| = \dfrac{S-1}{S+1}$ is substituted in the above equation.

Using compendo and dividendo method for equation (3.35),

(i.e.,) $\dfrac{N+D}{N-D}$

$$\frac{r_a + j x_a + 1}{r_a + j x_a - 1} = \frac{1 + |K|\angle\phi - 2\beta s + 1 - |K|\angle\phi - 2\beta s}{1 + |K|\angle\phi - 2\beta s - 1 + |K|\angle\phi - 2\beta s}$$

$$\frac{r_a + j x_a + 1}{r_a + j x_a - 1} = \frac{2}{2|K|\angle\phi - 2\beta s} = \frac{1}{|K|\angle\phi - 2\beta s} \qquad\qquad \text{... (3.36)}$$

$$[r_a + j x_a + 1] |K|\angle\phi - 2\beta s = r_a + j x_a - 1$$

$$[(r_a + 1) + j x_a] |K|\angle\phi - 2\beta s = (r_a - 1) + j x_a$$

Equating the magnitudes

$$\left(\sqrt{(r_a+1)^2 + x_a^2}\right) |K| = \sqrt{(r_a-1)^2 + x_a^2}$$

Squaring on both the sides

$$\left((r_a+1)^2 + x_a^2\right) K^2 = (r_a-1)^2 + x_a^2$$

$$\left(r_a^2 + 2r_a + 1 + x_a^2\right) K^2 = r_a^2 - 2r_a + 1 + x_a^2$$

$$r_a^2 (K^2 - 1) + 2r_a (K^2 + 1) + x_a^2 (K^2 - 1) + (K^2 - 1) = 0$$

Dividing by factor $(K^2 - 1)$

$$r_a^2 + 2r_a\left(\frac{K^2+1}{K^2-1}\right) + x_a^2 + 1 = 0 \qquad \qquad \text{... (3.37)}$$

$$K = \frac{S-1}{S+1}$$

We know that $K = \dfrac{S-1}{S+1}$

$(a-b)^2 = a^2 + b^2 - 2ab$

$$\frac{K^2+1}{K^2-1} = \frac{\left(\dfrac{S-1}{S+1}\right)^2 + 1}{\left(\dfrac{S-1}{S+1}\right)^2 - 1} = \frac{(S-1)^2 + (S+1)^2}{(S-1)^2 - (S+1)^2}$$

$$= \frac{S^2 - 2S + 1 + S^2 + 2S + 1}{S^2 - 2S + 1 - S^2 - 2S - 1}$$

$$= \frac{2(S^2+1)}{-4S} = \frac{(S^2+1)}{2S}$$

$$\frac{K^2+1}{K^2-1} = -\left[\frac{(S^2+1)}{2S}\right] \qquad \qquad \text{... (3.38)}$$

Sub. equation (3.38) is equation (3.37),

$$r_a^2 + 2r_a\left[-\left(\frac{S^2+1}{2S}\right)\right] + x_a^2 + 1 = 0$$

Add ing $\left(\dfrac{S^2+1}{2S}\right)^2$ **on both sides,**

$$r_a^2 - 2r_a\frac{(S^2+1)}{2S} + \left(\frac{S^2+1}{2S}\right)^2 + x_a^2 + 1 = \left(\frac{S^2+1}{2S}\right)^2$$

$$\left(r_a - \left(\frac{S^2+1}{2S}\right)\right)^2 + x_a^2 = \left(\frac{S^2+1}{2S}\right)^2 - 1$$

$$= \frac{S^4 + 2S^2 + 1}{4S^2} - 1$$

$$\left[r_a - \left(\frac{S^2 + 1}{2S} \right) \right]^2 = \frac{S^4 + 2S^2 + 1 - 4S^2}{4S^2}$$

$$= \frac{S^4 - 2S^2 + 1}{4S^2} = \left(\frac{S^2 - 1}{2S} \right)^2$$

$$\left[r_a - \left(\frac{S^2 + 1}{2S} \right) \right]^2 + x_a^2 = \left(\frac{S^2 - 1}{2S} \right)^2 \qquad \ldots (3.39)$$

This is an equation of the form

$$(x - c)^2 + y^2 = r^2. \qquad \ldots (3.40)$$

which is the equation of a family of circles of radius r and with centre shifted c units from the origin on the positive x-axis.

An actual circle will have a radius,

$$r = \frac{S^2 - 1}{2S} = \frac{S[S - 1/S]}{2S} = \frac{S - 1/S}{2} \qquad \ldots (3.41)$$

and a shift of the centre of the circle on the positive r_a axis,

$$c = \frac{S^2 + 1}{2S} = \frac{S(S + 1/S)}{2S} \qquad \ldots (3.42)$$
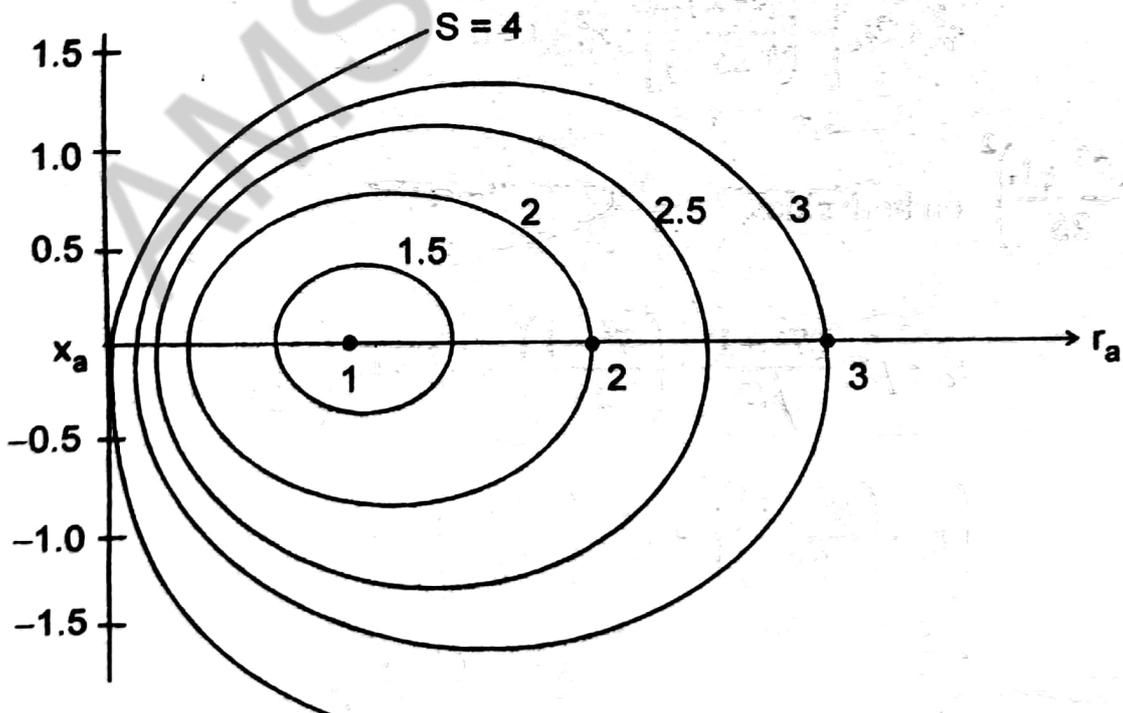
Fig. 3.6 : A family of constant S circles

→ The minimum value of S is unity. The circles for $S = 0$ represents the 1, 0 point.

→ The maximum value of S is infinity for the case of open and short circuit line termination.

→ As S increases, the radius of S circle increases and the center moves to the right and for the limiting case of $S = \infty$, the circle becomes x_a axis.

→ The line from the origin to a given point on the circle represents Z_S / R_0 in both magnitude and angle, with real and reactive components r_a and $j x_a$ respectively.

(b) Constant βs Circles:

From equation (3.36) we get,

$$\frac{(r_a + 1) + j x_a}{(r_a - 1) + j x_a} = \frac{1}{|K| \angle \phi - 2\beta s}$$

$$\frac{(r_a - 1) + j x_a}{(r_a + 1) + j x_a} = |K| \angle \phi - 2\beta s \qquad \dots (3.43)$$

Rationalizing LHS of the equation gives,

$$\frac{[(r_a - 1) + j x_a][(r_a + 1) - j x_a]}{(r_a + 1)^2 + x_a^2} = |K| \angle \phi - 2\beta s$$

$$(r_a - 1)(r_a + 1) = (r_a^2 - 1)$$

$$\frac{(r_a^2 - 1) + 2 j x_a + x_a^2}{(r_a + 1)^2 + x_a^2} = |K| \angle \phi - 2\beta s$$

⟹ real & imag

$$\frac{(r_a^2 - 1) + x_a^2}{(r_a + 1)^2 + x_a^2} + \frac{j 2 x_a}{(r_a + 1)^2 + x_a^2} = |K| \angle \phi - 2\beta s \qquad \dots (3.44)$$

The angle ϕ may be made zero inorder that the βs scale may start at 0° on the abscissa. Taking tangents of the angle on both the sides and $\phi = 0$, in equation (3.44).

$$\tan \left\{ \tan^{-1} \left[\frac{\dfrac{2 x_a}{(r_a^2 + 1) + x_a^2}}{\dfrac{(r_a^2 - 1) + x_a^2}{(r_a + 1)^2 + x_a^2}} \right] \right\} = \tan(-2\beta s)$$

$$\frac{2 x_a}{(r_a^2 - 1) + x_a^2} = \tan(-2\beta s)$$

$$(r_a^2 - 1) + x_a^2 = \frac{2 x_a}{\tan(-2\beta s)}$$

$$(r_a^2 - 1) + x_a^2 = \frac{-2x_a}{\tan 2\beta s} \qquad\qquad \because \ \tan(-\theta) = -\tan\theta$$

$$(r_a^2 - 1) + x_a^2 + \frac{2x_a}{\tan 2\beta s} = 0$$

Adding the term $\dfrac{1}{\tan^2(2\beta s)}$ on both the sides,

$$r_a^2 + x_a^2 + \frac{2x_a}{\tan 2\beta s} + \frac{1}{\tan^2 2\beta s} = 1 + \frac{1}{\tan^2 2\beta s}$$

$$r_a^2 + \left[x_a + \frac{1}{\tan 2\beta s} \right]^2 = 1 + \frac{1}{\tan^2 2\beta s}$$

$$r_a^2 + \left[x_a + \frac{1}{\tan 2\beta s} \right]^2 = \frac{1}{\sin^2 2\beta s}. \qquad\qquad ...\,(3.45)$$

The above equation is of the form

$$(x - c)^2 + y^2 = r^2. \qquad\qquad ...\,(3.46)$$

Lines of equal βs values are seen to be circles of radius, $r = \dfrac{1}{\sin 2\beta s}$ with a shift of center

downward on the x_a axis, $c = \dfrac{1}{\tan 2\beta s}$.
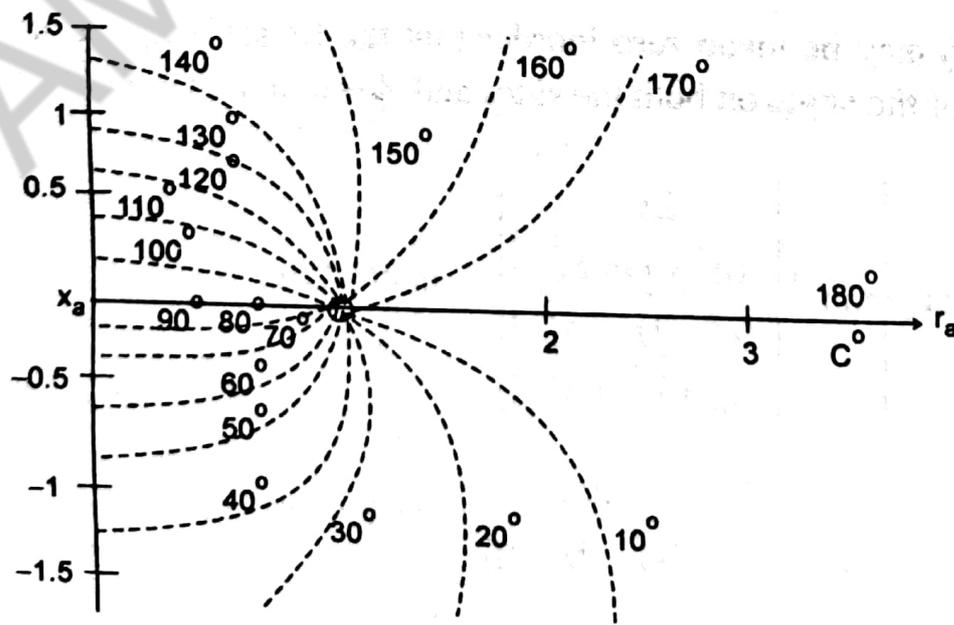
Fig. 3.7 : A family of constant βs circles

Drawback of Circle Diagram:

1) Constant S circles and βs circles are not concentric which makes it difficult to interpolate the circles.

2) Circle diagram can be used for the limited range of the impedance values with reasonable, practical chart size.

Applications of Circle Diagram:

1) To find input impedance of a line of any chosen length.

2) To find the input admittance of a length of a line.

3.8 SMITH CHART

It is a modified form of the circle diagram. Smith chart is a special polar diagram containing constant resistance circles, constant reactance circles, circles of constant SWR and used in solving transmission line and waveguide problems.

→ Smith Chart is the most useful graphical tool for solving transmission line problems very easily which would be difficult to solve by analytical methods.

→ The Smith Chart is obtained as follows.

The input impedance of a dissipationless line is,

$$Z_S = R_0 \left[\frac{1+|K| \angle \phi - 2\beta s}{1-|K| \angle \phi - 2\beta s} \right]$$

$$\frac{Z_S}{R_0} = \left[\frac{1+|K| \angle \phi - 2\beta s}{1-|K| \angle \phi - 2\beta s} \right] \qquad \ldots (3.47)$$

Normalized impedance $\frac{Z_S}{R_0}$ is a complex quantity and can be expressed as $r_a + j x_a$,

$$\frac{Z_S}{R_0} = r_a + j x_a = \frac{1+|K| \angle \phi - 2\beta s}{1-|K| \angle \phi - 2\beta s} \cdot \qquad \ldots (3.48)$$

The quantity $|K| \angle \phi - 2\beta s$ is a complex quantity and let it be given by,

$$|K| \angle \phi - 2\beta s = u + jv$$

$$\frac{Z_S}{R_0} = r_a + j x_a = \frac{1+(u+jv)}{1-(u+jv)}$$

On rationalizing the above equation,

$$r_a + j\,x_a = \frac{(1+u)+jv}{(1-u)-jv} \times \frac{(1-u)+jv}{(1+u)+jv}$$

$$= \frac{1-u+jv+u-u^2+juv+jv-juv+j^2v^2}{(1-u)^2+v^2}$$

$$= \frac{1-u^2-v^2+2jv}{(1-u)^2+v^2}$$

$$r_a + j\,x_a = \frac{1-u^2-v^2}{(1-u)^2+v^2} + j\frac{2v}{(1-u)^2+v^2}. \qquad \ldots (3.49)$$

Equating real and imaginary parts,

$$r_a = \frac{1-u^2-v^2}{(1-u)^2+v^2} \qquad \ldots (3.50)$$

$$x_a = \frac{2v}{(1-u)^2+v^2} \qquad \ldots (3.51)$$

a) Constant Resistance Circles: From equation of r we have, [eqn. 3.50]

$$r_a([1-u]^2+v^2)=1-u^2-v^2$$

$$r_a([1-2u+u^2]+v^2)=1-u^2-v^2$$

$$r_a - 2ur_a + u^2r_a + v^2r_a = 1-u^2-v^2$$

$$u^2(1+r_a)-2ur_a+v^2(1+r_a)=1-r_a$$

Dividng throughout by $(1+r_a)$

$$u^2 - 2u\left(\frac{r_a}{1+r_a}\right)+v^2=\frac{1-r_a}{1+r_a}$$

Completing the square by adding $\left(\dfrac{r_a}{1+r_a}\right)^2$ on both the sides, we get

$$u^2 - 2u\left(\frac{r_a}{1+r_a}\right)+\left(\frac{r_a}{1+r_a}\right)^2+v^2=\left(\frac{1-r_a}{1+r_a}\right)+\left(\frac{r_a}{1+r_a}\right)^2$$

$$\left(u - \frac{r_a}{1+r_a}\right)^2 + v^2 = \frac{1 - r_a^2 + r_a^2}{(1+r_a)^2}$$

$$\left(u - \frac{r_a}{1+r_a}\right)^2 + v^2 = \frac{1}{(1+r_a)^2}$$

$$\left(u - \frac{r_a}{1+r_a}\right)^2 + v^2 = \left(\frac{1}{1+r_a}\right)^2 \qquad \ldots (3.52)$$

Equation of a circle: $(x - c)^2 + y^2 = r^2$. Equation (3.52) represents the equation of a circle, r is the radius, centre is shifted c units from the origin.

The above equation (3.52) represents the equation of a circle of radius $\left(\dfrac{1}{1+r_a}\right)$ and centre $\left(\dfrac{r_a}{1+r_a}, 0\right)$.
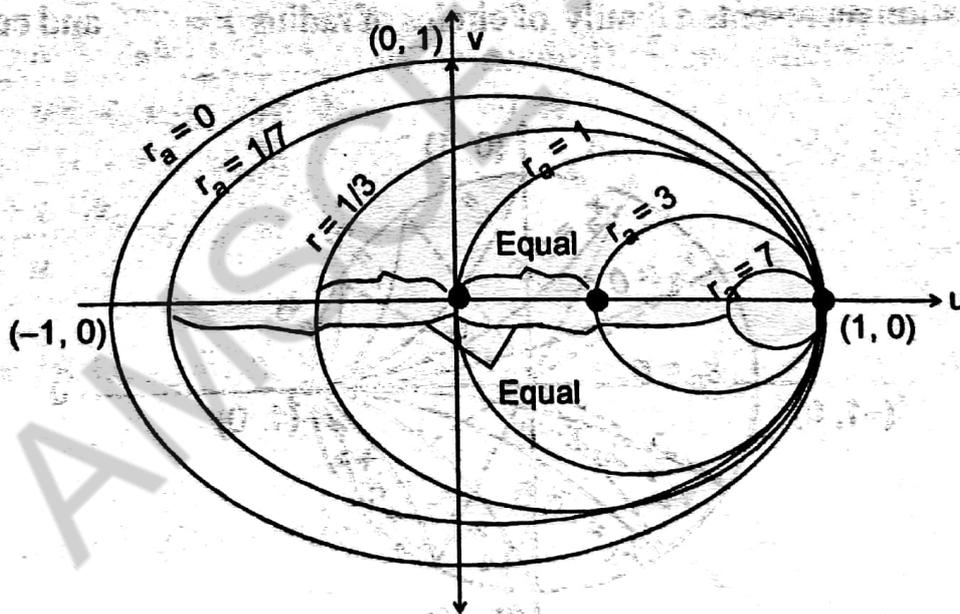
Fig. 3.8 : Constant r_a Circles

b) **Contant Reactance Circles:**

From the equation of x_a, equation 3.51

$$x_a[(1 - u)^2 + v^2] = 2v$$

$$x_a[1 - 2u + u^2 + v^2] - 2v = 0.$$

Dividing throughout by x_a

$$1 - 2u + u^2 + v^2 - \frac{2v}{x_a} = 0$$

$$(u - 1)^2 + v^2 - \frac{2v}{x_a} = 0.$$

Adding $\dfrac{1}{x_a^2}$ on both sides,

$$(u - 1)^2 + v^2 - \frac{2v}{x_a} + \frac{1}{x_a^2} = \frac{1}{x_a^2}$$

$$(u-1)^2 + \left(v - \frac{1}{x_a}\right)^2 = \frac{1}{x_a^2} \qquad \qquad \ldots (3.53)$$

The above equation represents a family of circles of radius $r = \dfrac{1}{x_a}$ and centre $c = \left(1, \dfrac{1}{x_a}\right)$.
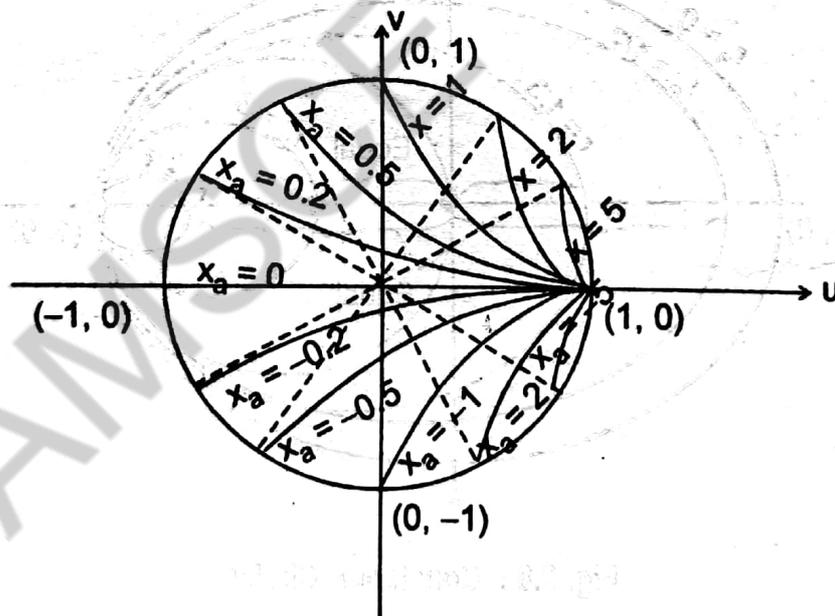
Fig. 3.9 : Constant reactance circles [x_a – circles]

The superposition of constant resistance circles and constant reactance circles forms the Smith chart.

c) **Constant S-Circles:** Consider the equation (3.52)

$$\left(u - \frac{r_a}{1+r_a}\right)^2 + v^2 = \left(\frac{1}{1+r_a}\right)^2.$$

Assume $V = 0$, centres of all the S circles lie on the horizontal real axis of the smith chart.

$$\left(u - \frac{r_a}{1 + r_a}\right)^2 = \left(\frac{1}{1 + r_a}\right)^2 \qquad \text{... (3.54)}$$

$$u - \frac{r_a}{1 + r_a} = \pm \frac{1}{1 + r_a}$$

$$u = \frac{r_a}{1 + r_a} \pm \frac{1}{1 + r_a} = \frac{r_a \pm 1}{1 + r_a}$$

$$u = \frac{r_a + 1}{1 + r_a} \quad \text{(or)} \quad \frac{r_a - 1}{1 + r_a}$$

i.e., $u = 1$ (or) $\dfrac{r_a - 1}{1 + r_a}$. $\qquad$ ... (3.55)

Consider $u = \dfrac{r_a - 1}{1 + r_a}$

$$\frac{1 + u}{1 - u} = \frac{r_a + 1 + r_a - 1}{r_a + 1 - (r_a - 1)} \quad \text{by Compendo-Dividendo Method.}$$

$$\frac{1 + u}{1 - u} = \frac{2r_a}{2} = r_a. \qquad \text{... (3.56)}$$

But $u + jv = |K| \angle \phi - 2\beta s$, since $V = 0$.

... (3.57)

$$u = |K|$$

Substituting the value in equation 3.56 we get,

$$r_a = \frac{1 + |K|}{1 - |K|} = S. \qquad \text{... (3.58)}$$

The constant S circle can be drawn with radius equal to distance between the centre of the Smith chart and the point $r_a = S$. The constant S circle cuts the horizontal axis at point $r_a = S$ on the right hand side of the centre while at point $r_a = \dfrac{1}{S}$ to the left hand side of the centre.

Applications of Smith Chart:

1) Plotting of impedance.

2) Finding the input impedance of the line.

3) Measurement of voltage standing wave ratio.

4) Measurement of reflection coefficient K [magnitude and phase].

5) Impedance to admittance conversion.

3.9 SINGLE STUB MATCHING USING SMITH CHART
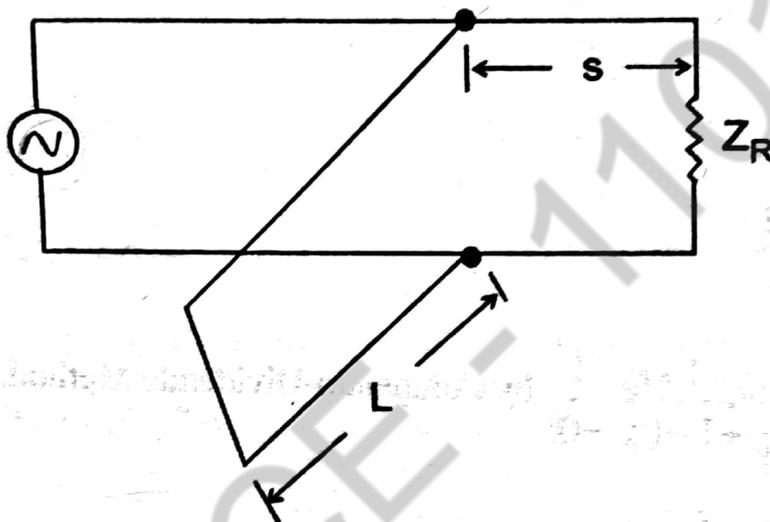
Fig. 3.10 : Single Stub Matching

The stub is connected in parallel with the line, hence it is convenient to work with admittances. Following are the steps to be followed while using smith chart for single stub matching analysis:

(1) First locate the normalized load impedance point on the smith chart, say point P.

(2) Draw a constant S circle with center of the chart O as the center and OP as radius.

(3) Locate the point Q, diametrically opposite to P. The point Q is the normalized load admittance.

(4) Locate the point of intersection of constant S circle with $\left(\dfrac{Y}{G_0}=1\right)$ unit conductance circle $g_i = 1$. There are two points of intersection. Take R the point of intersection nearer to Q, when moving from Q towards generator.

(5) Measure the distance between Q' and R' by moving from Q to R. (i.e) from load to generator. This distance corresponds to the distance (s) from the load at which the stub is connected.

(6) The susceptance at point R is $\pm jb$ of the line at the point of stub connection. This susceptance must be resonated by the stub line having a susceptance of $\mp jb$.

(7) The subsceptance of the stub is marked at intersection of $S = \infty$ circle and $\mp jb$ susceptance circle. This point is denoted as T.

(8) The length of the stub L is measured by moving from the short circuit end (extreme right) to point T.
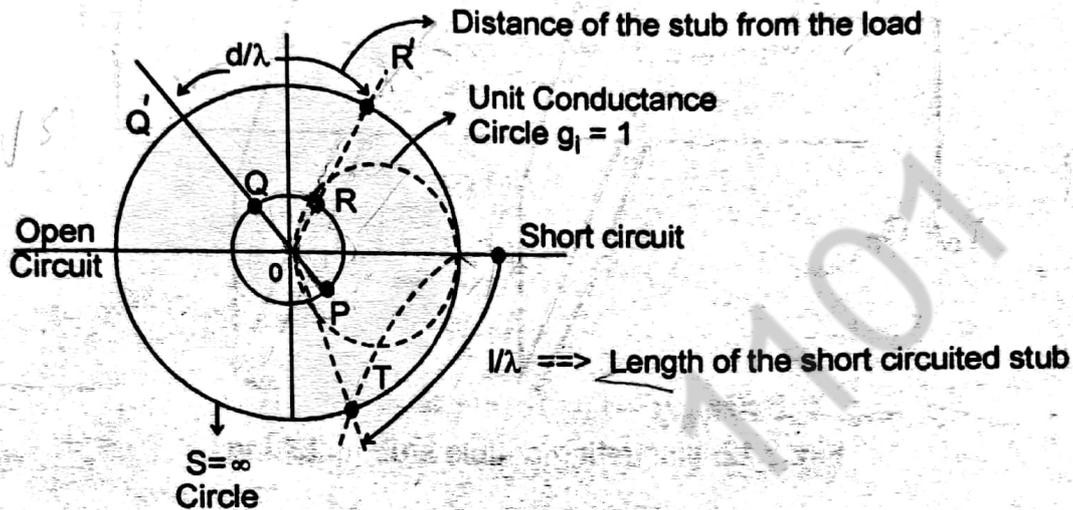
Fig. 3.11 : Single stub matching using Smith Chart

In the Figure 3.11,

P represents normalized load impedance point.

Q represents normalized load admittance point.

R represents intersection of constant S circle and with conductance circle.

T represents intersection of $S = \infty$ circle and $\mp jb$ susceptance circle.

3.10 DOUBLE STUB MATCHING

Limitations of Single Stub Matching

i) The stub has to be located at a definite point on the line. Single stub is adequate for a openwire line. For coaxial lines, placement of stub at exact point is difficult.

ii) Two adjustments were required in single stub, these being the location and length of the stub.

iii) This technique is suitable for fixed frequency only. If frequency changes, the location of the stub has to be changed.

To overcome this difficulty, double stub impedance matching is used. In this system, two different short circuited stubs are used for impedance matching, location of the stubs is arbitrary and the spacing between the two stubs is made $\frac{\lambda}{4}$.

The spacing d_2 is frequently made ($\lambda/4$). A spacing of ($3\lambda/8$) is also sometimes used, if the first stub is to be located very near to the load. ($\lambda/2$) spacing is avoided becuase it places the two stubs effectively in parallel at the same point.
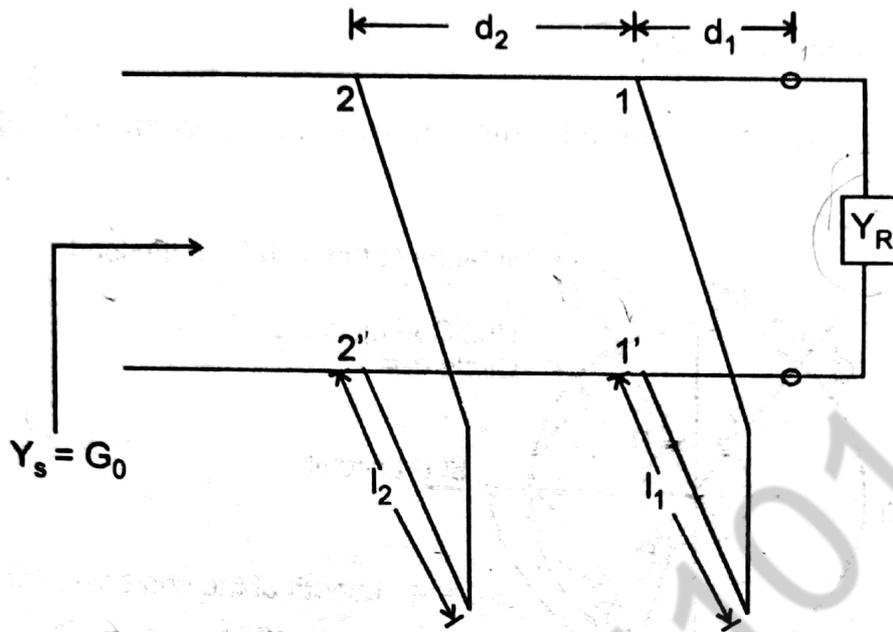
Fig. 3.12: Illustrating double stub matching

Let the first stub whose length is l_1 be located at the point 1 1' at a distance d_1 from the laod end.

The input impedance at any point on the line is given by:

$$Z_s = Z_o \left[\frac{Z_R + j Z_o \tan \beta s}{Z_o + j Z_R \tan \beta s} \right].$$

$$Z_s = \frac{1}{Y_s}, \quad Z_o = \frac{1}{Y_o} \text{ and } Z_R = \frac{1}{Y_R}$$

Substituting all these expressions into the expression for Z_s, we get

$$\frac{1}{Y_s} = \frac{1}{Y_o} \left[\frac{\left(\dfrac{1}{Y_R}\right) + \dfrac{1}{Y_o} j \tanh \beta s}{\dfrac{1}{Y_o} + \dfrac{1}{Y_R} j \tanh \beta s} \right]$$

$$\frac{1}{Y_s} = \frac{1}{Y_o} \left[\frac{1 + \dfrac{Y_R}{Y_o} j \tanh \beta s}{\dfrac{Y_R}{Y_o} j \tanh \beta s} \right] \quad \text{(by multiplying and dividing by } Y_R)$$

$$\therefore \frac{Y_s}{Y_o} = \frac{\dfrac{Y_R}{Y_o} + j\tan\beta s}{1 + \dfrac{Y_R}{Y_o} j\tan\beta s}$$

We have $\dfrac{Y_s}{Y_o} = Y_s$ = normalized input admittance

$$\frac{Y_R}{Y_o} = Y_R = \text{normalized load admittance}$$

$$\therefore Y_s = \frac{Y_R + j\tan\beta s}{1 + jy_R\tan\beta s}$$

conjugate

Rationalizing, we get

$$y_s = \frac{(y_R + j\tan\beta s)(1 - jy_R\tan\beta s)}{(1 + jy_R\, j\tan\beta s)(1 - jy_R\tan\beta s)}$$

$$y_s = \frac{y_R - jy_R^2\tan\beta s + j\tan\beta s + y_R\tan^2\beta s}{1 + y_R^2\tan^2\beta s}$$

$$= \frac{y_R(1 + \tan^2\beta s)}{1 + y_R^2\tan^2\beta s} + \frac{(1 - y_R^2)\tan\beta s}{1 + y_R^2\tan^2\beta s}$$

At point $11'$, we have $s = d_1$. Substituting we get

$$y_s = \frac{y_R(1 + \tan^2\beta d_1)}{1 + y_R^2\tan^2\beta d_1} + j\frac{(1 - y_R^2)\tan\beta d_1}{1 + y_R^2\tan^2\beta d_1} = g_1 + b_1 \qquad \ldots (3.59)$$

When a stub having a susceptance of $\pm b_{11}$ is added at this point, the new admittance value will be

$$y_R^1 = g_1 + jb_1^1 \qquad \ldots (3.60)$$

where $b_1{}^1 = b_1 \pm b_{11}$ and g_1 remains unchanged.

Now the input admittance of the line looking toward the load at $22'$ location should be:

$$Y_s = G_o \qquad \ldots (3.61)$$

Or the line should appear terminated in its characteristic impedance at that point. Thus, the 22' point should be at a location on the line having normalized admittance of:

$$Y_{22'} = \frac{Y_s}{G_o} = 1 \pm b_{22'} \qquad \qquad \ldots (3.62)$$

The $g = 1$ circle on the Smith chart will be in the locus of all such admittance resonated by a stub of susceptance $\mp j\, b_{22'}$ to give the desired normalized admittance $(Y_s/G_o) = 1$ at 22' on the line. Two cases are considered.

Case (i): Quarter-Wavelength spacing between two stubs ($d_2 = \lambda/4$)

The transformer formed by the quarter-wave line between 22' and 11' will transform all admittances which will lie on a second locus circle B found by displacing each point of circle A anticlockwise by a quarter wavelength on the chart (180° rotation of the chart). This second locus circle B is indicated in the Figure 3.13.

(a)

(b)
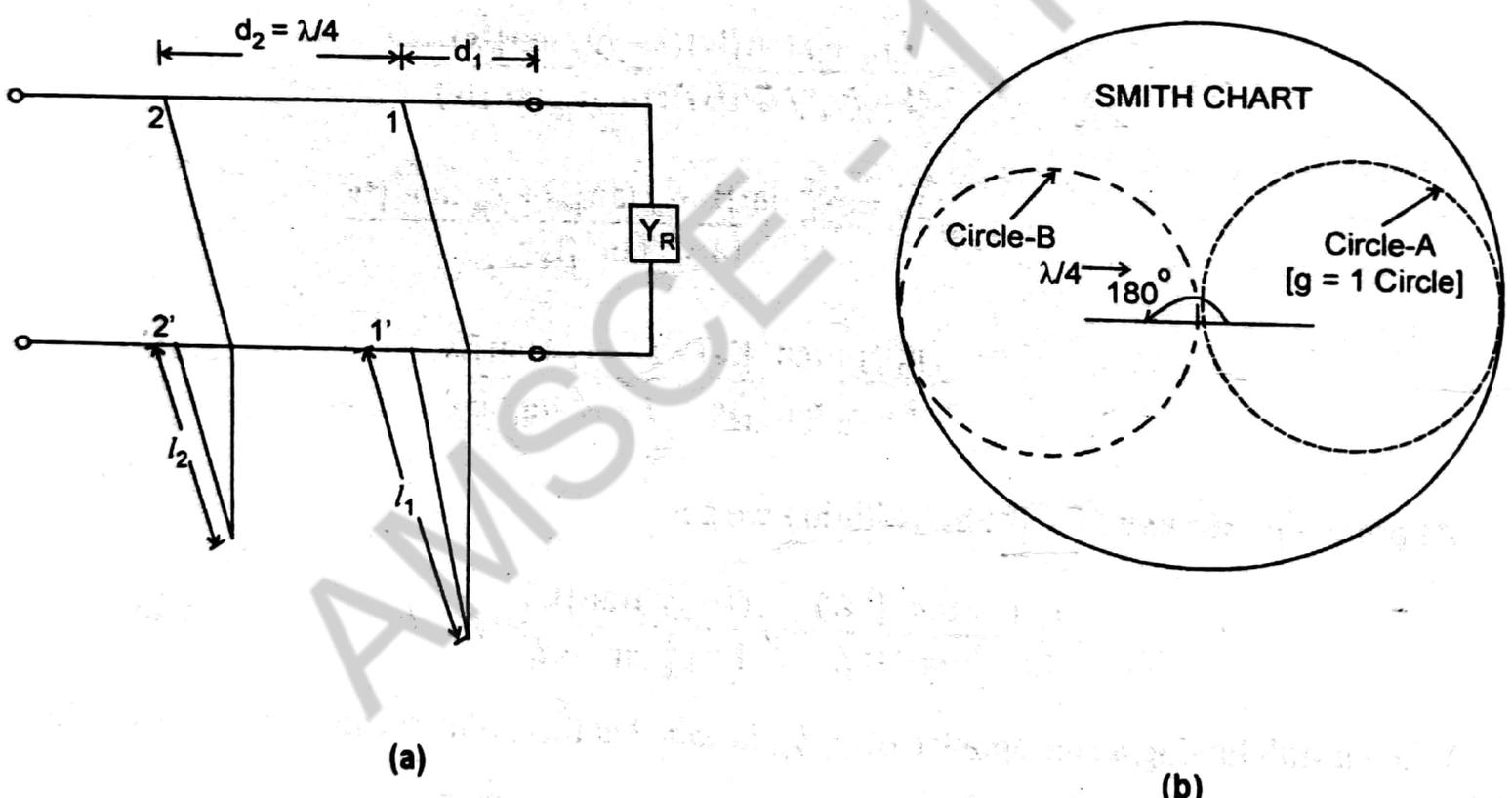
Fig. 3.13: (a) Double stub matching with $d_2 = \dfrac{\lambda}{4}$
(b) Corresponding Smith chart

Therefore, if stub-1 succeeds in transforming the input admittance of the line and load to the right of 11' into an admittance which will plot on the circle B locus, the quarter wavelength line will further transform the admittance into a value at 22' which will plot on the locus circle A and will have a value $y_{22'} = 1 \pm jb_{22'}$. Stub-2 can then resonate the admittance at 22' to the desired value $(Y_s/G_o) = 1 + j\,0$ for properly terminating the main line to the left of 22'.

In this case, there is a restriction on the value of g_1 conductance. The largest conductance component that can be transformed on to circle B is $g_1 = 1$. If g_1 is greater than 1, then that g_1 circle never cuts the B circle. If $g_1 > 1$, then a point on the line toward the source having normalized conductance less than 1 must be selected for connection of the stub. Such a point will be available since normalized conductance varies between S and $\frac{1}{S}$ in a $(\lambda/4)$ distance along the line.

Case (ii) : Three-eighth's wavelength spacing between two stubs $(d_2 = 3\lambda/8)$

When $(\lambda/4)$ spacing is not suitable, the same procedure holds good by drawing a different locus circle B. The desired circle B is now drawn by rotating circle A through $(3/8)$ λ $(= 270°)$ counter clock-wise as shon in Figure 3.14 (b). Now the diameter of the circle B will project vertically downwards from the centre of the chart.
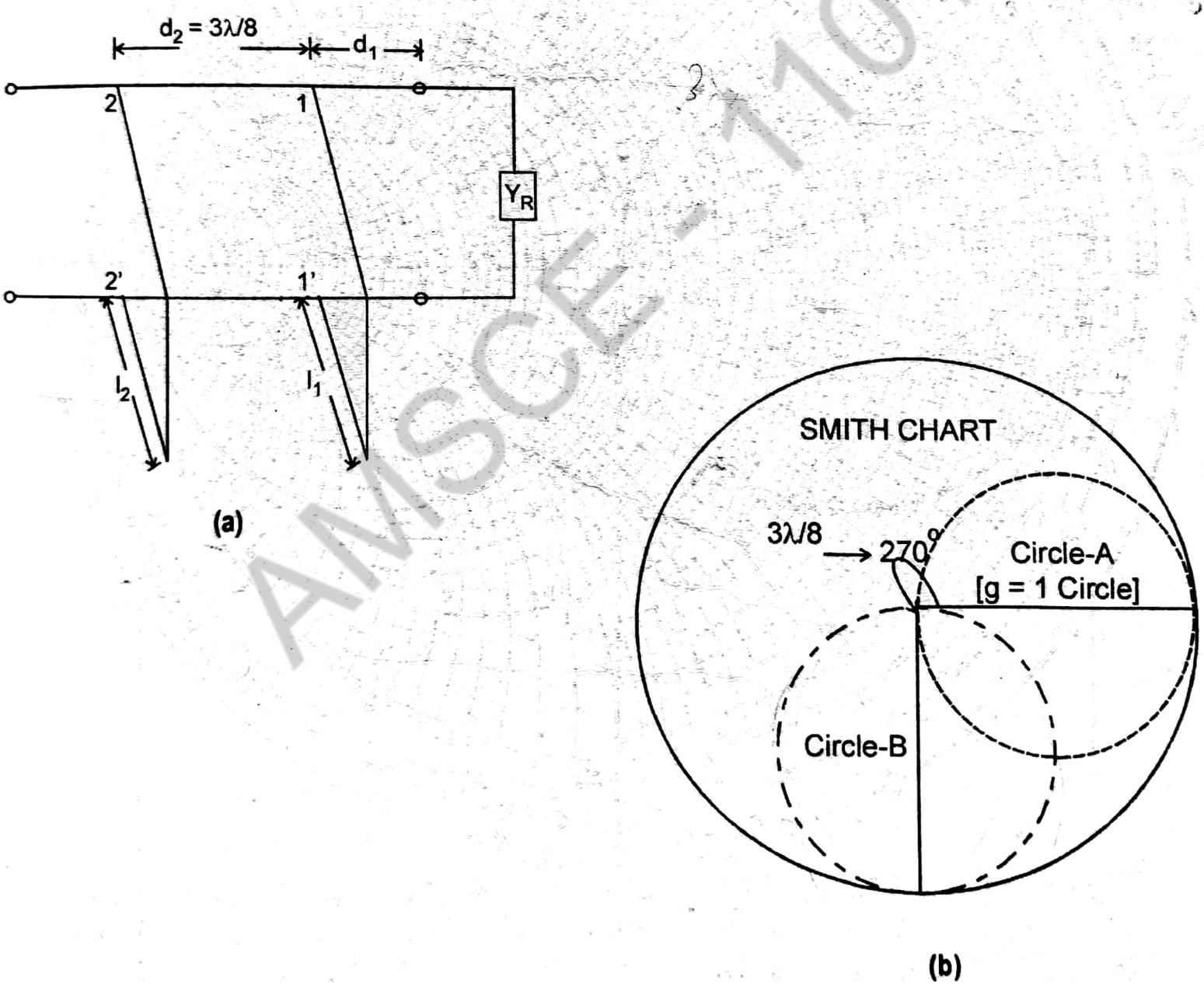
Fig. 3.14: (a) Double stub matching with $d_2 = (3\lambda/8)$

(b) Corresponding Smith chart

Smith Chart for Double Stub Matching:

Looking at terminals 11' towards the load, the normalized admittance y_R is given by $y_R = 0.5 \ \lambda \ j \ 1.0$. Design a double-stub matching system with ($\lambda/4$) spacing between the stubs.

Solution:

1. The given admittance y_R is plotted as point A on the Smith chart of Figure 3.15 as shown.
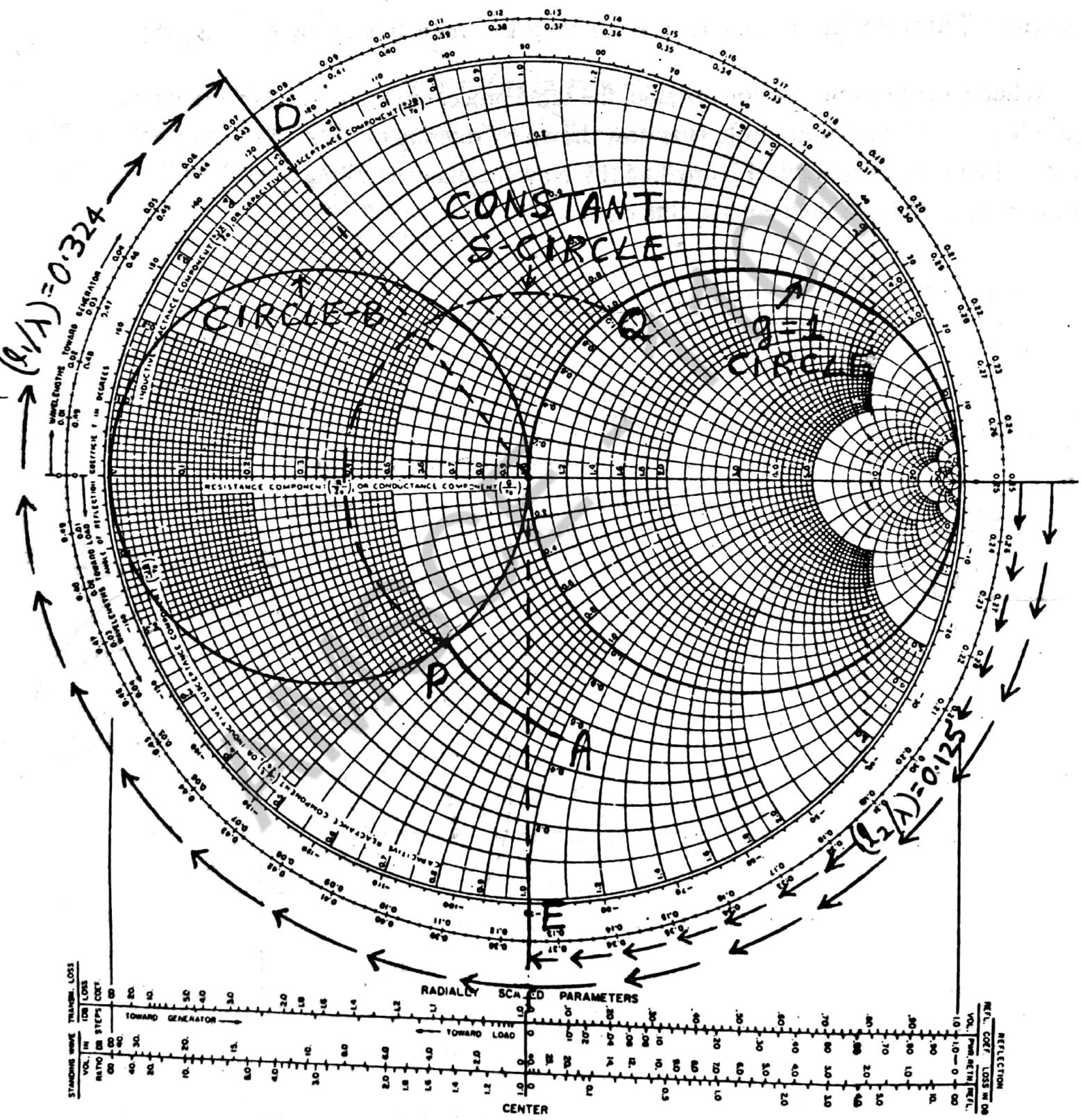
Fig. 3.15: Chart for Double Stub Matching

2. Stub-1 adds a susceptance in parallel and this must change y_R to a value $y_{11'}$ which lies on the locus circle B. Stub-1 cannot alter the conductance. Hence the constant conductance path from point A locates point P on circle B. At point P, $y_{11'} = 0.5 - j0.5$.

Thus stub-1 must contribute a susceptance of $b_{11'} = +0.5$. This value of capacitive susceptance is plotted as point D on the chart. The length l_1 of the stub-1 is the distance of point D from the short circuit point C in clockwise direction.

∴ From the chart, $l_1 = (0.25 + 0.074) \lambda$ or $l_1 = 0.324 \lambda$.

3. If y_R has a conductance greater than 1, a constant conductance circle from y_R plot would miss circle B and point P cannot be established, which confirms the limitations on conductance.
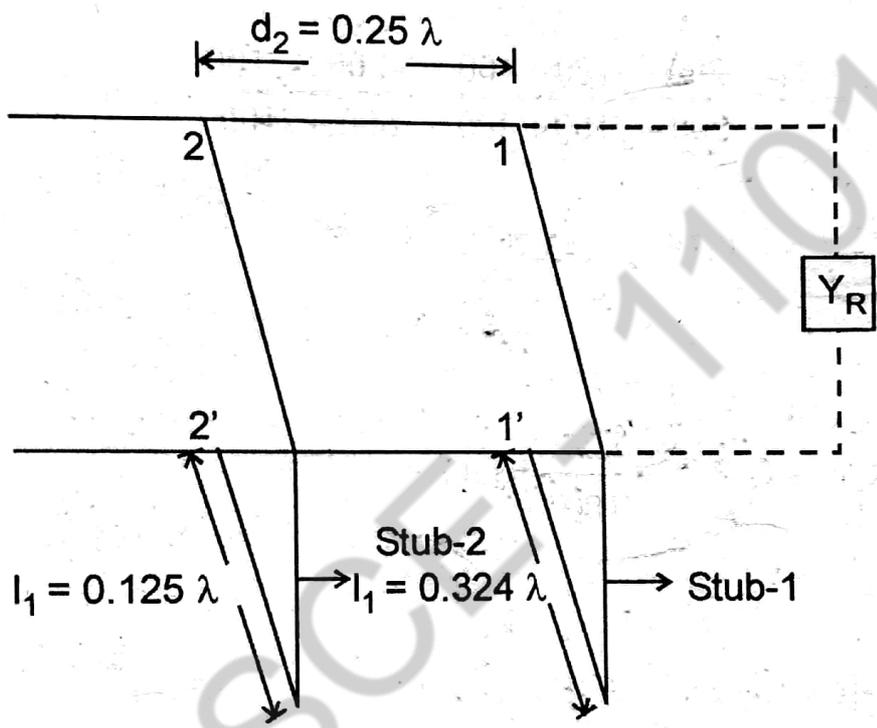
Fig. 3.16 : Illustration double-stub matching

4. The section of line between points 11' and 22' changes the admittance at 11' to that of 22' along a constant-S-circle. The admittance at 22' on the line, without stub-2 connected, can be read from the chart at point Q on $g = 1$ circle as shown in chart of Figure 3.16. At Q,

$$y_{22} = 1.0 + j\,1.0.$$

5. Stub-2 can be the adjusted to provide an inductive susceptance of $b_{22'} = -1.0$, after which the admittance at 22' will be $1 + j\,0$ and the main line to the left of 22' will be properly terminated.

6. The inductive susceptance of $b_{22'} = -1.0$ is plotted as point E on the chart of Figure. The length l_2 of the stub-2 is the distance of point E from the short-circuit point C in clock-wise direction. From the chart, l_2 is read as

$$l_2 = (0.375 - 0.25) \lambda \quad \text{or} \quad l_2 = 0.125 \lambda.$$

The complete configuration of double stub matched system is shown in Figure 3.16.

Chapter 6 provides a concise summary of the most important semiconductor fundamentals that are encountered at high frequencies.

By analyzing the *pn*-junction and the Schottky contact, we gain a more complete picture of electronic circuit functions that form the foundation of rectifier, amplifier, tuning, and switching systems. In particular, the metal-semiconductor interface is shown to be especially useful for high-frequency operations. It is the RF domain that has seen many specialized diode developments. Chief among them are the Schottky, *PIN*, and tunnel diode, to name but a few.

Next, our attention is turned toward the bipolar and field effect transistors, which are more complex implementations of the previously investigated *pn*-junction and Schottky contact. We learn about the construction, functionality, temperature, and noise performance of the bipolar and the metal-semiconductor field effect transistors.

6.1 Semiconductor Basics

6.1.1   Physical Properties of Semiconductors

The operation of semiconductor devices is naturally dependent on the physical behavior of the semiconductors themselves. This section presents a brief introduction to the basic building blocks of semiconductor device modeling, particularly the operation of the ***pn*-junction**. In our discussion we will concentrate on the three most commonly used semiconductors: germanium (Ge), silicon (Si), and gallium arsenide (GaAs). Figure 6-1(a) schematically shows the bonding structure of pure silicon: Each silicon atom shares its four valence electrons with the four neighboring atoms, forming four covalent bonds.

In the absence of thermal energy (i.e., when the temperature is equal to zero degree Kelvin [$T°K = 0$ or $T°C = -273.15$, where $T°K = 273.15 + T°C$]) all electrons are bonded to the corresponding atoms and the semiconductor is not conductive. However, when the temperature increases, some of the electrons obtain sufficient energy to break up the covalent bond and cross the energy gap $W_g = W_C - W_V$, as shown in Figure 6-1(b) (at room temperature $T = 300°K$ the **bandgap energy** is equal to 1.12 eV for Si, 0.62 eV for Ge, and 1.42 eV for GaAs). These free electrons form negative charge carriers that allow electric current conduction. The concentration of the conduction electrons in the semiconductor is denoted as n. When an electron breaks the covalent bond it leaves behind a positively charged vacancy, which can be occupied by another free electron. These types of vacancies are called **holes** and their concentration is denoted by p.

Electrons and holes undergo random motion through the semiconductor lattice as a result of the presence of thermal energy ($T > 0°K$). If an electron happens to meet a

(a) Planar representation of covalent bonds

(b) Energy band levels
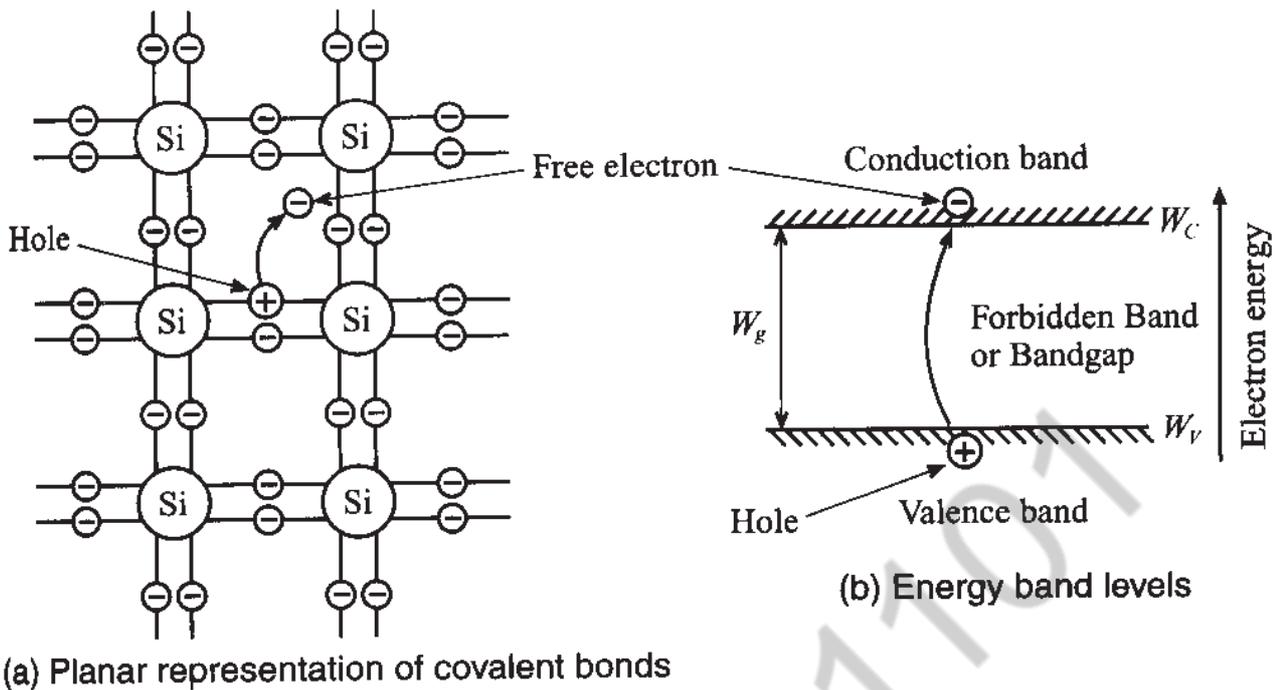
Figure 6-1 Lattice structure and energy levels of silicon.
(a) schematic planar crystal arrangement with thermal breakup of one valent bond resulting in a hole and a moving electron for T > 0°K.
(b) equivalent energy band level representation whereby a hole is created in the valence band W_V and an electron is produced in the conduction band W_C. The energy gap between both bands is indicated by W_g.

hole, they recombine and both charge carriers disappear. In thermal equilibrium we have equal number of recombinations and generations of holes and electrons. The concentrations obey the Fermi statistics according to

$$n = N_C \exp\left[-\frac{W_C - W_F}{kT}\right] \tag{6.1a}$$

$$p = N_V \exp\left[-\frac{W_F - W_V}{kT}\right] \tag{6.1b}$$

where

$$N_{C,V} = 2(2m^*_{n,p}\pi kT/h^2)^{3/2} \tag{6.2}$$

are the **effective carrier concentration** in the conduction (N_C) and valence (N_V) bands, respectively. The terms W_C and W_V denote the energy levels associated with the conduction and valence bands and W_F is the **Fermi energy level**, which indicates the energy level that has a 50% probability of being occupied by an electron. For **intrinsic** (i.e., pure) semiconductors at room temperature the Fermi level is very close

to the middle of the bandgap. In (6.2), m_n^* and m_p^* refer to the effective mass of electrons and holes in the semiconductor that are different from the free electron rest mass due to interaction with the crystal lattice; k is Boltzmann's constant; h is Planck's constant; and T is the absolute temperature measured in Kelvin.

In an intrinsic semiconductor the number of free electrons produced by thermal excitation is equal to the number of holes (i.e. $n = p = n_i$). Therefore, electron and hole concentrations are described by the concentration law

$$np = n_i^2 \tag{6.3}$$

where n_i is the intrinsic concentration. Equation (6.3) is true not only for intrinsic but also for doped semiconductors, which are discussed later in this section.

Substitution of (6.1) into (6.3) results in the expression for the intrinsic carrier concentration:

$$n_i = \sqrt{N_C N_V} \exp\left[-\frac{W_C - W_V}{2kT}\right] = \sqrt{N_C N_V} \exp\left[-\frac{W_g}{2kT}\right] \tag{6.4}$$

The effective electron and hole masses as well as the concentrations N_C, N_V, and n_i for $T = 300°\text{K}$ are summarized in Table 6-1 and are also listed in Table E-1 in Appendix E.

Table 6-1   Effective concentrations and effective mass values at $T = 300°\text{K}$

Semiconductor	m_n^*/m_0	m_p^*/m_0	$N_C(\text{cm}^{-3})$	$N_V(\text{cm}^{-3})$	$n_i(\text{cm}^{-3})$
Silicon (Si)	1.08	0.56	2.8×10^{19}	1.04×10^{19}	1.45×10^{10}
Germanium (Ge)	0.55	0.37	1.04×10^{19}	6.0×10^{18}	2.4×10^{13}
Gallium Arsenide (GaAs)	0.067	0.48	4.7×10^{17}	7.0×10^{18}	1.79×10^6

Classical electromagnetic theory specifies the electrical conductivity in a material to be $\sigma = J/E$, where J is the current density and E is the applied electric field. The conductivity in the classical model (Drude model) can be found through the carrier concentration N, the associated elementary charge q, the drift velocity v_d, and the applied electric field E:

$$\sigma = qNv_d/E \tag{6.5}$$

In semiconductors, we have both electrons and holes contributing to the conductivity of the material. At low electric fields the drift velocity v_d of the carriers is proportional to

the applied field strength through a proportionality constant known as **mobility** μ. Thus, for semiconductors we can rewrite (6.5) as

$$\sigma = qn\mu_n + qp\mu_p \qquad (6.6)$$

where μ_n, μ_p are the mobilities of electrons and holes, respectively. For intrinsic semi-conductors we can simplify (6.6) further by recalling that $n = p = n_i$, that is,

$$\sigma = qn_i(\mu_n + \mu_p) = q\sqrt{N_C N_V}\exp\left[-\frac{W_g}{2kT}\right](\mu_n + \mu_p) \qquad (6.7)$$
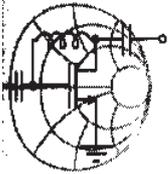
RF&MW

Example 6-1:   Computation of the temperature dependence of the intrinsic semiconductor conductivity

It is desired to find the conductivities for the intrinsic materials of Si, Ge, and GaAs as a function of temperature. To make the computations not too difficult, we assume that the bandgap energy and the mobilities for holes and electrons are temperature independent over the range of interest $-50°C \le T \le 200°C$.

Solution: As a first step it is convenient to combine into one parameter $\sigma_0(T)$ all factors without the exponential term in (6.7); that is,

$$\sigma_0(T) = q\sqrt{N_C N_V}(\mu_n + \mu_p)$$

where electron and hole mobilities are found from Table E-1:

$$\mu_n = 1350(\text{Si}), 3900(\text{Ge}), 8500(\text{GaAs})$$

$$\mu_p = 480(\text{Si}), 1900(\text{Ge}), 400(\text{GaAs})$$

All values are given in units of $\text{cm}^2/(\text{V}\cdot\text{s})$. N_C, N_V are computed according to (6.2) as

$$N_{C,V}(T) = N_{C,V}(300°\text{K})\left(\frac{T}{300}\right)^{3/2}$$

This leads to the form

$$\sigma = \sigma_0(T)\exp\left(-\frac{W_g}{2kT}\right) = q(\mu_n + \mu_p)\sqrt{N_C N_V}\left(\frac{T}{300}\right)^{3/2}\exp\left(-\frac{W_g}{2kT}\right)$$

where the bandgap energy $W_g = W_C - W_V$ is, respectively, 1.12 eV (Si), 0.62 eV (Ge), and 1.42 eV (GaAs). The three conductivities are plotted in Figure 6-2.
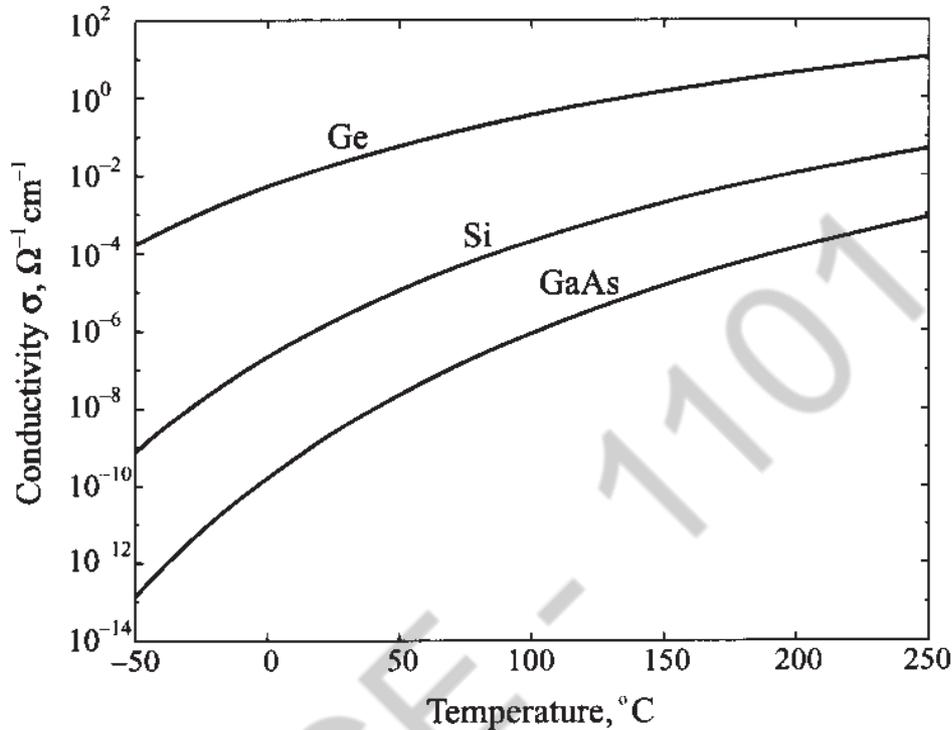
Figure 6-2   Conductivity of Si, Ge, GaAs in the range from –50°C to 250°C.

The electric properties of semiconductors are strongly influenced by the ambient temperature. In this example we have neglected the temperature dependence of the bandgap energy, which is discussed in Chapter 7. Knowledge of the temperature behavior of active devices is an important design consideration where internal heating, due to power dissipation, can easily result in temperature values exceeding 100–150°C.

A major change in the electrical properties of a semiconductor can be initiated by introducing impurity atoms. This process is called **doping**. To achieve **n-type** doping (which supplies additional electrons to the conduction band) we introduce atoms with a larger number of valence electrons than the atoms in the intrinsic semiconductor lattice that they substitute. For instance, the implantation of phosphorous (P) atoms into Si introduces loosely bound electrons into the neutral crystal lattice, as shown in Figure 6-3(b).
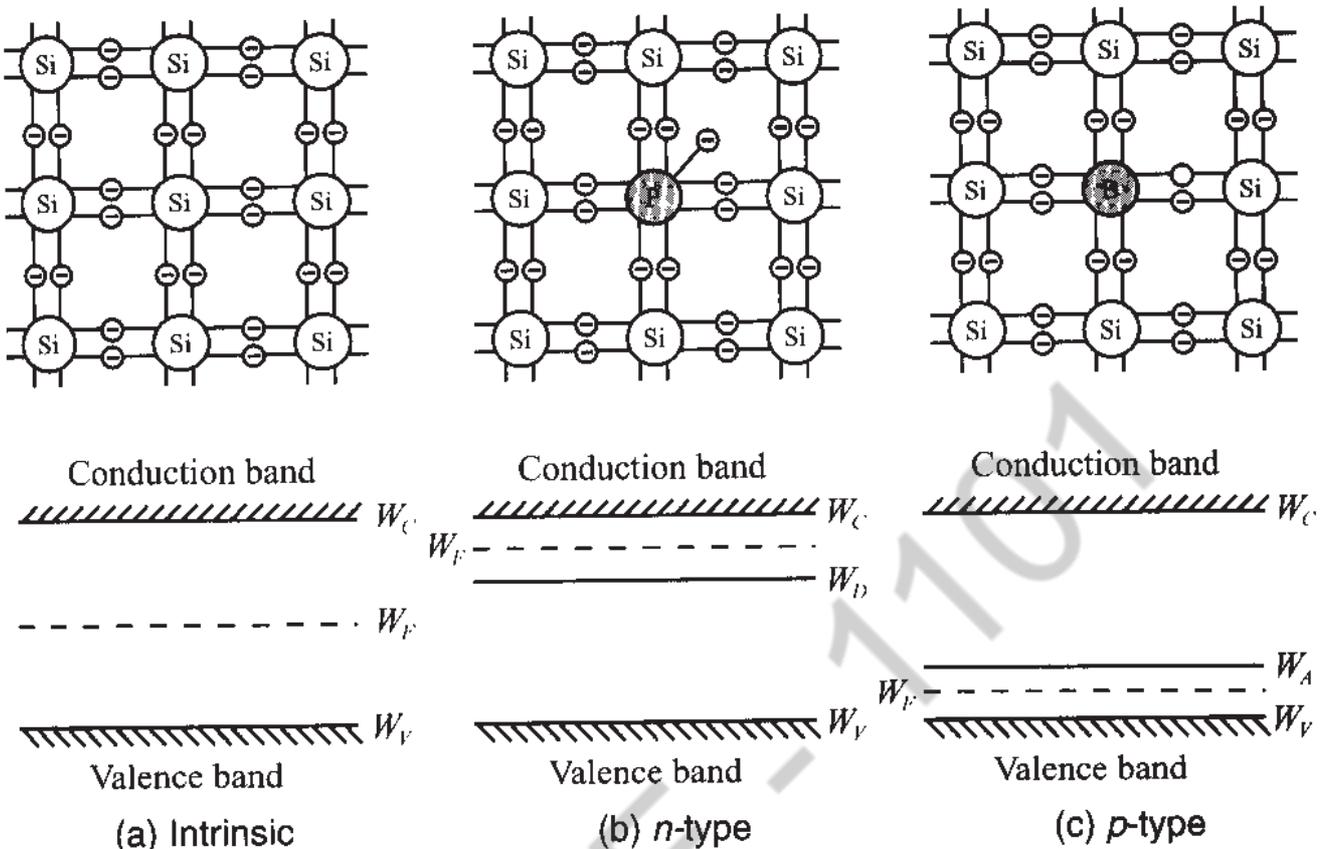
Figure 6-3 Lattice structure and energy band model for (a) intrinsic, (b) *n*-type, and (c) *p*-type semiconductors at no thermal energy. W_D and W_A are donor and acceptor energy levels.

It is intuitively apparent that the energy level of this "extra" electron is closer to the conduction band than the energy of the remaining four valence electrons. When the temperature is increased above absolute zero, the loosely bound electron separates from the atom, forming a free negative charge and leaving behind the fixed positive ion of phosphorous. Thus, while still maintaining charge neutrality, the atom has donated an electron to the conduction band without creating a hole in the valence band. This results in an increase in the Fermi level since more electrons are located in the conduction band. Contrary to the intrinsic semiconductor (n_i, p_i) we now have an **n-type** semiconductor in which the electron concentration is related to the hole concentration as

$$n_n = N_D + p_n \tag{6.8}$$

where N_D is the donor concentration and p_n represents the minority hole concentration. To find n_n and p_n we have to solve (6.8) in conjunction with (6.3). The result is

$$n_n = \frac{N_D + \sqrt{N_D^2 + 4n_i^2}}{2} \tag{6.9a}$$

$$p_n = \frac{-N_D + \sqrt{N_D^2 + 4n_i^2}}{2} \tag{6.9b}$$

If the donor concentration N_D is much greater than the intrinsic electron concentration n_i, then

$$n_n \approx N_D \tag{6.10a}$$

$$p_n \approx \frac{-N_D + N_D(1 + 2n_i^2/N_D^2)}{2} = \frac{n_i^2}{N_D} \tag{6.10b}$$

Let us now consider adding impurity atoms with fewer valence electrons than the atoms forming the intrinsic semiconductor lattice. These types of elements are called **acceptors**, and an example of such an element for the Si lattice is boron (B). As seen in Figure 6-3(c), one of the covalent bonds appears to be empty. This empty bond introduces additional energy states in the bandgap that are closely situated to the valence band. Again, when the temperature is increased from absolute zero, some electrons gain extra energy to occupy empty bonds but do not possess sufficient energy to cross the bandgap. Thus, impurity atoms will accept additional electrons, forming negative net charges. At the sites where the electrons are removed, holes will be created. These holes are free to migrate and will contribute to the conduction current of the semiconductor. By doping the semiconductor with acceptor atoms we have created a *p*-type semiconductor with

$$p_p = N_A + n_p \tag{6.11}$$

where N_A, n_p are the acceptor and minority electron concentrations. Solving (6.11) together with (6.3), we find hole p_p and electron n_p concentrations in the *p*-type semiconductor:

$$p_p = \frac{N_A + \sqrt{N_A^2 + 4n_i^2}}{2} \tag{6.12a}$$

$$n_p = \frac{-N_A + \sqrt{N_A^2 + 4n_i^2}}{2} \tag{6.12b}$$

Similar to (6.9), for high doping levels, when $N_A \gg n_i$, we observe

$$p_p \approx N_A \tag{6.13a}$$

$$n_p \approx \frac{-N_A + N_A(1 + 2n_i^2/N_A^2)}{2} = \frac{n_i^2}{N_A} \tag{6.13b}$$

Minority and majority concentrations play key roles in establishing the current flow characteristics in the semiconductor materials.

6.1.2 *PN*-Junction

The physical contact of a *p*-type with an *n*-type semiconductor leads to one of the most important concepts when dealing with active semiconductor devices: the ***pn*-junction.** Because of the difference in the carrier concentrations between the two types of semiconductors a current flow will be initiated across the interface. This current is commonly known as a **diffusion current** and is composed of electrons and holes. To simplify our discussion we consider a one-dimensional model of the *pn*-junction as seen in Figure 6-4.
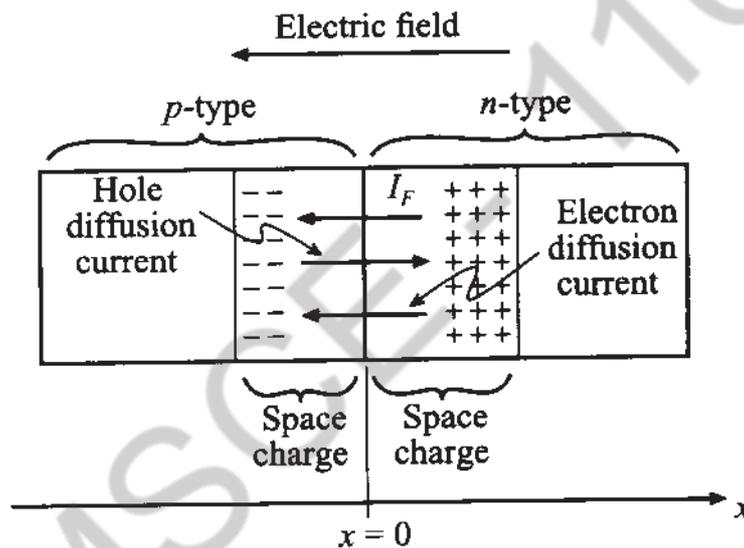
Figure 6-4   Current flow in the *pn*-junction

The diffusion current is composed of $I_{n_{\text{diff}}}$ and $I_{p_{\text{diff}}}$ components:

$$I_{\text{diff}} = I_{n_{\text{diff}}} + I_{p_{\text{diff}}} = qA\left(D_n\frac{dn}{dx} + D_p\frac{dp}{dx}\right) \tag{6.14}$$

where A is the semiconductor cross-sectional area orthogonal to the *x*-axis, and D_n, D_p are the diffusion constants for electrons and holes in the form (Einstein relation)

$$D_{n,\,p} = \mu_{n,\,p}\frac{kT}{q} = \mu_{n,\,p}V_T \tag{6.15}$$

The thermal potential $V_T = kT/q$ is approximately 26 mV at room temperature of 300°K.

Since the *p*-type semiconductor was initially neutral, the diffusion current of holes is going to leave behind a negative space charge. Similarly, the electron current flow

from the n-semiconductor will leave behind positive space charges. As the diffusion current flow takes place, an electric field E is created between the net positive charge in the n-semiconductor and the net negative charge in the p-semiconductor. This field in turn induces a current $I_F = \sigma AE$ which opposes the diffusion current such that $I_F + I_{\text{diff}} = 0$. Substituting (6.6) for the conductivity, we find

$$I_F = qA(n\mu_n + p\mu_p)E = I_{n_F} + I_{p_F} \tag{6.16}$$

Since the total current is equal to zero, the electron portion of the current is also equal to zero; that is,

$$I_{n_{\text{diff}}} + I_{n_F} = qD_n A \frac{dn}{dx} + qn\mu_n AE = q\mu_n A\left(V_T \frac{dn}{dx} - n\frac{dV}{dx}\right) = 0 \tag{6.17}$$

where the electric field E has been replaced by the derivative of the potential $E = -dV/dx$. Integrating (6.17), we obtain the **diffusion barrier voltage** or, as it is often called, the **built-in potential**:

$$\int_0^{V_{\text{diff}}} dV = V_{\text{diff}} = V_T \int_{n_p}^{n_n} n^{-1} dn = V_T \ln\left(\frac{n_n}{n_p}\right) \tag{6.18}$$

where again n_n is the electron concentration in the n-type and n_p is the electron concentration in the p-type semiconductor. The same diffusion barrier voltage could have been found had we considered the hole current flow from the p to the n-semiconductor and the corresponding balancing field-induced current flow I_{p_F}. The resulting equation describing the barrier voltage is

$$V_{\text{diff}} = V_T \ln\left(\frac{p_p}{p_n}\right) \tag{6.19}$$

If the concentration of acceptors in the p-semiconductor is $N_A \gg n_i$ and the concentration of donors in the n-semiconductor is $N_D \gg n_i$, then $n_n = N_D$, $n_p = n_i^2/N_A$, see (6.13b), and by using (6.18) we obtain

$$V_{\text{diff}} = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{6.20}$$

Exactly the same result is obtained from (6.19) if we substitute $p_p = N_A$ and $p_n = n_i^2/N_D$.
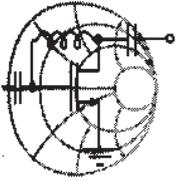
Example 6-2: Determining the diffusion barrier or built-in voltage of a *pn*-junction

For a particular Si *pn*-junction the doping concentrations are given to be $N_A = 10^{18} cm^{-3}$ and $N_D = 5 \times 10^{15} cm^{-3}$ with an intrinsic concentration of $n_i = 1.5 \times 10^{10} cm^{-3}$. Find the barrier voltages for $T = 300°K$.

Solution: The barrier voltage is directly determined from (6.20):

$$V_{\text{diff}} = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) = \frac{kT}{q}\ln\left(\frac{N_A N_D}{n_i^2}\right) = 0.796 \text{ (V)}$$

We note that the built-in potential is strongly dependent on the doping concentrations and temperature.

For different semiconductor materials such as GaAs, Si, Ge, the built-in voltage will be different even if the doping densities are the same. This is due to significantly different intrinsic carrier concentrations.

If we want to determine the potential distribution along the *x*-axis, we can employ Poisson's equation, which for a one-dimensional analysis is written as

$$\frac{d^2V(x)}{dx^2} = -\frac{\rho(x)}{\varepsilon_r \varepsilon_0} = -\frac{dE}{dx} \qquad (6.21)$$

where $\rho(x)$ is the charge density and ε_r is the relative dielectric constant of the semiconductor. Assuming uniform doping and the **abrupt junction approximation**, as shown in Figure 6-5(b), the charge density in each material is

$$\rho(x) = -qN_A, \quad \text{for } -d_p \le x \le 0 \qquad (6.22a)$$

$$\rho(x) = qN_D, \quad \text{for } 0 \le x \le d_n \qquad (6.22b)$$

where d_p and d_n are the extents of the space charges in the *p*- and *n*-type semiconductors.
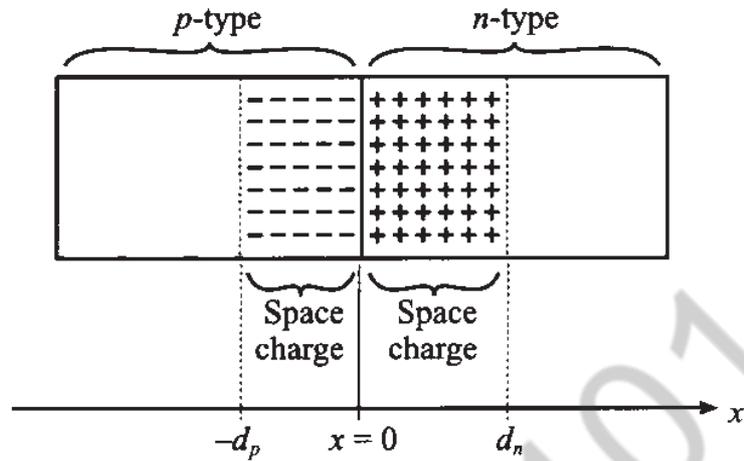
(a) *pn*-junction with space charge extent
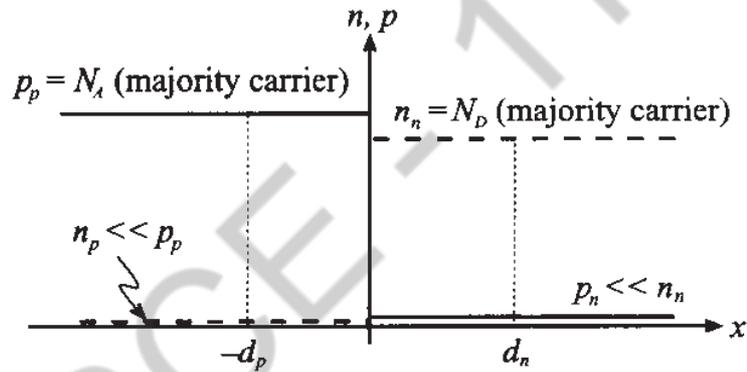
(b) Acceptor and donor concentrations
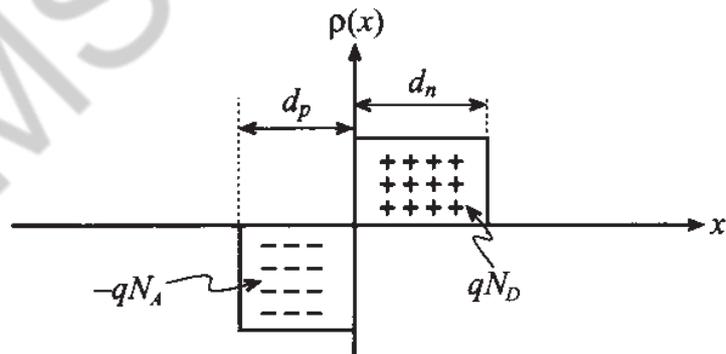
(c) Polarity of charge density distribution

Figure 6-5   The *pn*-junction with abrupt charge carrier transition in the absence of an externally applied voltage.
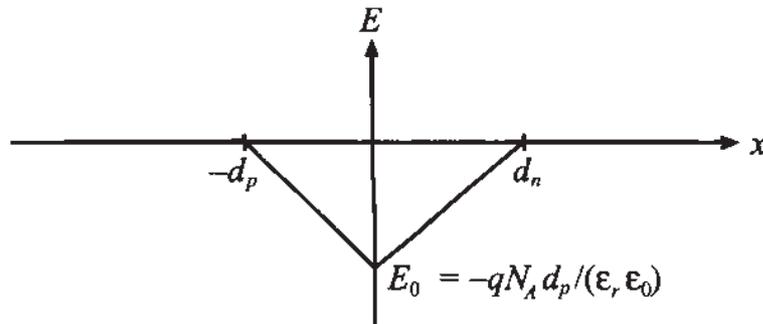
(d) Electric field distribution
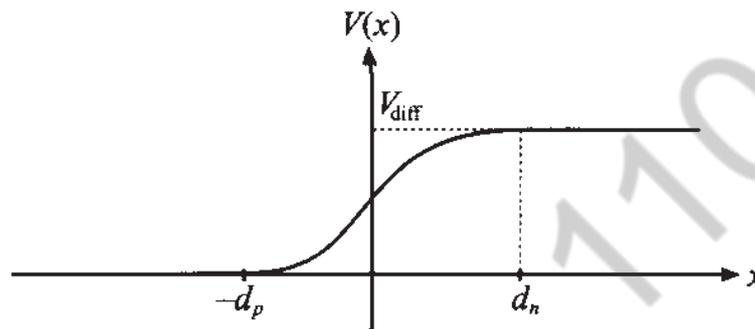
(e) Barrier voltage distribution

Figure 6-5   The *pn*-junction with abrupt charge carrier transition in the absence of an externally applied voltage. (Continued)

The electric field in the semiconductor is found by integrating (6.21) in the spatial limits $-d_p \leq x \leq d_n$ such that

$$E(x) = \int_{-d_p}^{x} \frac{\rho(x)}{\varepsilon_r \varepsilon_0} dx = \begin{cases} -\dfrac{qN_A}{\varepsilon_r \varepsilon_0}(x + d_p), & \text{for } -d_p \leq x \leq 0 \\[2mm] -\dfrac{qN_D}{\varepsilon_r \varepsilon_0}(d_n - x), & \text{for } 0 \leq x \leq d_n \end{cases} \tag{6.23}$$

The resulting electric field profile is depicted in Figure 6-5(d). In deriving (6.23) we used the fact that the charge balance law demands that the total space charge in the semiconductor equals zero, which for highly doped semiconductors is equivalent to the condition

$$N_A \cdot d_p = N_D \cdot d_n \tag{6.24}$$

To obtain the voltage distribution profile we now carry out the integration of (6.23) as follows:

$$V(x) = -\int_{-d_p}^{x} E(x)dx = \begin{cases} \dfrac{qN_A}{2\varepsilon_r\varepsilon_0}(x+d_p)^2, & \text{for } -d_p \leq x \leq 0 \\[3mm] \dfrac{q}{2\varepsilon_r\varepsilon_0}(N_A d_p^2 + N_D d_n^2) - \dfrac{qN_D}{2\varepsilon_r\varepsilon_0}(d_n - x)^2, & \text{for } 0 \leq x \leq d_n \end{cases} \tag{6.25}$$

Since the total voltage drop must be equal to the diffusion voltage V_{diff}, it is found that

$$V(d_n) = V_{\text{diff}} = \frac{qN_A d_p^2}{2\varepsilon_r\varepsilon_0} + \frac{qN_D d_n^2}{2\varepsilon_r\varepsilon_0} \tag{6.26}$$

Substituting $d_p = d_n N_D/N_A$ and solving (6.26) for d_n, we obtain the extent of the positive space charge domain into the n-semiconductor:

$$d_n = \left\{ \frac{2\varepsilon V_{\text{diff}}}{q}\frac{N_A}{N_D}\left(\frac{1}{N_A + N_D}\right) \right\}^{1/2} \tag{6.27}$$

where $\varepsilon = \varepsilon_0\varepsilon_r$. An identical derivation involving $d_n = d_p N_A/N_D$ gives us the space charge extent into the p-semiconductor:

$$d_p = \left\{ \frac{2\varepsilon V_{\text{diff}}}{q}\frac{N_D}{N_A}\left(\frac{1}{N_A + N_D}\right) \right\}^{1/2} \tag{6.28}$$

The entire length is then the addition of (6.27) and (6.28):

$$d_S = \left\{ \frac{2\varepsilon V_{\text{diff}}}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right) \right\}^{1/2} \tag{6.29}$$

We next turn our attention to the computation of the **junction capacitance**. This is an important parameter for RF devices, since low capacitances imply rapid switching speeds and suitability for high-frequency operations. The junction capacitance can be found via the well-known one-dimensional capacitor formula

$$C = \frac{\varepsilon A}{d_S}$$

Substituting (6.29) for the distance d_S, we express the capacitance as

$$C = A\left\{ \frac{q\varepsilon}{2V_{\text{diff}}}\frac{N_A N_D}{N_A + N_D} \right\}^{1/2} \tag{6.30}$$

If an external voltage V_A is applied across the junction, two situations arise that explain the rectifier action of the diode, as shown in Figure 6-6.

Space charge distribution in the *pn*-junction

Electric field distribution in the *pn*-junction

Voltage distribution in the *pn*-junction

(a) Reverse biasing ($V_A < 0$)   (b) Forward biasing ($V_A > 0$)
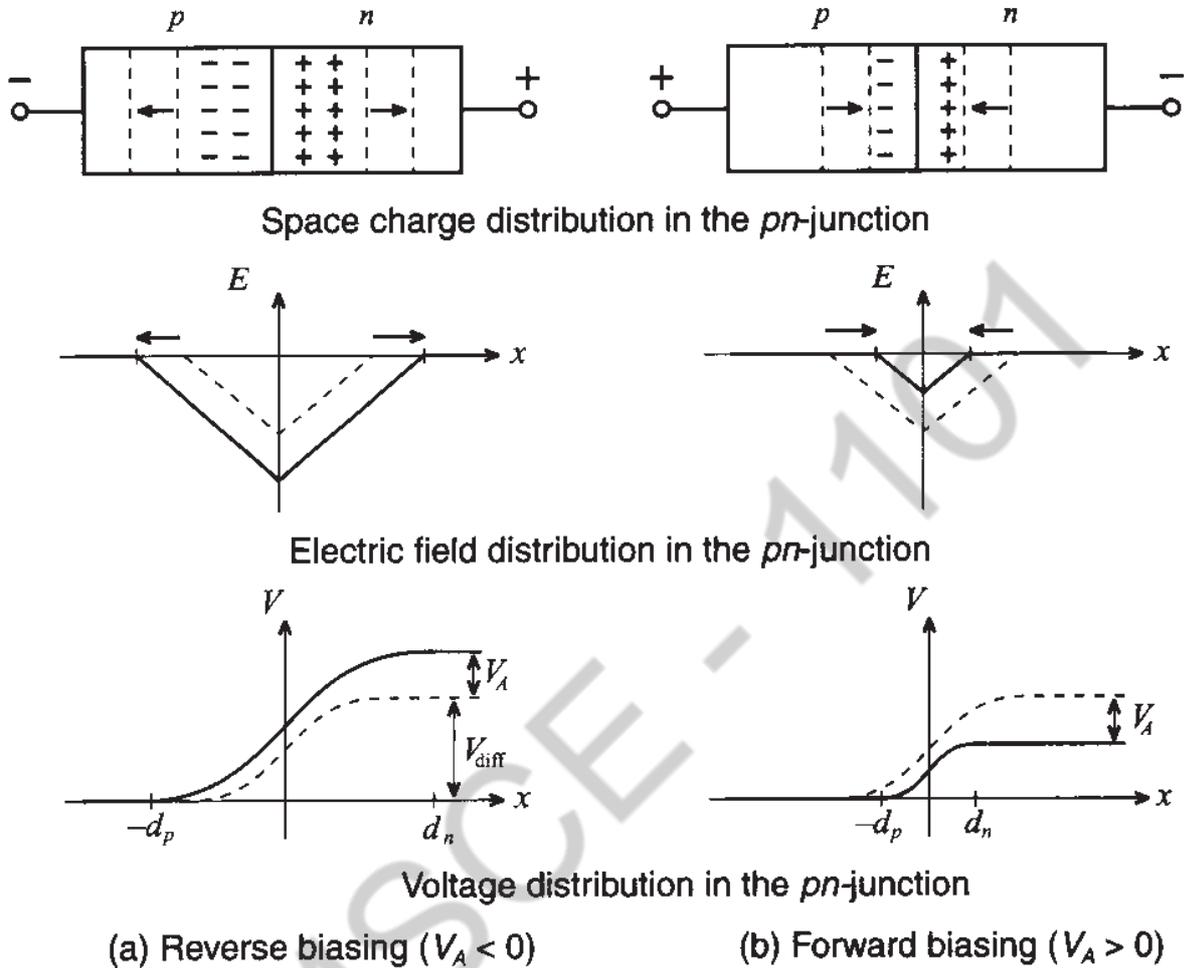
Figure 6-6   External voltage applied to the *pn*-junction in reverse and forward directions.

The reverse polarity [Figure 6-6(a)] increases the space charge domain and prevents the flow of current except for a small leakage current involving the minority carrier concentration (holes in the *n*-semiconductor, and electrons in the *p*-semiconductor). In contrast, the forward polarity reduces the space charge domain by injecting excessive electrons into the *n*- and holes into the *p*-type semiconductor. To describe these situations, the previously given equations (6.27) and (6.28) have to be modified by replacing the barrier voltage V_{diff} with $V_{\text{diff}} - V_A$; that is,

$$d_p = \left\{ \frac{2\varepsilon(V_{\text{diff}} - V_A)N_D}{q} \frac{1}{N_A}\left(\frac{1}{N_A + N_D}\right) \right\}^{\frac{1}{2}} \qquad (6.31)$$

$$d_n = \left\{\frac{2\varepsilon(V_{\text{diff}} - V_A)}{q}\frac{N_A}{N_D}\left(\frac{1}{N_A + N_D}\right)\right\}^{\frac{1}{2}} \tag{6.32}$$

which leads to a total length of the space charge or depletion domain

$$d_S = \left\{\frac{2\varepsilon(V_{\text{diff}} - V_A)}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right\}^{\frac{1}{2}} \tag{6.33}$$

Depending on the polarity of V_A, we notice from (6.31)–(6.33) that either the space charge domain is enlarged or diminished.

<hr>
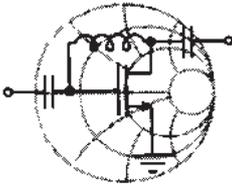
RF&MW

Example 6-3: **Computation of the junction capacitance and the space charge region length of a pn-junction**

For an abrupt pn-junction Si semiconductor at room temperature ($\varepsilon_r = 11.9$, $n_i = 1.5 \times 10^{10}$ cm^{-3}) with donor and acceptor concentrations equal to $N_D = 5 \times 10^{15}$ cm^{-3} and $N_A = 10^{15}$ cm^{-3}, we desire to find the space charge regions d_p and d_n and the junction capacitance at zero biasing voltage. Show that the depletion-layer capacitance of a pn-junction can be cast into the form

$$C_J = C_{J0}\left(1 - \frac{V_A}{V_{\text{diff}}}\right)^{-1/2}$$

and determine C_{J0}. Sketch the depletion capacitance as a function of applied voltage. Assume that the cross-sectional area of the pn-junction is $A = 10^{-4}$ cm^2.

Solution: We return to the capacitance expression (6.30) where we introduce the applied voltage V_A. Thus,

$$C_J = A\left[\frac{q\varepsilon}{2V_{\text{diff}}(1 - V_A/V_{\text{diff}})}\frac{N_A N_D}{N_A + N_D}\right]^{1/2}$$

which is immediately recognized as the preceding formula, if we set

$$C_{J0} = A\left[\frac{q\varepsilon}{2V_{\text{diff}}}\frac{N_A N_D}{N_A + N_D}\right]^{1/2}$$

Substituting $V_{\text{diff}} = V_T \ln(N_A N_D / n_i^2) = 0.6159\,\text{V}$, it is found that $C_{J0} = 10.68\,\text{pF}$.

For the space charge extents we use (6.28) and (6.29):

$$d_n = \left\{\frac{2\varepsilon V_{\text{diff}}}{q}\frac{N_A}{N_D}\left(\frac{1}{N_A + N_D}\right)\right\}^{1/2} = 0.1643\,\mu\text{m}$$

$$d_p = \left\{\frac{2\varepsilon V_{\text{diff}}}{q}\frac{N_D}{N_A}\left(\frac{1}{N_A + N_D}\right)\right\}^{1/2} = 0.8214\,\mu\text{m}$$

The dependence of the junction capacitance on the applied voltage is depicted in Figure 6-7.
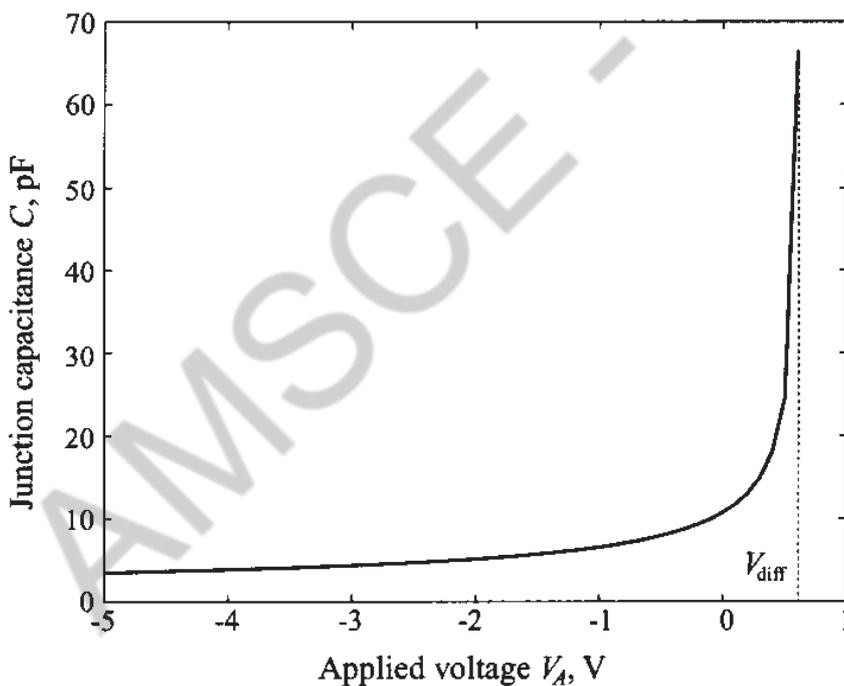
Figure 6-7 The *pn*-junction capacitance as a function of applied voltage.

In Figure 6-7 the junction capacitance for applied voltages near the built-in potential will approach infinity. However, in reality the value begins to saturate, as further discussed in Chapter 7.

For the current flow through the diode we list the Shockley diode equation, which is derived in Appendix F:

$$I = I_0(e^{V_A/V_T} - 1) \tag{6.34}$$

where I_0 is the **reverse saturation** or **leakage current**. The current-voltage characteristic, often called the *I-V* **curve**, is generically depicted in Figure 6-8.
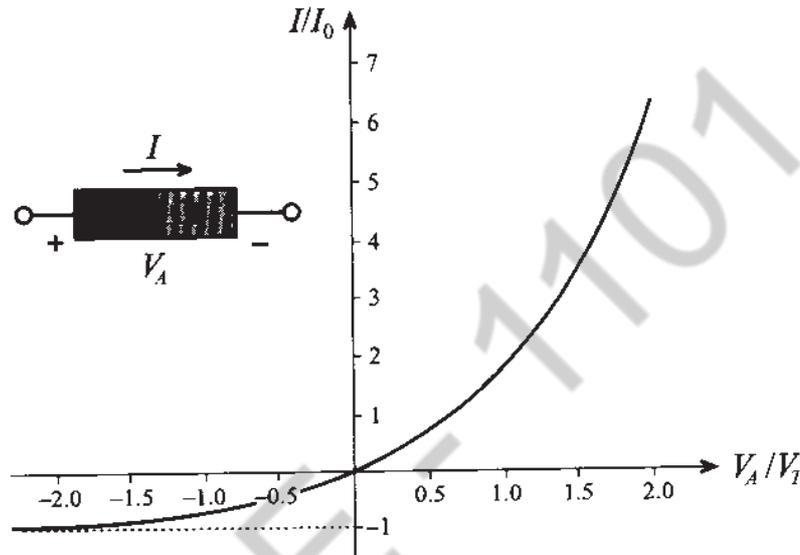
Figure 6-8   Current-voltage behavior of *pn*-junction based on Shockley equation.

This curve reveals that for negative voltages a small, voltage independent, current $-I_0$ will flow, whereas for positive voltages an exponentially increasing current is observed. The function shown in Figure 6-8 is an idealization since it does not take into account breakdown phenomena. Nonetheless (6.34) reveals clearly the rectifier property of the *pn*-junction when an alternating voltage is applied.

The existence of the depletion layer or junction capacitance requires a reverse-biased *pn*-junction diode. This implies, with reference to Example 6-3, the condition that $V_A < V_{\text{diff}}$. However, under forward bias condition we encounter an additional **diffusion capacitance** due to the presence of diffusion charges Q_d (minority carriers) stored in the semiconductor layers which become dominant if $V_A > V_{\text{diff}}$. This charge can be quantified by realizing that the charge Q_d can be computed as diode current I multiplied by the transition time of carriers through the diode τ_T or

$$Q_d = I\tau_T = \tau_T I_0(e^{V_A/V_T} - 1) \tag{6.35}$$

It is apparent that the diffusion capacitance assumes a nonlinear relation with the applied voltage and the junction temperature. The diffusion capacitance is computed as

$$C_d = \frac{dQ_d}{dV_A} = \frac{I_0 \tau_T}{V_T} e^{V_A/V_T} \qquad (6.36)$$

and is seen to be strongly dependent on the operating voltage.

In general, the total capacitance C of a pn-diode can roughly be divided into three regions:

1. $V_A < 0$: only the depletion capacitance is significant: $C = C_J$
2. $0 < V_A < V_{\text{diff}}$: depletion and diffusion capacitances combine: $C = C_J + C_d$
3. $V_A > V_{\text{diff}}$: only the diffusion capacitance is significant: $C = C_d$

The influence of the diffusion capacitance is appreciated if we consider a diode that is operated at $V_A = 1$ V and that has an assumed transition time of $\tau_T = 100$ ps $= 10^{-10}$ s and a reverse saturation current of $I_0 = 1$ fA $= 10^{-15}$ A measured at room temperature of 300°K (i.e., $V_T = 26$ mV). Substituting these values into (6.36), we find $C = C_d = 194$ nF which is rather large and for typical resistances of $R = 0.1 \ldots 1$ Ω results in large RC time constants that restrict the high-frequency use of conventional pn-junction diodes.

6.1.3   Schottky Contact

W. Schottky analyzed the physical phenomena involved when a metallic electrode is contacting a semiconductor. For instance, if a p-semiconductor is in contact with a copper or aluminum electrode, there is a tendency for the electrons to diffuse into the metal, leaving behind an increased concentration of holes in the semiconductor. The consequences of this effect are modified valence and conduction band energy levels near the interface. This can be displayed by a local change in the energy band structure depicted in Figure 6-9(a).
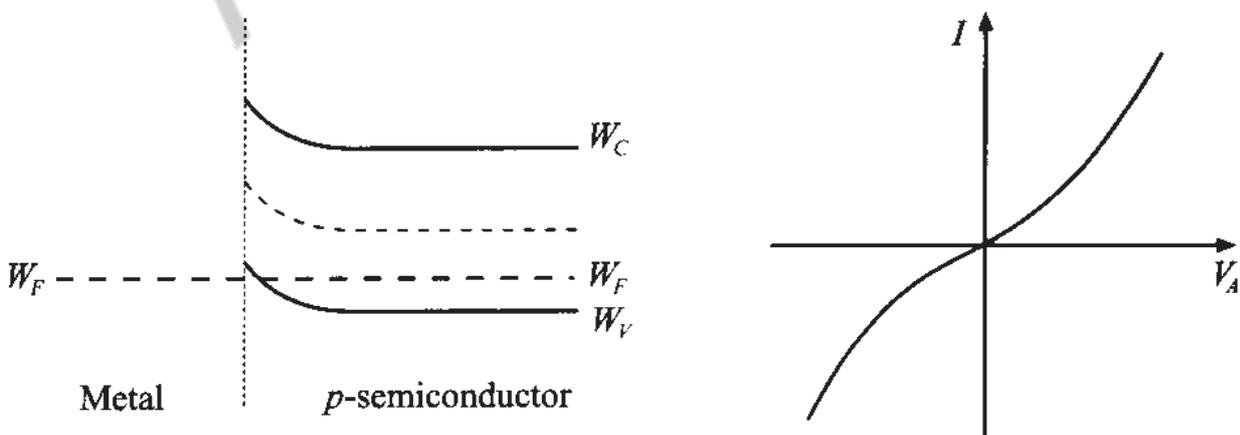
(a) Energy band model   (b) Voltage-current characteristic

Figure 6-9   Metal electrode in contact with p-semiconductor.

Because of the higher concentration of holes, the valence band bends toward the Fermi level. The conduction band, as the result of a lower electron concentration, bends away from the Fermi level. For such a configuration we always obtain a low resistance contact [(see Figure 6-9(b)], irrespective of the polarity of the applied voltage.

The situation becomes more complicated, but technologically much more interesting, when a metallic electrode is brought in contact with an n-semiconductor. Here the more familiar behavior of a pn-junction emerges: A small positive volume charge density is created in the semiconductor due to electron migration from the semiconductor to the metal. This mechanism is due to the fact that the Fermi level is higher in the semiconductor (lower work function) than in the metal (higher work function) when the two materials are apart. However, as both materials are contacted, the Fermi level again has to be the same and band distortions are created. Electrons diffuse from the n-semiconductor and leave behind positive space charges. The depletion zone grows until the electrostatic repulsion of the space charges prevents further electron diffusion. To clarify the issues associated with a metal n-semiconductor contact, Figure 6-10 shows the two materials before and after bonding.
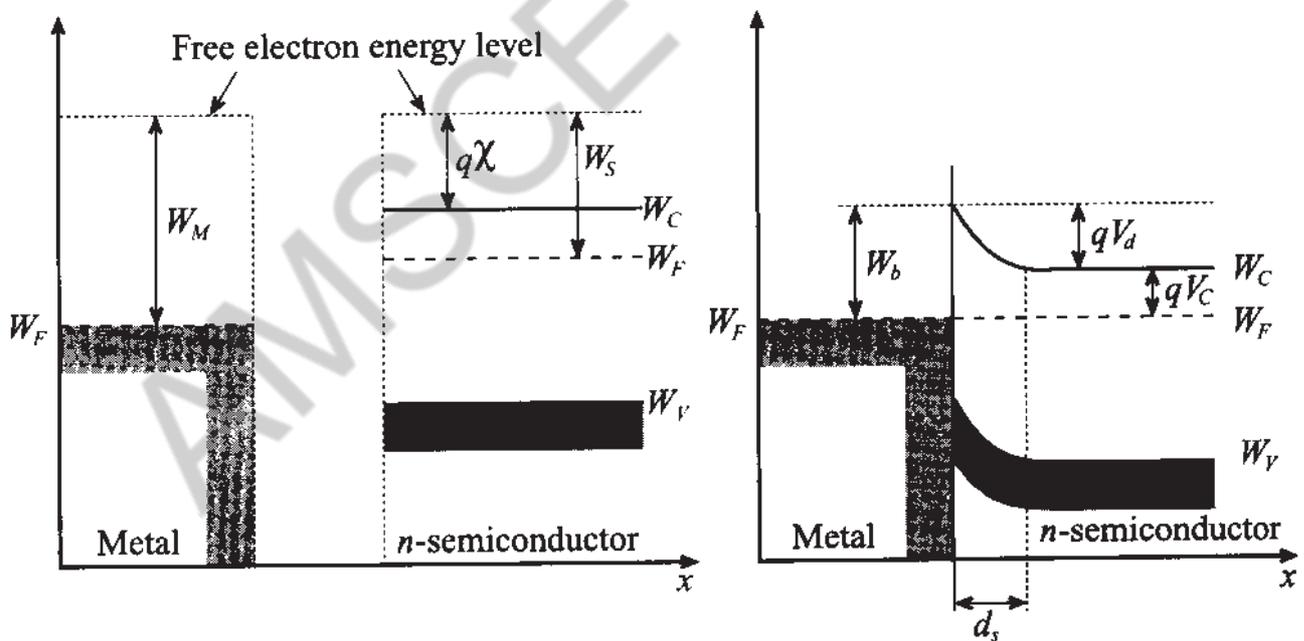
(a) Metal and semiconductor do not interact   (b) Metal-semiconductor contact

Figure 6-10   Energy-band diagram of Schottky contact, (a) before and (b) after contact.

The energy $W_b = qV_b$ is related to the metal work function $W_M = qV_M$ (V_M is recorded from the Fermi level to the reference level where the electron becomes a detached free particle; values of V_M for some commonly used metals are summarized in Table 6-2) and the electron affinity $q\chi$, where χ is 4.05V for Si, 4.0V for Ge, and

4.07V for GaAs and is measured from the conduction band to the same reference level where the electron becomes a free carrier, according to

$$W_b = q(V_M - \chi) \tag{6.37}$$

Table 6-2 Work function potentials of some metals

Material	Work function potential, V_M
Silver (Ag)	4.26 V
Aluminum (Al)	4.28 V
Gold (Au)	5.1 V
Chromium (Cr)	4.5 V
Molybdenum (Mo)	4.6 V
Nickel (Ni)	5.15 V
Palladium (Pd)	5.12 V
Platinum (Pt)	5.65 V
Titanium (Ti)	4.33 V

An expression for a built-in Schottky barrier voltage V_d is established just as in the *pn*-junction, which involves (6.37) and the additional voltage V_C between conduction and Fermi levels:

$$V_d = (V_M - \chi) - V_C \tag{6.38}$$

where V_C is dependent on the doping N_D and the concentration of states in the conduction band N_C according to $N_C = N_D \exp(V_C/V_T)$. Solving for the voltage gives $V_C = V_T \ln(N_C/N_D)$. Although real metal-semiconductor interfaces usually involve an additional very narrow isolation layer, we will neglect the influence of this layer and only deal with the length of the space charge in the semiconductor:

$$d_S = \left\{ \frac{2\varepsilon}{q} \frac{(V_d - V_A)}{N_D} \right\}^{\frac{1}{2}} \tag{6.39}$$

Therefore, it is found that the junction capacitance of the Schottky contact

$$C_J = A \frac{\varepsilon}{d_S} = A \left\{ \frac{q\varepsilon}{2(V_d - V_A)} N_D \right\}^{\frac{1}{2}} \tag{6.40}$$

is almost identical to (6.30). A simple computation now can predict a typical value for V_d as illustrated in the following example.
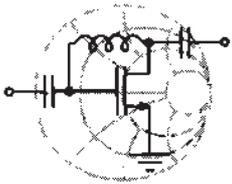
RF&MW

Example 6-4: Computation of the barrier voltage, depletion capacitance, and space charge region width for a Schottky diode

A Schottky diode is created as an interface between a gold contact material and an n-type silicon semiconductor. The semiconductor is doped to $N_D = 10^{16} \text{cm}^{-3}$ and the work function V_M for gold is 5.1 V. Also, as mentioned above, the affinity for Si is $\chi = 4.05 \text{V}$. Find the Schottky barrier V_d, space charge width d_S, and capacitance C_J if the dielectric constant of silicon is $\varepsilon_r = 11.9$. Assume the cross-sectional diode area to be $A = 10^{-4} \text{cm}^2$ and the temperature equal to $300°\text{K}$.

Solution: Since the concentration of states in the conduction band of silicon is $N_C = 2.8 \times 10^{19} \text{cm}^{-3}$, we can compute the conduction band potential as

$$V_C = V_T \ln\left(\frac{N_C}{N_D}\right) = \frac{1.38 \times 10^{-23} 300}{1.6 \times 10^{-19}} \ln\left(\frac{2.8 \times 10^{19}}{10^{16}}\right) \text{V} = 0.21 \text{V}$$

Substituting the obtained value for V_C into (6.38), we find the built-in barrier voltage

$$V_d = (V_M - \chi) - V_C = (5.1 \text{ V} - 4.05 \text{ V}) - 0.21 \text{ V} = 0.84 \text{ V}$$

The space charge width is obtained from (6.39)

$$d_S = \sqrt{\frac{2\varepsilon_0 \varepsilon_r}{q} \frac{V_d}{N_D}} = \sqrt{\frac{2(8.85 \times 10^{-12})11.9}{1.6 \times 10^{-19}} \frac{0.84}{10^{16}}} \text{ m} = 332 \text{ μm}$$

Finally, the junction capacitance according to the formula for the parallel-plate capacitor, see (6.40), gives us

6.3 Bipolar-Junction Transistor

The transistor was invented in 1948 by Bardeen and Brattain at the former AT&T Bell Laboratories and has over the past 50 years received a long lists of improvements and refinements. Initially developed as a point-contact, single device, the transistor has proliferated into a wide host of sophisticated types ranging from the still popular **bipolar junction transistors** (BJTs) over the modern **GaAs field effect transistors** (GaAs FETs) to the most recent **high electron mobility transistors** (HEMTs). Although transistors are often arranged in the millions in integrated circuits (ICs) as part of microprocessor, memory, and peripheral chips, in RF and MW applications the single transistor has retained its importance. Many RF circuits still rely on discrete transistors in low-noise, linear, and high-power configurations. It is for this reason that we need to investigate both the DC and RF behavior of the transistors in some detail.

The constituents of a bipolar transistor are three alternatively doped semiconductors, in *npn* or *pnp* configuration. As the word *bi*polar implies, the internal current flow is due to both minority and majority carriers. In the following we recapitulate some of the salient characteristics.

6.3.1 Construction

The BJT is one of the most widely used active RF elements due to its low-cost construction, relatively high operating frequency, low-noise performance, and high-power handling capacity. The high-power capacity is achieved through a special interdigital emitter-base construction as part of a planar structure. Figure 6-28 shows both the cross-sectional planar construction and the top view of an interdigital emitter-base connection.

Because of the interleaved construction shown in Figure 6-28(b) the base-emitter resistance is kept at a minimum while not compromising the gain performance. As we will see, a low base resistance directly improves the signal-to-noise ratio by reducing the current density through the base-emitter junction (shot noise) and by reducing the random thermal motion in the base (thermal noise), see Chapter 7 for more details.

For frequency applications exceeding 1 GHz it is important to reduce the emitter width to typically less than 1 μm size while increasing the doping to levels of $10^{20} \ldots 10^{21}$ cm^{-3} to both reduce base resistance and increase current gain. Unfortunately, it becomes extremely difficult to ensure the tight tolerances, and self-aligning processes are required. Furthermore, the acceptor and donor doping concentrations reach quickly the solubility limits of the Si or GaAs semiconductor materials, providing a physical limitation of the achievable current gain. For these reasons, **heterojunction bipolar transistors** (HBTs) are becoming increasingly popular. HBTs achieve high
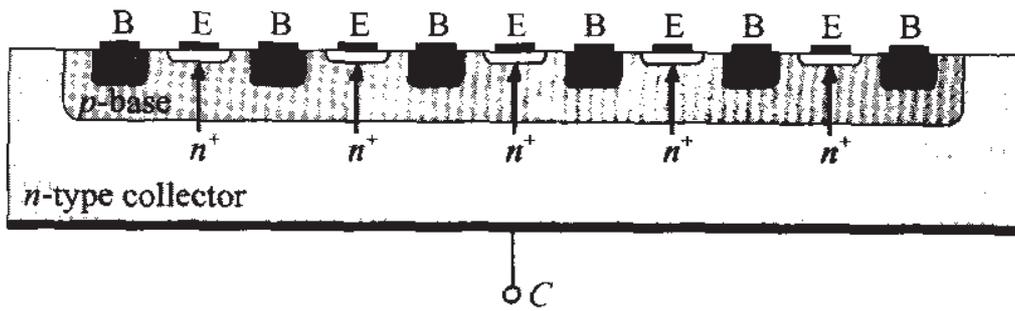
(a) Cross-sectional view of a multifinger bipolar junction transistor
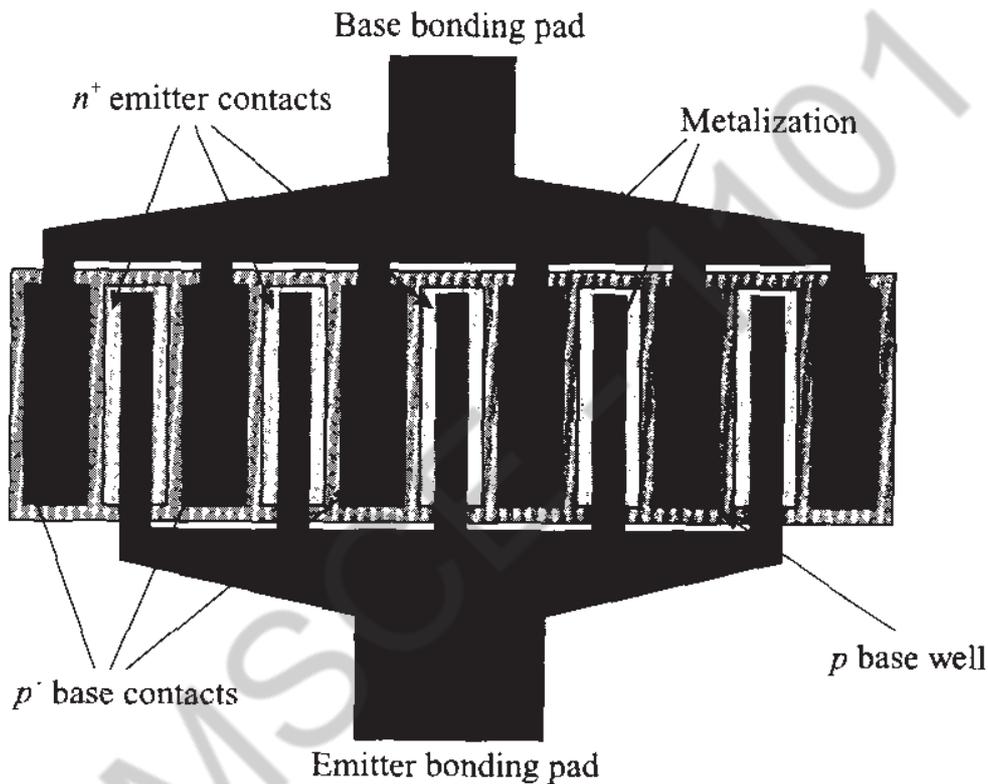
(b) Top view of a multifinger bipolar junction transistor

Figure 6-28   Interdigitated structure of high-frequency BJT.

current gains without having to dope the emitter excessively. Due to additional semiconductor layers (for instance, GaAlAs-GaAs sandwich structures) an enhanced electron injection into the base is achieved while the reverse hole injection into the emitter is suppressed. The result is an extremely high **emitter efficiency** as defined by the ratio of electron current into the base to the sum of the same electron current and reverse emitter hole current. Figure 6-29 shows a cross-sectional view of such a structure.

Besides GaAs, heterojunctions have been accomplished with InP emitter and InGaAs base interfaces; even additional heterojunction interfaces between the GaInAs base and InP collector (double heterojunctions) have been fabricated. The material InP has the advantage of high breakdown voltage, larger bandgap, and higher thermal conductivity compared to GaAs. Operational frequencies exceeding 100 GHz, and a carrier
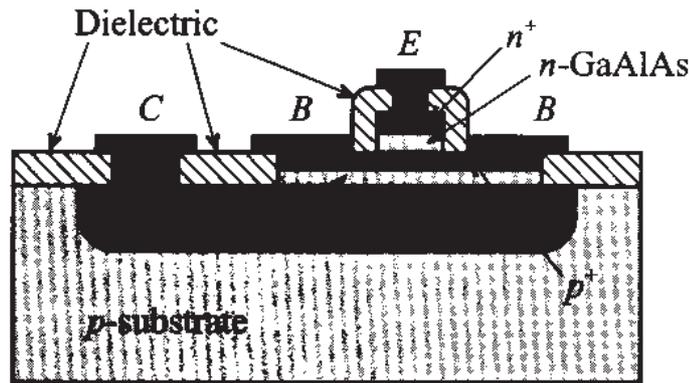
Figure 6-29   Cross-sectional view of a GaAs heterojunction bipolar transistor involving a GaAlAs-GaAs interface.

transition time between base and collector of less than 0.5 ps have been achieved. Unfortunately, InP is a difficult material to handle and the manufacturing process has not yet matured to a level that allows it to compete with the Si and GaAs technologies.

6.3.2   Functionality

In general, there are two types of BJTs: *npn* and *pnp* transistors. The difference between these two types lies in the doping of the semiconductor used to produce base, emitter, and collector. For an *npn*-transistor, collector and emitter are made of *n*-type semiconductor, while the base is of *p*-type. For a *pnp*-transistor, the semiconductor types are reversed (*n*-type for base, and *p*-type for emitter and collector). Usually, the emitter has the highest and the base has the lowest concentration of doping atoms. The BJT is a *current-controlled device* that is best explained by referring to Figure 6-30, which shows the structure, electrical symbol, and diode model with associated voltage and current convention for the *npn*-structure. We omit the discussion of the *pnp*-transistor since it requires only a reversal of voltage polarity and diode directions.

The first letter in the voltage designation always denotes the positive and the second letter gives the negative voltage reference points. Under normal mode of operation (i.e., the **forward active** mode), the emitter-base diode is operated in forward direction (with $V_{BE} \approx 0.7 \text{ V}$) and the base-collector diode in reverse. Thus the emitter injects electrons into the base, and conversely from the base a hole current reaches the emitter. If we maintain the collector emitter voltage to be larger than the so-called **saturation voltage** (typically around 0.1 V), and since the base is a very thin (on the order of $d_B \leq 1 \text{ } \mu\text{m}$) and lightly doped *p*-type layer, only a small amount of electrons recombine with the holes supplied through the base current. The vast majority of electrons reach the base-collector junction and are collected by the applied reverse voltage V_{BC}.

For the **reverse active mode**, the collector-emitter voltage is negative (typically $V_{CE} < -0.1 \text{ V}$) and the base-collector diode is forward biased, while the base-emitter
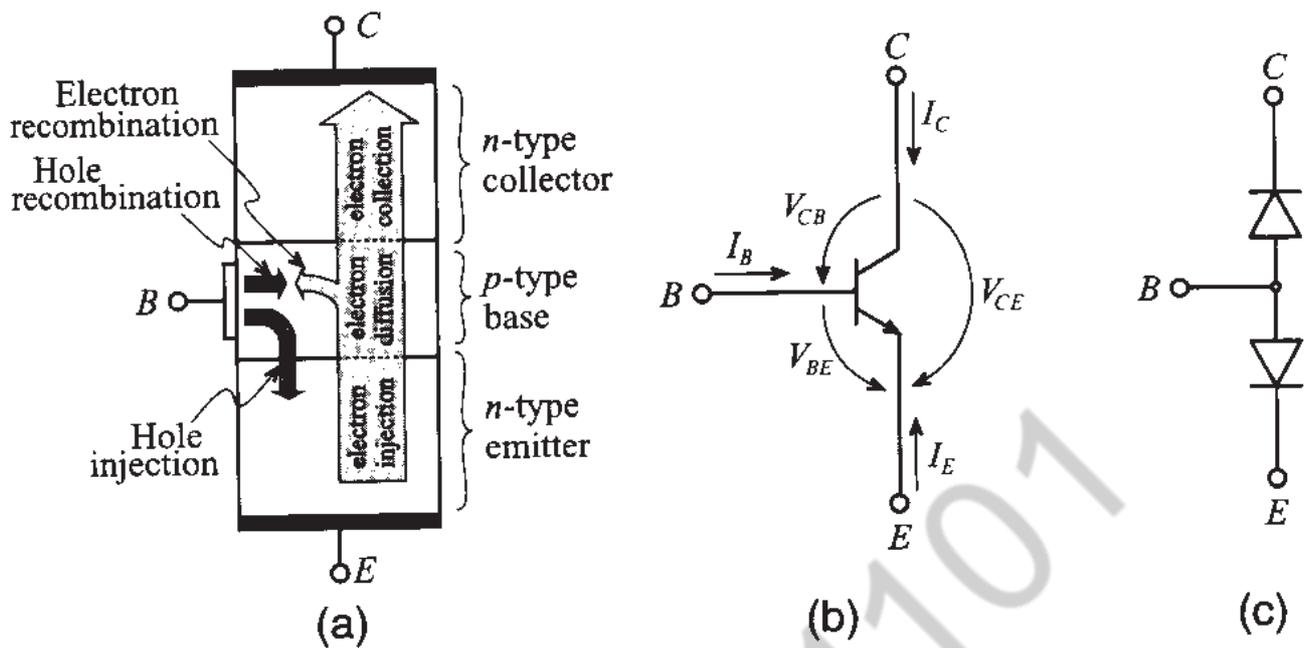
Figure 6-30 *npn* transistor: (a) structure with electrical charge flow under forward active mode of operation, (b) transistor symbol with voltage and current directions, and (c) diode model.

diode is now operated in reverse direction. Unlike the forward active mode, it is now the electron flow from the collector that bridges the base and reaches the emitter.

Finally, the **saturation mode** involves the forward biasing of both the base-emitter and base-collector junctions. This mode typically plays an important role when dealing with switching circuits.

For a common emitter configuration, Figure 6-31(a) depicts a generic biasing arrangement, where the base current is fixed through an appropriate choice of biasing resistor R_B and voltage source V_{BB}, resulting in a suitable Q-point. The base current versus base-emitter voltage, Figure 6-31(b), follows a typical diode I-V behavior, which constitutes the input characteristic of the transistor. The base current and base-emitter voltage at the intersection point between the load line and the transistor input characteristic are identified as I_B^Q and V_{BE}^Q. The collector current versus collector-emitter voltage behavior as part of the transistor output characteristic follows a more complicated pattern since the collector current must be treated as a parametric curve dependent on the base current ($I_{B1} < I_{B2} \ldots$) as seen in Figure 6-31(c).

The quantitative BJT behavior is analyzed by investigating the three modes of operation in terms of setting appropriate operating points and formulating the various current flows. For simplicity, we will neglect the spatial extent of the individual space charge domains and assume typical representative voltage and current conditions. To keep track of the different minority/majority and doping conditions in the three semi-conductor layers, Table 6-3 summarizes the parameters and corresponding notation.
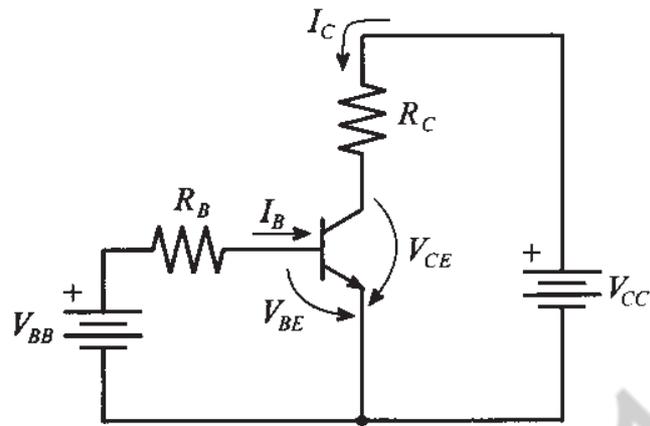
(a) Biasing circuit for *npn* BJT in common-emitter configuration
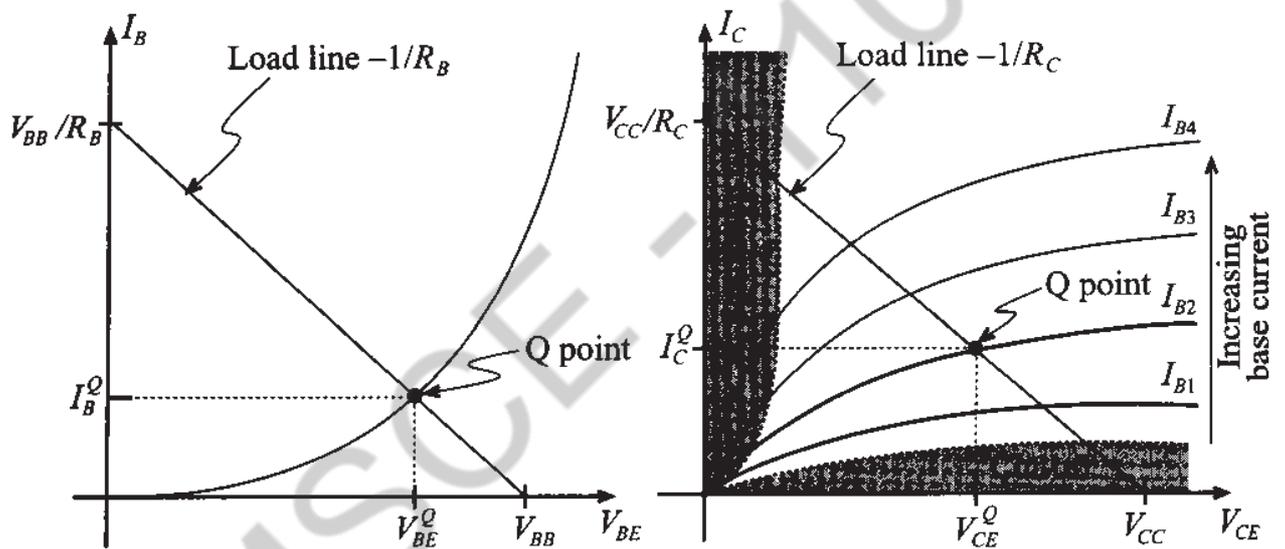
(b) Input characteristic of transistor

(c) Output characteristic of transistor

Figure 6-31   Biasing and input, output characteristics of an *npn* BJT.

Table 6-3   BJT parameter nomenclature

Parameter description	Emitter (*n*-type)	Base (*p*-type)	Collector (*n*-type)
Doping level	N_D^E	N_A^B	N_D^C
Minority carrier concentration in thermal equilibrium	$p_{n_0}^E = n_i^2/N_D^E$	$n_{p_0}^B = n_i^2/N_A^B$	$p_{n_0}^C = n_i^2/N_D^C$
Majority carrier concentration in thermal equilibrium	$n_{n_0}^E$	$p_{p_0}^B$	$n_{n_0}^C$
Spatial extent	d_E	d_B	d_C

For the following BJT analysis, it is implicitly understood that the concentrations obey the inequality $p_{n0}^E \ll n_{p0}^B \ll p_{n0}^C$.

Forward Active Mode ($V_{CE} > V_{CEsat} = 0.1$ V, $I_B > 0$)

To find the minority charge concentrations, we consider the configuration shown in Figure 6-32. Here the concentration is plotted as a function of distance across the three semiconductor layers. For predicting the spatial minority carrier concentrations in the respective layer, we rely on the so-called **short diode** (see Appendix F) analysis, which approximates the exponentials as linear charge concentration gradients.
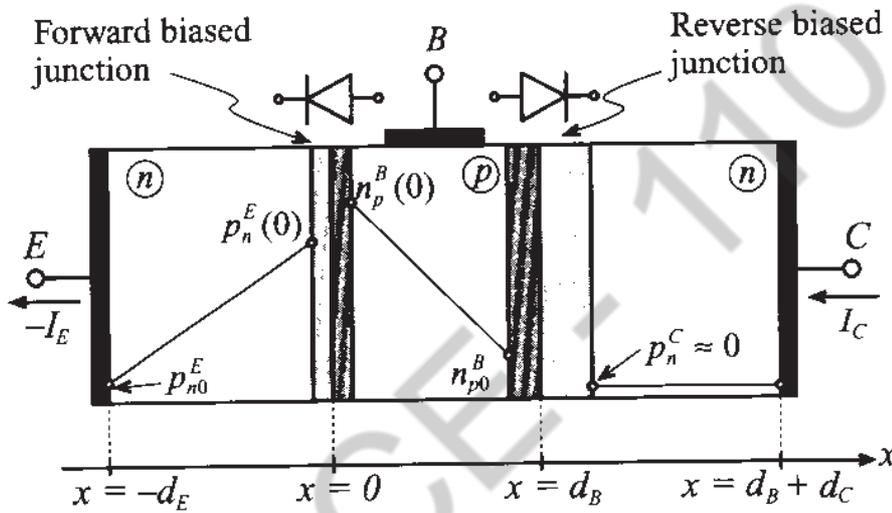
Figure 6-32 Minority carrier concentrations in forward active BJT.

The minority charge concentrations in each layer are given as follows:

- Emitter: $p_n^E(-d_E) = p_{n0}^E$ and $p_n^E(0) = p_{n0}^E e^{V_{BE}/V_T}$

- Base: $n_p^B(0) = n_{p0}^B e^{V_{BE}/V_T}$ and $n_p^B(d_B) = n_{p0}^B e^{V_{BC}/V_T} \approx 0$

- Collector: $p_n^C(d_B) = p_{n0}^C e^{V_{BC}/V_T} \approx 0$

The last two concentrations are zero because the base-collector voltage is negative (for instance, for typical transistor values of $V_{CE} = 2.5$ V and $V_{BE} = 0.7$ V we find $V_{BC} = -1.8$ V, which yields $\exp[V_{BC}/V_T] = \exp[-1.8/0.026] \to 0$). Based on the aforementioned carrier concentrations we can now predict the diffusion current density of holes J_{pdiff}^E in the emitter:

$$J_{pdiff}^E = -qD_p^E \frac{dp_n^E(x)}{dx} = -\frac{qD_p^E}{d_E}[p_n^E(0) - p_n^E(-d_E)] = -\frac{qD_p^E p_{n0}^E}{d_E}(e^{V_{BE}/V_T} - 1) \quad (6.58)$$

For the diffusion current density of electrons in the base layer $J_{n\,\text{diff}}^B$ we similarly obtain

$$J_{n\,\text{diff}}^B = qD_n^B \left[\frac{dn_p^B(x)}{dx} \right] = \frac{qD_n^B}{d_B} [n_p^B(d_B) - n_p^B(0)] = -\frac{qD_n^B n_{p0}^B}{d_B} e^{V_{BE}/V_T} \quad (6.59)$$

From the preceding two equations, the collector and base currents can be established as

$$I_{FC} = -J_{n\,\text{diff}}^B A = \frac{qD_n^B n_{p0}^B}{d_B} A e^{V_{BE}/V_T} = I_S e^{V_{BE}/V_T} \quad (6.60)$$

and

$$I_{FB} = -J_{p\,\text{diff}}^E A = \frac{qD_p^E p_{n0}^E}{d_E} A (e^{V_{BE}/V_T} - 1) \quad (6.61)$$

where index F denotes forward current, A is the junction cross-sectional area, and $I_S = (qD_n^B n_{p0}^B A)/d_B$ is the **saturation current**. The emitter current is directly found by adding (6.60) and (6.61). The forward current gain β_F under constant collector emitter voltage is defined as

$$\beta_F = \frac{I_{FC}}{I_{FB}}\bigg|_{V_{CE}} = \frac{D_n^B n_{p0}^B d_E}{D_p^E p_{n0}^E d_B} \quad (6.62)$$

To arrive at (6.62) it is assumed that the exponential function in (6.61) is much larger than 1, allowing us to neglect the factor −1. Moreover, the ratio between collector and emitter currents, or α_F, is expressed as

$$\alpha_F = \frac{I_{FC}}{(-I_{FE})} = \frac{\beta_F}{1 + \beta_F} \quad (6.63)$$

RF&MW
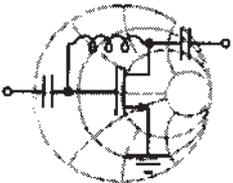
Example 6-7: Computation of the maximum forward current gain in a bipolar-junction transistor

Find the maximum forward current gain for a silicon-based BJT with the following parameters: donor concentration in the emitter, $N_D^E = 10^{19} \text{ cm}^{-3}$; acceptor concentration in the base, $N_A^B = 10^{17} \text{ cm}^{-3}$; space charge extent in the emitter, $d_E = 0.8 \ \mu\text{m}$; and space charge extent in the base, $d_B = 1.2 \ \mu\text{m}$.

Solution: To apply (6.62), we need to determine the diffusion constants in base and emitter as described by the Einstein relation (6.15). Substituting this relation into (6.62), we obtain the forward current gain:

$$\beta_F = \frac{\mu_n n_{p0}^B d_E}{\mu_p p_{n0}^E d_B}$$

Furthermore, using the expressions for the minority carrier concentrations in base and emitter from Table 6-3, we arrive at the final expression for β_F:

$$\beta_F = \frac{\mu_n N_D^E d_E}{\mu_p N_A^B d_B} = 187.5$$

As discussed in Section 6.3.3 and in the following chapter, the current gain is only approximately constant. In general, it depends on the transistor operating conditions and temperature behavior.

Reverse Active Mode ($V_{CE} < -0.1$ V, $I_B > 0$)

The minority carrier concentrations are shown in Figure 6-33 with the associated space charge domains (i.e., the base-emitter diode is reversed biased whereas the base-collector diode is forward biased).
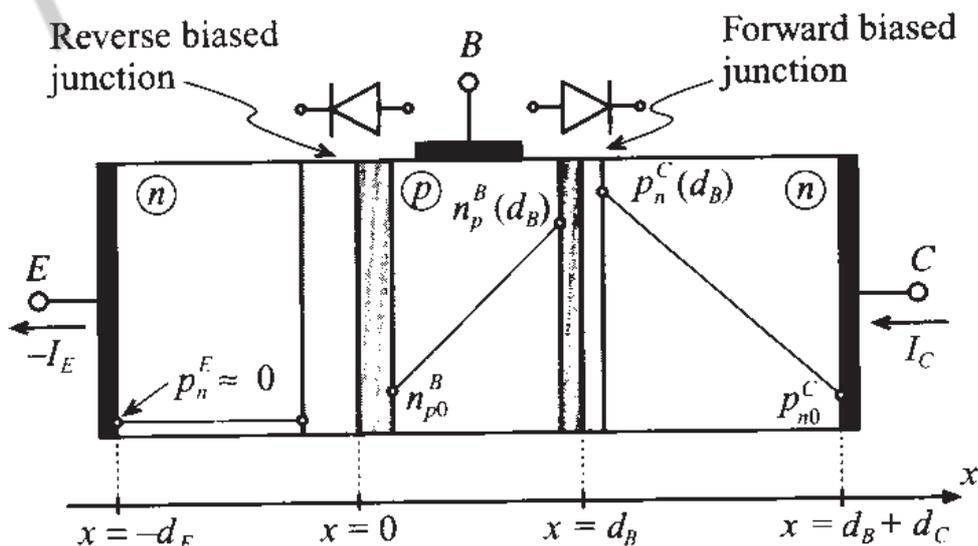
Figure 6-33 Reverse active mode of BJT.

The minority charge concentrations in each layer are as follows:

- **Emitter:** $p_n^E(-d_E) \approx 0$ and $p_n^E(0) = p_{n0}^E e^{V_{BE}/V_T} \approx 0$

- **Base:** $n_p^B(0) = n_{p0}^B e^{V_{BE}/V_T} \approx 0$ and $n_p^B(d_B) = n_{p0}^B e^{V_{BC}/V_T}$

- **Collector:** $p_n^C(d_B) = p_{n0}^C e^{V_{BC}/V_T}$ and $p_n^C(d_B + d_C) \approx p_{n0}^C$

From the diffusion current density, we can find the reverse emitter current

$$I_{RE} = -J_{n\text{diff}}^B A = -q D_n^B \left(\frac{dn_p^B}{dx} \right) A = \frac{q D_n^B n_{p0}^B}{d_B} A e^{V_{BC}/V_T} = I_S e^{V_{BC}/V_T} \tag{6.64}$$

and the reverse base current

$$I_{RB} = -J_{p\text{diff}}^C A = -q D_p^C \left(\frac{dp_n^C}{dx} \right) A = \frac{q D_p^C p_{n0}^C A}{d_C} \left(e^{V_{BC}/V_T} - 1 \right) \tag{6.65}$$

In a similar manner as done for the forward current gain, we define the **reverse current gain** β_R

$$\beta_R = \left. \frac{I_{RE}}{I_{RB}} \right|_{V_{BC}} = \frac{D_n^B n_{p0}^B d_C}{D_p^C p_{n0}^C d_B} \tag{6.66}$$

and the collector emitter ratio α_R

$$\alpha_R = \left. \frac{I_{RC}}{(-I_{RE})} \right|_{V_{BC}} = \frac{\beta_R}{1 + \beta_R} \tag{6.67}$$

Saturation Mode ($V_{BE}, V_{BC} > V_T, I_B > 0$)

    This mode of operation implies the forward bias of both diodes, so that the diffusion current density in the base is the combination of forward and reverse carrier flows; that is, with (6.60) and (6.64):

$$J_{n\text{diff}}^B = J_{RE} - J_{FC} = -\frac{I_S}{A} e^{V_{BE}/V_T} + \frac{I_S}{A} e^{V_{BC}/V_T} \tag{6.68}$$

From (6.68) it is possible to find the emitter current by taking into account the forward base current. This forward base current (6.61) injects holes into the emitter and thus has to be taken with a negative sign to comply with our positive emitter current direction convention. Making the exponential expressions in (6.68) compatible with (6.61), we add and subtract unity and finally obtain

$$I_E \cong -I_S\left(e^{V_{BE}/V_T} - 1\right) - \frac{I_S}{\beta_F}\left(e^{V_{BE}/V_T} - 1\right) + I_S\left(e^{V_{BC}/V_T} - 1\right) \qquad (6.69)$$

Because the BJT can be treated as a symmetric device, the collector current is expressible in a similar manner as the contribution of three currents: the forward collector and reverse emitter currents, given by the negative of (6.68), and an additional hole diffusion contribution as the result of the reverse base current I_{RB}. The resulting equation is

$$I_C = I_S\left(e^{V_{BE}/V_T} - 1\right) - \frac{I_S}{\beta_R}\left(e^{V_{BC}/V_T} - 1\right) - I_S\left(e^{V_{BC}/V_T} - 1\right) \qquad (6.70)$$

Finally, the base current $I_B = -I_C - I_E$ is found from the preceding two equations:

$$I_B = I_S\left\{\frac{1}{\beta_R}\left(e^{V_{BC}/V_T} - 1\right) + \frac{1}{\beta_F}\left(e^{V_{BE}/V_T} - 1\right)\right\} \qquad (6.71)$$

Here again, it is important to recall that the internal emitter current flow is denoted opposite in sign to the customary external circuit convention.

6.3.3 Frequency Response

The **transition frequency** f_T (also known as the **cut-off frequency**) of a microwave BJT is an important figure of merit since it determines the operating frequency at which the common-emitter, short-circuit current gain h_{fe} decreases to unity. The transition frequency f_T is related to the transit time τ that is required for carriers to travel through the emitter-collector structure:

$$f_T = \frac{1}{\tau} \qquad (6.72)$$

This transition time is generally composed of three delays:

$$\tau = \tau_E + \tau_B + \tau_C \qquad (6.73)$$

where τ_E, τ_B, and τ_C are delays in emitter, base, and collector, respectively. The base-emitter depletion region charging time is given by

$$\tau_E = r_E C = \frac{V_T}{I_E}(C_E + C_C) \cong \frac{V_T}{I_C}(C_E + C_C) \qquad (6.74a)$$

where C_E, C_C are emitter and collector capacitances, and r_E is the emitter resistance obtained by differentiation of the emitter current with respect to base-emitter voltage. The second delay in (6.73) is the base layer charging time, and its contribution is given as

$$\tau_B = \frac{d_B^2}{\eta D_n^B} \tag{6.74b}$$

where the factor η is doping profile dependent and ranges from $\eta = 2$ for uniformly doped base layers up to $\eta = 60$ for highly nonuniform layers. Finally, the transition time τ_C through the base-collector junction space charge zone w_C can be computed as

$$\tau_C = \frac{w_C}{v_S} \tag{6.74c}$$

with v_S representing the saturation drift velocity. In the preceding formulas we have neglected the collector charging time $\tau_{CC} = r_C C_C$, which is typically very small when compared with τ_E.

As seen in (6.74a), the emitter charging time is inversely proportional to the collector current, resulting in higher transition frequencies for increasing collector currents. However, as the current reaches sufficiently high values, the concentration of charges injected into the base becomes comparable with the doping level of the base, which causes an increase of the effective base width and, in turn, reduces the transition frequency. Usually, BJT data sheets provide information about the dependence of the transition frequency on the collector current. For instance, Figure 6-34 shows the transition frequency as a function of collector current for the wideband npn-transistor BFG403W measured at $V_{CE} = 2$ V, $f = 2$ GHz, and at an ambient temperature of 25°C.
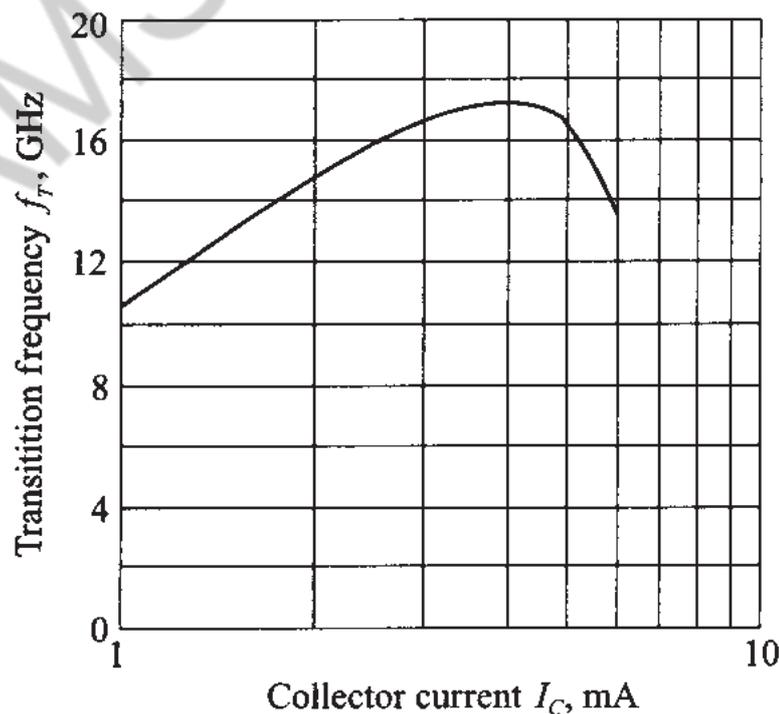
Figure 6-34   Transition frequency as a function of collector current for the 17 GHz npn wideband transistor BFG403W (courtesy of Philips Semiconductors).

Another aspect of the BJT operated at RF and MW frequencies is that at high frequencies the skin effect physically restricts current flow to the outer perimeter of the emitter (see also Section 1.4). To keep the charging time as low as possible, the emitter is constructed in a grid pattern of extremely narrow (less than 1 μm) strips. Unfortunately, the trade-off is a high current density over the small surface area, limiting the power handling capabilities. Additional ways to increase the cut-off frequency are to reduce the base transition time constant τ_B by high doping levels and concomitantly fabricate very short base layers of less than 100 nm. In addition, a small base thickness has as an advantage a reduction in power loss.

6.3.4 Temperature Behavior

We have seen in this chapter that almost all parameters describing both the static and dynamic behavior of semiconductor devices are influenced by the junction temperature T_j. As an example of such a dependence, in Figure 6-35 the forward current gain β_F for a given V_{CE} is plotted as a function of collector current I_C for various junction temperatures T_j. As we can see from this graph, the current gain raises from 40 at $I_C = 3.5$ mA and $T_j = -50°C$ to more than 80 at $T_j = 50°C$.
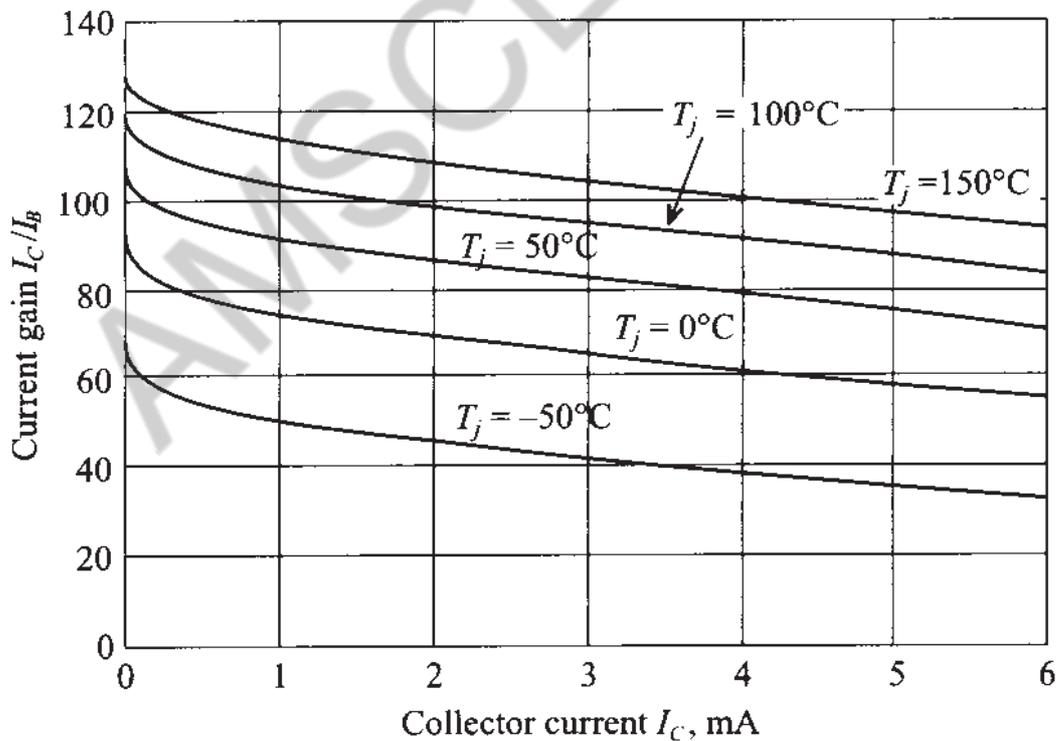
Figure 6-35 Current gain $\beta_F = \alpha_F/(1-\alpha_F)$ as a function of collector current for various junction temperatures at a fixed V_{CE}.

Another example that shows the strong temperature influence is the dependence of the input characteristic of a transistor described by the base current as a function of base-emitter voltage, as depicted in Figure 6-36.
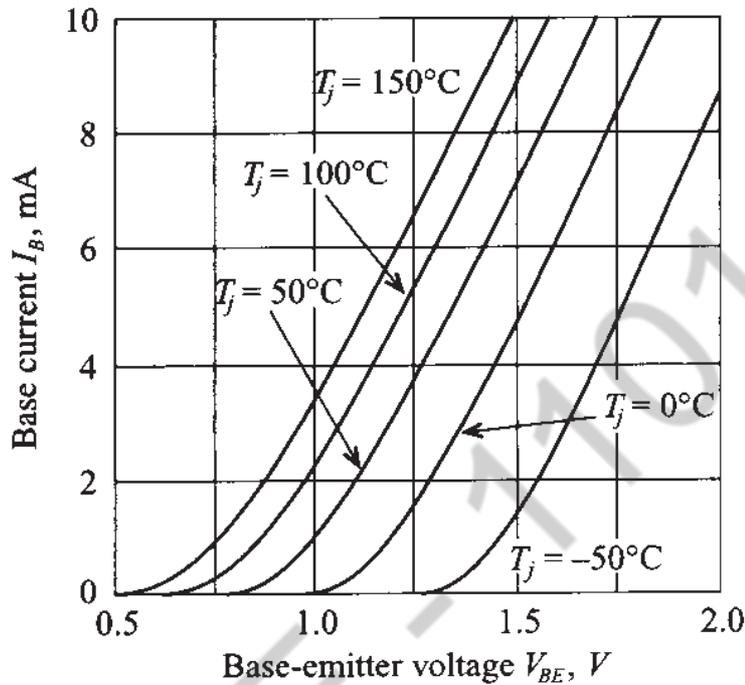
Figure 6-36 Typical base current as a function of base-emitter voltage for various junction temperatures at a fixed V_{CE}.

Again, if we compare the behavior of the transistor at $T_j = -50°C$ and $T_j = 50°C$, we notice that at $T_j = -50°C$ and a base-emitter voltage of 1.25 V the transistor is in cut-off state, whereas at $T_j = 50°C$ the BJT already conducts 4 mA base current. These two examples underscore the importance of temperature considerations in the design of RF circuits. For instance, the design of a cellular phone for worldwide use must ensure that our circuit preforms according to specifications under all temperature conditions encountered by the operator. Standard specifications usually cover the temperature range from $-50°C$ to $80°C$.

The junction temperature also plays an important role when dealing with the maximum power dissipation. In general, the manufacturer provides a **power derating curve** that specifies the temperature T_S up to which the transistor can be operated at the maximum available power P_{tot}. For junction temperatures T_j exceeding this value, the power has to be reduced to values dictated by the thermal resistance between the junction and the soldering point (or case) R_{thjs} according to

$$P = P_{tot}\frac{T_{jmax} - T_j}{T_{jmax} - T_S} = \frac{T_{jmax} - T_j}{R_{thjs}} \qquad (6.75)$$

where $T_{j\max}$ is the maximum junction temperature. Typical BJT values vary between 150 and 200°C.

For the RF transistor BFG403W the maximum total power P_{tot} of 16 mW can be maintained up to $T_S = 140°C$. For higher temperatures $T_S \leq T_j \leq T_{j\max}$, the power must be derated until the maximum junction temperature $T_{j\max}$ of 150°C is reached. The corresponding slope is 820°K/W. This value implies that if the power dissipation of the device decreases by 10 mW, the junction temperature can be increase by 8.2°C up to the maximum junction temperature. Obviously, transistor cases with such a high slope (or high thermal resistance) are not acceptable for high-power applications and manufacturers have to develop effective ways to dissipate the thermal energy generated by the transistor. Usually, this is done by employing heat sinks and using materials with high thermal conductivity. Instead of the thermal resistance at the soldering point R_{thjs}, the manufacturer may supply additional information involving heat resistances between junction-to-case (R_{thjc}), case-to-sink (R_{thcs}), and sink-to-air (R_{thha}) interfaces.

To simplify the thermal analysis it is convenient to resort to a thermal equivalent circuit with the following correspondences:

- Thermal power dissipation = electric current
- Temperature $\quad\quad\quad\quad\quad$ = electric voltage

A typical thermal circuit in equilibrium is shown in Figure 6-37, where the total electric power supplied to the device is balanced through a thermal circuit involving thermal resistances. In particular, we recognize the thermal resistance of junction to soldering point which is assumed to be equal to R_{thjc}. Therefore

$$R_{thjc} = R_{thjs} = \frac{T_j - T_s}{P_W} = \frac{1}{\gamma_{th} A_{BJT}} \qquad (6.76)$$

where junction and soldering point temperatures T_j and T_s and thermal power P_W determine the thermal resistance in Kelvin per Watt (°K/W), and whose value can also be expressed in terms of the thermal conductivity γ_{th} and the surface area A_{BJT} of the BJT. The solder point temperature is affected by the transition between casing and heat sink. This constitutes a thermal resistance R_{thcs} with values up to 5 °K/W. Finally, the heat sink represents a thermal resistance of

$$R_{thha} = \frac{1}{\delta_{hs} A_{hs}} \qquad (6.77)$$

where δ_{hs} is a convection coefficient that can vary widely between 10 W/(K·m^2) for still air, 100 W/(K·m^2) for forced air, up to 1000 W/(K·m^2) for water cooling, and A_{hs} is the total area of the heat sink.
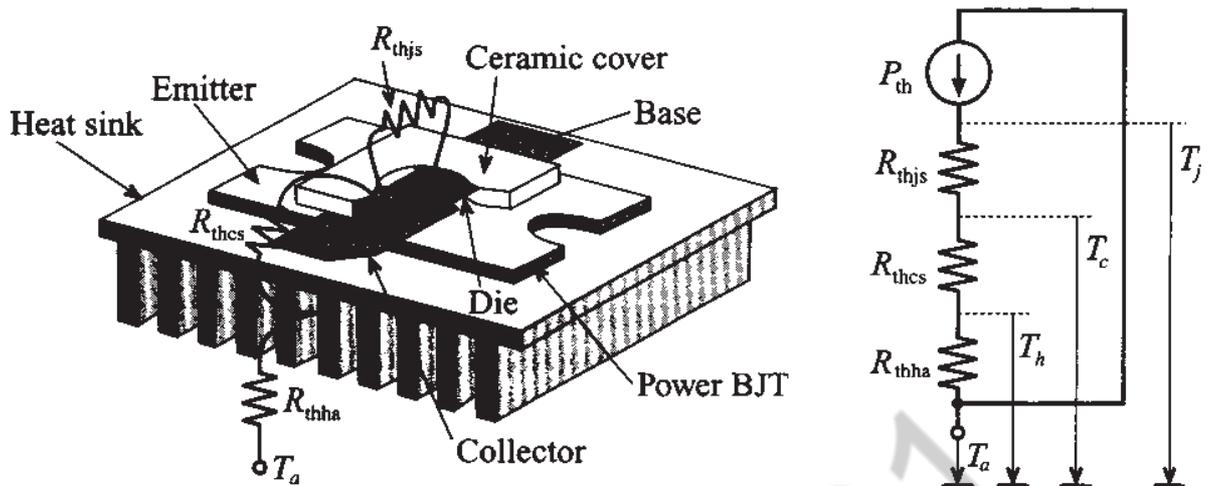
Figure 6-37 Thermal equivalent circuit of BJT.

The following example provides an often encountered design problem.
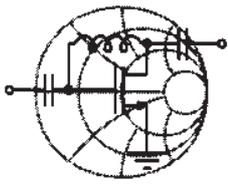
RF&MW

Example 6-8: **Thermal analysis involving a BJT mounted on a heat sink**

An RF power BJT generates a total power P_W of 15 W at case temperature of $25°C$. The maximum junction temperature is $150°C$ and the maximum ambient operating temperature is specified by the user to be $T_a = 60°C$. What is the maximum dissipated power if the thermal resistances between case-to-sink and sink-to-air is $2°K/W$ and $10°K/W$, respectively.

Solution: With reference to Figure 6-37, we are dealing with three thermal resistances: R_{thjs}, R_{thcs}, and R_{thha}. The junction-to-soldering resistance can be found based on equation (6.76):

$$R_{thjs} = \frac{T_j - T_s}{P_W} = \frac{150°C - 25°C}{15\ W} = 8.33°K/W$$

Adding up all resistances gives us a total thermal resistance of

$$R_{thtot} = R_{thjs} + R_{thca} + R_{thhs} = 20.333°K/W$$

The dissipated power P_{th} follows from the temperature drop (junction temperature T_j minus ambient temperature T_a) divided by the total thermal resistance:

$$P_{th} = \frac{T_j - T_a}{R_{thtot}} = \frac{150°C - 60°C}{20.33° \, K/W} = 4.43 \, W$$

To operate the BJT in thermal equilibrium, we have to reduce the total electric power $P_{tot} = P_W$ to the point where it is in balance with the computed thermal power $P_{tot} = P_{th}$. Thus a reduction from 15 W to 4.43 W is required.

> *While the design engineer cannot influence the junction-to-soldering point heat resistance, it is the choice of casing and heat sink that typically allows major improvements in thermal performance.*

6.3.5 Limiting Values

The total power dissipation capabilities at a particular temperature restrict the range of safe operation of the BJT. In our discussion we will exclusively focus on the active mode in the common-emitter configuration and will neglect the switch-mode behavior whereby the BJT is operated either in saturation or cut-off mode. For a given maximum BJT power rating, we can either vary the collector-emitter voltage V_{CE} and plot the allowable collector current $I_C = P_{tot}/V_{CE}$ (here we assume that base current is negligibly small compared to the collector current due to high β) or vary I_C and plot the allowable collector-emitter voltage $V_{CE} = P_{tot}/I_C$. The result is the **maximum power hyperbola**. This does not mean that I_C and V_{CE} can be increased without bound. In fact, we need to ensure that $I_C \leq I_{Cmax}$ and $V_{CE} \leq V_{CEmax}$, as depicted in Figure 6-38. The **safe operating area** (SOAR) is defined as a set of biasing points where the transistor can be operated without risk of unrecoverable damage to the device. The SOAR domain, shown as a shaded region in Figure 6-38, is more restrictive than a subset bounded by the maximum power hyperbola, since we have to take into account two more breakdown mechanisms:

1. **Breakdown of first kind**. Here the collector current density exhibits a nonuniform distribution that results in a local temperature increase, which in turn lowers the resistance of a portion of the collector domain, creating a channel. The consequence is a further increase in current density through this channel until the positive feedback begins to destroy the crystal structure (**avalanche breakdown**), ultimately destroying the transistor itself.

2. **Breakdown of second kind**. This breakdown mechanism can take place independently of the first mechanism and affects primarily power BJTs. Internal overheating may cause an abrupt increase in the collector current for constant V_{CE}. This breakdown mechanism usually occurs at the base-collector junction when the temperature increases to such high values that the intrinsic concentration is equal to the collector doping concentration. At this point the resistance of the junction is abruptly reduced, resulting in a dramatic current increase and melting of the junction.
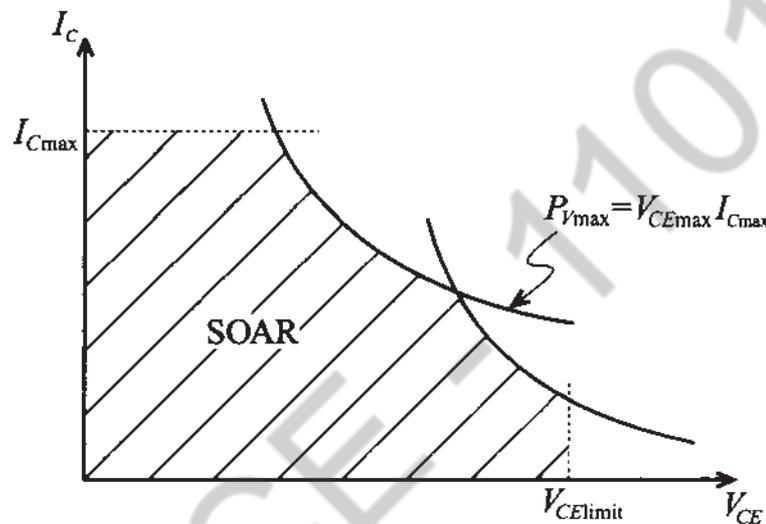
Figure 6-38 Operating domain of BJT in active mode with breakdown mechanisms.

It is interesting to point out that the BJT can exceed the SOAR, indeed even the maximum power hyperbola, for a short time since the temperature response has a much larger time constant (on the order of microseconds) in comparison with the electric time constants.

Additional parameters of importance to a design engineer are the maximum voltage conditions for open emitter, base and collector conditions; that is, V_{CBO} (collector-base voltage, open emitter), V_{CEO} (collector-emitter, open base), and V_{EBO} (emitter-base voltage, open collector). For instance, values for the BFG403W are as follows: $V_{CBO}|_{max} = 10 \text{ V}$, $V_{CEO}|_{max} = 4.5 \text{ V}$, and $V_{EBO}|_{max} = 1.0 \text{ V}$.

6.4 RF Field Effect Transistors

Unlike BJTs, **field effect transistors** (FETs) are **monopolar devices**, meaning that only one carrier type, either holes or electrons, contributes to the current flow through the channel. If hole contributions are involved we speak of **p-channel**, otherwise of **n-channel** FETs. Moreover, the FET is a voltage-controlled device. A variable

electric field controls the current flow from **source** to **drain** by changing the applied voltage on the **gate** electrode.

6.4.1   Construction

Traditionally FETs are classified according to how the gate is connected to the conducting channel. Specifically, the following four types are used:

1. **Metal Insulator Semiconductor FET** (MISFET). Here the gate is separated from the channel through an insulation layer. One of the most widely used types, the **Metal Oxide Semiconductor FET** (MOSFET), belongs to this class.
2. **Junction FET** (JFET). This type relies on a reverse biased *pn*-junction that isolates the gate from the channel.
3. **MEtal Semiconductor FET** (MESFET). If the reverse biased *pn*-junction is replaced by a Schottky contact, the channel can be controlled just as in the JFET case.
4. **Hetero FET**. As the name implies (and unlike the previous three cases, whose constructions rely on a single semiconductor material such as Si, GaAs, SiGe, or InP) the hetero structures utilize abrupt transitions between layers of different semiconductor materials. Examples are GaAlAs to GaAs or GaInAs to GaAlAs interfaces. The **High Electron Mobility Transistor** (HEMT) belongs to this class.

Figure 6-39 provides an overview of the first three types. In all cases the current flow is directed from the **source** to **drain**, with the **gate** controlling the current flow.

Due to the presence of a large capacitance formed by the gate electrode and the insulator or reverse biased *pn*-junction, MISFETs and JFETs have a relatively low cut-off frequency and are usually operated in low and medium frequency ranges of typically up to 1 GHz. GaAs MESFETs find applications up to 60–70 GHz, and HEMT can operate beyond 100 GHz. Since our interest is geared toward RF applications, the emphasis will be on the last two types.

In addition to the above physical classification, it is customary to electrically classify FETs according to **enhancement** and **depletion** types. This means that the channel either experiences an increase in carriers (for instance the *n*-type channel is injected with electrons) or a depletion in carriers (for instance the *n*-type channel is depleted of electrons). In Figure 6-39 (a) the FET is nonconducting, or **normally-off**, until a sufficiently positive gate voltage sets up a conduction channel. Normally-off FETs can only be operated in enhancement mode. Alternatively, **normally-on** FETs can be of both enhancement and depletion types.
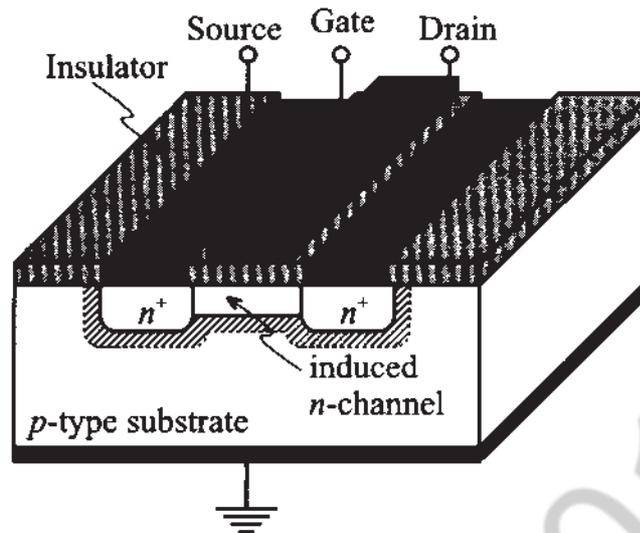
(a) Metal insulator semiconductor FET (MISFET)
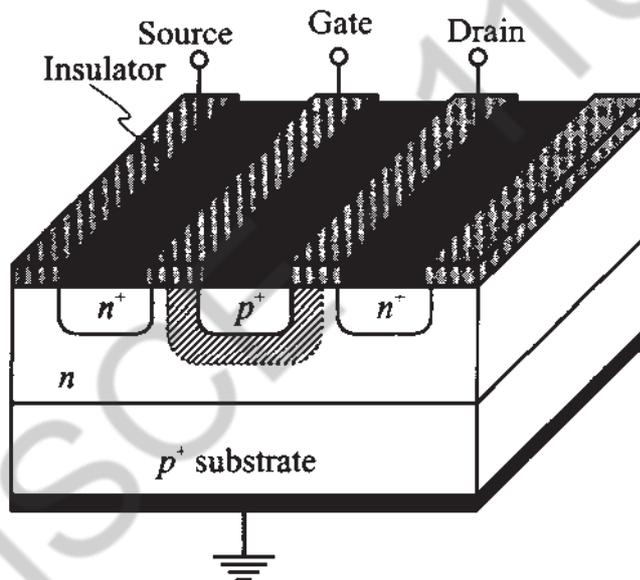
(b) Junction field effect transistor (JFET)
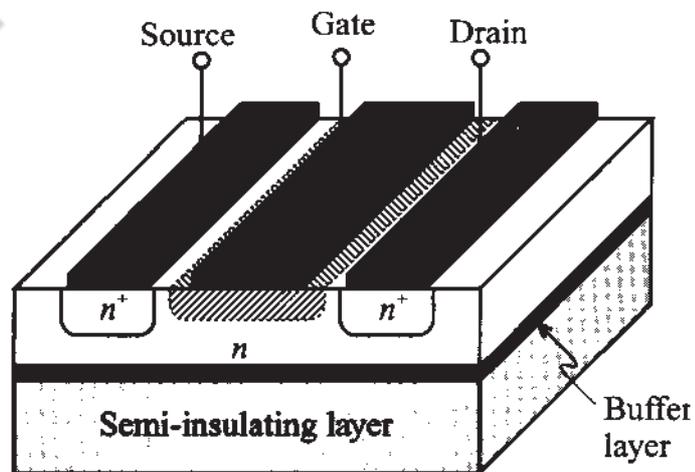
(c) Metal semiconductor FET (MESFET)

Figure 6-39 Construction of (a) MISFET, (b) JFET, and (c) MESFET. The shaded areas depict the space charge domains.

6.4.2 Functionality

Because of its importance in RF and MW amplifier, mixer, and oscillator circuits, we focus our analysis on the MESFET, whose physical behavior is in many ways similar to the JFET. The analysis is based on the geometry shown in Figure 6-40 where the transistor is operated in depletion mode.
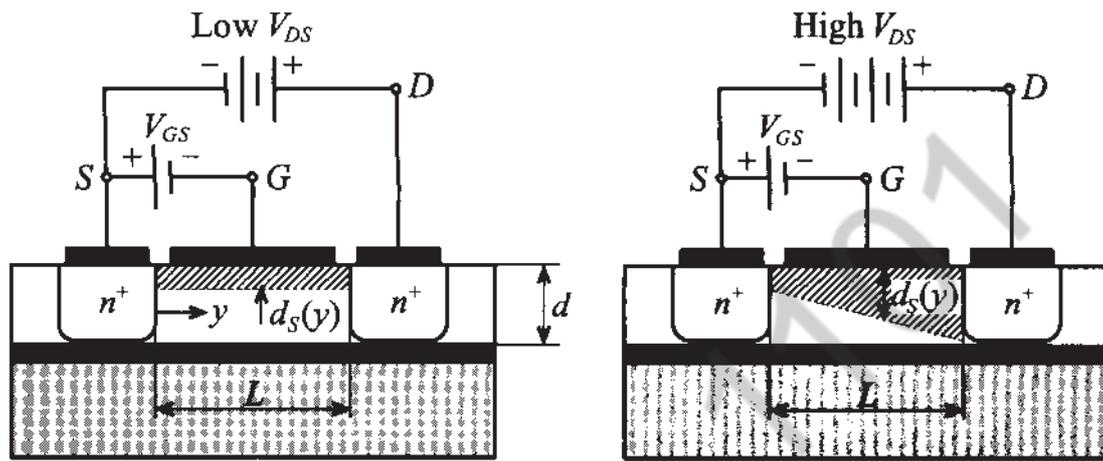
(a) Operation in the linear region.   (b) Operation in the saturation region.

Figure 6-40   Functionality of MESFET for different drain-source voltages.

The Schottky contact builds up a channel space charge domain that affects the current flow from source to drain. The space charge extent d_S can be controlled via the gate voltage in accordance to our discussion in Section 6.1.3, where (6.39) is adjusted such that V_A is replaced by the gate source voltage V_{GS}:

$$d_S = \sqrt{\frac{2\varepsilon}{q}\left(\frac{V_d - V_{GS}}{N_D}\right)} \tag{6.78}$$

For instance, the barrier voltage V_d is approximately 0.9 V for a GaAs-Au interface. The resistance R between source and drain is predicted by

$$R = \frac{L}{\sigma(d - d_S)W} \tag{6.79}$$

with the conductivity given by $\sigma = q\mu_n N_D$ and W being the gate width. Substituting (6.78) into (6.79) yields the drain-current equation:

$$I_D = \frac{V_{DS}}{R} = G_0\left[1 - \sqrt{\frac{2\varepsilon}{qd^2}\left(\frac{V_d - V_{GS}}{N_D}\right)}\right]V_{DS} \tag{6.80}$$

where we have defined the conductance $G_0 = \sigma W d / L$. This equation shows that the drain current depends linearly on the drain source voltage, a fact that is only true for small V_{DS}.

As the drain-source voltage increases, the space charge domain near the drain contact increases as well, resulting in a nonuniform distribution of the depletion region along the channel; see Figure 6-40(b). If we assume that the voltage along the channel changes from 0 at the source location to V_{DS} at the drain end, then we can compute the drain current for the nonuniform space charge region. This approach is also known as the **gradual-channel approximation**. The approximation rests primarily on the assumption that the cross-sectional area at a particular location y along the channel is given by $A(y) = \{d - d_S(y)\}W$ and the electric field E is only y-directed. The channel current is thus

$$I_D = -\sigma E A(y) = \sigma \frac{dV(y)}{dy} \{d - d_S(y)\} W \tag{6.81}$$

where the difference between V_d and V_{GS} in the expression for $d_S(y)$ has to be augmented by the additional drop in voltage $V(y)$ along the channel; that is, (6.78) becomes

$$d_S(y) = \left[\frac{2\varepsilon}{qN_D}(V_d - V_{GS} + V(y)) \right]^{1/2} \tag{6.82}$$

Substituting (6.82) into (6.81) and carrying out the integration on both sides of the equation yields

$$\int_0^L I_D dy = I_D L = \sigma W \int_0^{V_{DS}} \left(d - \sqrt{\frac{2\varepsilon}{qN_D}} \sqrt{V + V_d - V_{GS}} \right) dV \tag{6.83}$$

The result is the **output characteristic** of the MESFET in terms of the drain current as a function of V_{DS} for a given fixed V_{GS}, or

$$I_D = G_0 \left(V_{DS} - \frac{2}{3}\sqrt{\frac{2\varepsilon}{qN_D d^2}}[\{V_{DS} + V_d - V_{GS}\}^{3/2} - \{V_d - V_{GS}\}^{3/2}] \right) \tag{6.84}$$

This equation reduces for small V_{DS} to (6.80).

An interesting phenomenon occurs when the space charge extends over the entire channel depth d. The drain-source voltage for this situation is called **drain saturation voltage** $V_{D\text{sat}}$ and is given by

$$d_S(L) = d = \sqrt{\frac{2\varepsilon}{qN_D}(V_d - V_{GS} + V_{D\text{sat}})} \tag{6.85}$$

or, explicitly,

$$V_{Dsat} = \frac{qN_D d^2}{2\varepsilon} - (V_d - V_{GS}) = V_P - V_d + V_{GS} = V_{GS} - V_{T0} \qquad (6.86)$$

where we introduced the so-called **pinch-off voltage** $V_P = qN_D d^2/(2\varepsilon)$ and **threshold voltage** $V_{T0} = V_d - V_P$. The associated drain saturation current is found by inserting (6.86) into (6.84) with the result

$$I_{Dsat} = G_0 \left[\frac{V_P}{3} - (V_d - V_{GS}) + \frac{2}{3\sqrt{V_P}}(V_d - V_{GS})^{3/2} \right] \qquad (6.87)$$

The maximum saturation current in (6.87) is obtained when $V_{GS} = 0$, which we define as $I_{Dsat}(V_{GS} = 0) = I_{DSS}$. In Figure 6-41 the typical input/output transfer as well as the output characteristic behavior is shown.

(a) Circuit symbol   (b) Transfer characteristic   (c) Output characteristic
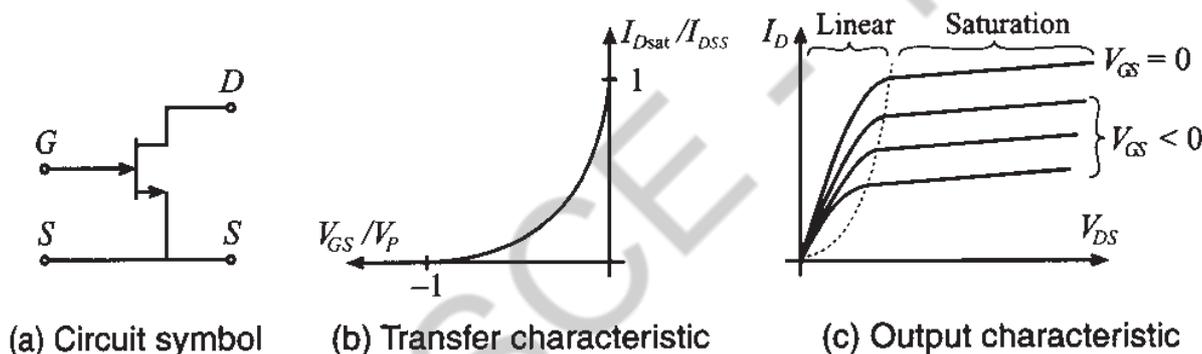
Figure 6-41   Transfer and output characteristics of an *n*-channel MESFET.

The saturation drain current (6.87) is often approximated by the simple relation

$$I_{Dsat} = I_{DSS} \left(1 - \frac{V_{GS}}{V_{T0}} \right)^2 \qquad (6.88)$$

How well (6.88) approximates (6.87) is discussed in the following example.
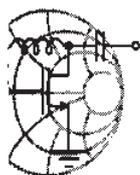
RF&MW

Example 6-9:   Drain saturation current in a MESFET

A GaAs MESFET has the following parameters: $N_D = 10^{16} \text{cm}^{-3}$, $d = 0.75\ \mu m$, $W = 10\ \mu m$, $L = 2\ \mu m$, $\varepsilon_r = 12.0$, $V_d = 0.8\ \text{V}$, and $\mu_n = 8500\ \text{cm}^2/(\text{Vs})$. Determine (a) the pinch-

off voltage, (b) the threshold voltage, (c) the maximum saturation current I_{DSS}; and plot the drain saturation current based on (6.87) and (6.88) for V_{GS} ranging from -4 to 0 V.

Solution: The pinch-off voltage for the FET is independent of the gate-source voltage and is computed as

$$V_p = \frac{qN_D d^2}{2\varepsilon} = 4.24 \text{ V}$$

Knowing V_p and the barrier voltage $V_d = 0.8 \text{ V}$, we find the threshold voltage to be $V_{T0} = V_d - V_P = -3.44 \text{ V}$. The maximum saturation drain current is again independent of the applied drain-source voltage and, based on (6.87), is equal to

$$I_{DSS} = G_0 \left[\frac{V_P}{3} - V_d + \frac{2}{3\sqrt{V_P}} V_d^{3/2} \right] = 6.89 \text{ A}$$

where $G_0 = \sigma q N_D W d / L = q^2 \mu_n N_D^2 W d / L = 8.16 \text{ S}$.

Figure 6-42 shows results for the saturation drain current computed by using the exact formula (6.87) and by using the quadratic law approximation given by (6.88).
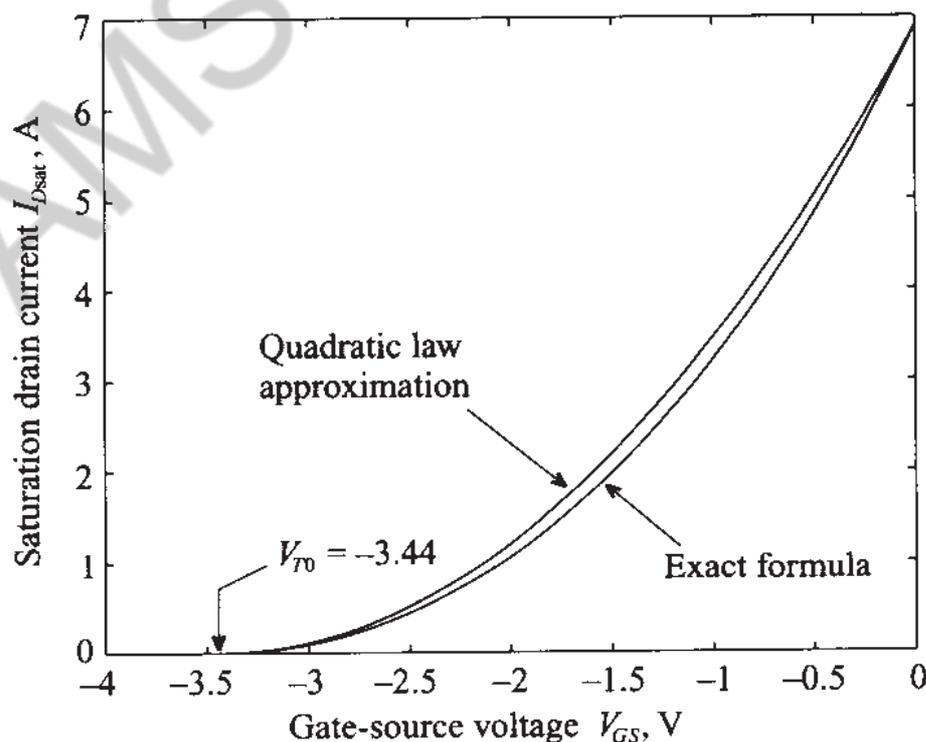
Figure 6-42   Drain current versus V_{GS} computed using the exact and the approximate equations (6.87) and (6.88).

Because of the excellent agreement, the quadratic law approximation (6.88) is more widely used in the literature and data sheets than the exact equation.

If V_{DS} reaches the saturation voltage $V_{D\text{sat}}$ for a given V_{GS}, the space charges pinch off the channel. This implies that the drain current saturates. Interestingly, *pinch-off does not imply a zero* I_D since there is no charge barrier impeding the flow of carriers. It is the electric field as a result of the applied voltage V_{DS} that "pulls" the electrons across the depletion space charge domain. Any additional increase $V_{DS} > V_{D\text{sat}}$ will result in a shortening of the channel length from the original length L to the new length $L' = L - \Delta L$. The result is that (6.87) must be modified to

$$I'_D = I_D\left(\frac{L}{L - \Delta L}\right) = I_D\left(\frac{L}{L'}\right) \tag{6.89}$$

The change in channel length as a function of V_{DS} is heuristically taken into account through the so-called **channel length modulation parameter** $\lambda = \Delta L/(L'V_{DS})$. This is particularly useful when expressing the drain current in the saturation region:

$$I'_{D\text{sat}} = I_{D\text{sat}}(1 + \lambda V_{DS}) \tag{6.90}$$

where measurements show a slight increase in drain current as V_{DS} is increased.
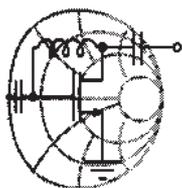
RF&MW→

Example 6-10: *I-V* characteristic of a MESFET

For discrete gate-source voltages $V_{GS} = -1, -1.5, -2,$ and -2.5 V, plot the drain current I_D of a MESFET as a function of drain-source voltage V_{DS} in the range from 0 to 5 V. Assume that the device parameters are the same as in the previous example and that the channel length modulation parameter λ is set to be 0.03 V^{-1}. Compare your results with the case where $\lambda = 0$.

Solution: In the analysis of the MESFET behavior we have to be careful about choosing the appropriate formulas. At very low drain-source voltages, the drain current can be described by a simple linear relation (6.80). As the voltage increases, this approximation

becomes invalid and a more complicated expression for I_D has to be employed; see (6.84). Further increase in V_{DS} ultimately leads to channel pinch-off, where $V_{DS} \geq V_{Dsat} = V_{GS} - V_{T0}$. In this case the drain current is equal to the saturation current given by (6.87). Additional increases in V_{DS} beyond the saturation voltage result only in minor increases of the drain current due to a shortening of the channel. At this point, I_D is linearly dependent on V_{DS}. Substituting (6.87) into (6.90) for $V_{DS} \geq V_{Dsat}$, we obtain

$$I_D = G_0 \left\{ \frac{V_P}{3} - (V_d - V_{GS}) + \frac{2}{3} \frac{\sqrt{(V_d - V_{GS})^3}}{\sqrt{V_P}} \right\} (1 + \lambda V_{DS})$$

To provide a smooth transition from normal to saturation region for nonzero λ we multiply (6.84) by $(1 + \lambda V_{DS})$. Thus, the final expression for the drain current for $V_{DS} \leq V_{Dsat}$ is

$$I_D = G_0 \left\{ V_{DS} - \frac{2}{3} \frac{\sqrt{(V_{DS} + V_d - V_{GS})^3} - \sqrt{(V_d - V_{GS})^3}}{\sqrt{V_p}} \right\} (1 + \lambda V_{DS})$$

The results of applying these formulas to predict I_D for zero (dashed line) as well as nonzero λ (solid line) are shown in Figure 6-43.
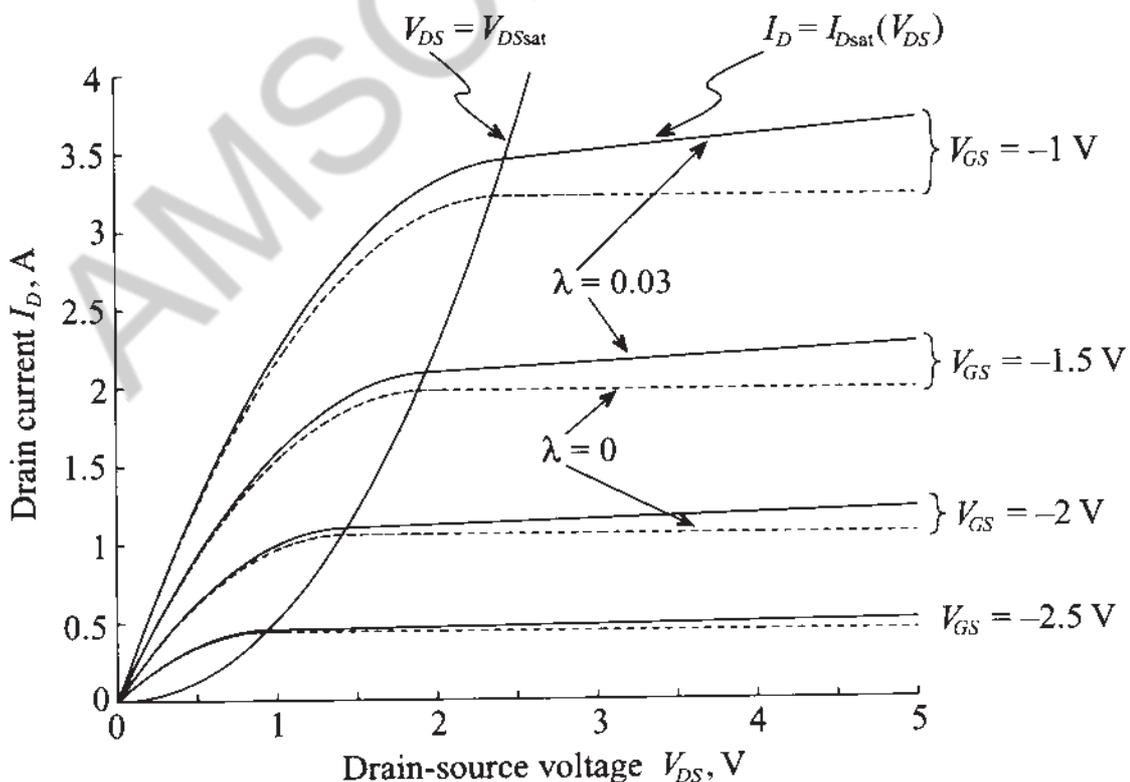
Figure 6-43   Drain current as a function of applied drain-source voltage for different gate-source biasing conditions.

The channel length modulation is similar to the Early effect encountered in a BJT, where the collector current in saturation mode increases slightly for increasing collector emitter voltage as discussed in Chapter 7.

6.4.3 Frequency Response

The high-frequency MESFET performance is determined by the transit time of charge carriers traveling between source and drain and the RC time constant of the device. Here we will focus our attention on the transit time only and defer the time constant computation, which requires knowledge of the channel capacitance, to Chapter 7. Since electrons in silicon and GaAs have much higher mobility than holes, n-channel MESFETs are used in RF and MW applications almost exclusively. Furthermore, since the electron mobility of GaAs is roughly five times higher than that of Si, GaAs MESFETs are usually preferred over Si devices.

The transit time τ of the electrons traveling through the channel of gate length L is computed as

$$\tau = \frac{L}{v_{\text{sat}}} \tag{6.91}$$

where we have assumed a fixed saturation velocity v_{sat}. As an example, the transition frequency $f_T = 1/(2\pi\tau)$ for a gate length of 1.0 μm and a saturation velocity of approximately 10^7 cm/s is 15 GHz.

6.4.4 Limiting Values

The MESFET must be operated in a domain limited by maximum drain current $I_{D\text{max}}$, maximum gate-source voltage $V_{GS\text{max}}$, and maximum drain-source voltage $V_{DS\text{max}}$. The maximum power P_{max} is dictated by the product of V_{DS} and I_D, or

$$P_{\text{max}} = V_{DS}I_D \tag{6.92}$$

which in turn is related to the channel temperature T_C and ambient temperature T_a and the thermal resistance between channel and soldering point R_{thjs}, according to

$$T_C = T_a + R_{\text{thjs}}P \tag{6.93}$$

Figure 6-44 clarifies this point. Also shown in this figure are three possible operating points. Bias point 3 indicates low amplification and possible clipping of the output current. However, the power consumption is at a minimum. Bias point 2 reveals accept-
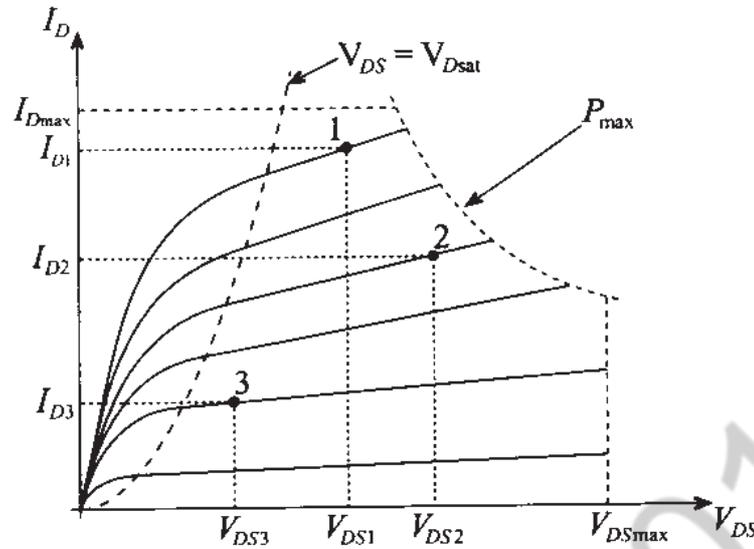
Figure 6-44   Typical maximum output characteristics and three operating points of MESFET.

able amplification at substantially increased power consumption. Finally, bias point 1 shows high amplification at high power consumption and low output current swing. Choosing appropriate bias points for specific applications will be investigated in-depth in subsequent chapters.

6.5 High Electron Mobility Transistors

The **high electron mobility transistor** (HEMT), also known as **modulation-doped field effect transistor** (MODFET), exploits the differences in band gap energy between dissimilar semiconductor materials such as GaAlAs and GaAs in an effort to *substantially surpass the upper frequency limit of the MESFET while maintaining low noise performance and high power rating*. At present, transit frequencies of 100 GHz and above have been achieved. The high frequency behavior is due to a separation of the carrier electrons from their donor sites at the interface between the doped GaAlAs and undoped GaAs layer (**quantum well**), where they are confined to a very narrow (about 10 nm thick) layer in which motion is possible only parallel to the interface. Here we speak of a **two-dimensional electron gas** (2DEG) or plasma of very high mobility, up to 9000 $cm^2/(V \cdot s)$. This is a major improvement over GaAs MESFETs with $\mu_n \approx 4500$ $cm^2/(V \cdot s)$. Because of the thin layer, the carrier density is often specified in terms of a surface density, typically on the order of 10^{12}–10^{13} cm^{-2}.

To further reduce carrier scattering by impurities it is customary to insert a spacer layer ranging between 20 and 100 nm of undoped GaAlAs. The layer is grown through a molecular beam epitaxial process and has to be sufficiently thin so as to allow the gate voltage V_{GS} to control the electron plasma through electrostatic force mechanism. Besides single layer heterostructures (GaAlAs on GaAs), multilayer heterostructures

involving several 2DEG channels have also been proposed. As can be expected, manufacturing an HEMT is significantly more expensive when compared with the relatively inexpensive GaAs MESFET due to the precisely controlled thin-layer structures, steep doping gradients, and the use of more difficult to fabricate semiconductor materials.

6.5.1 Construction

The basic heterostructure is shown in Figure 6-45, where a GaAlAs n-doped semiconductor is followed by an undoped GaAlAs spacer layer of the same material, an undoped GaAs layer, and a high resistive semi-insulating (s.i.) GaAl substrate.
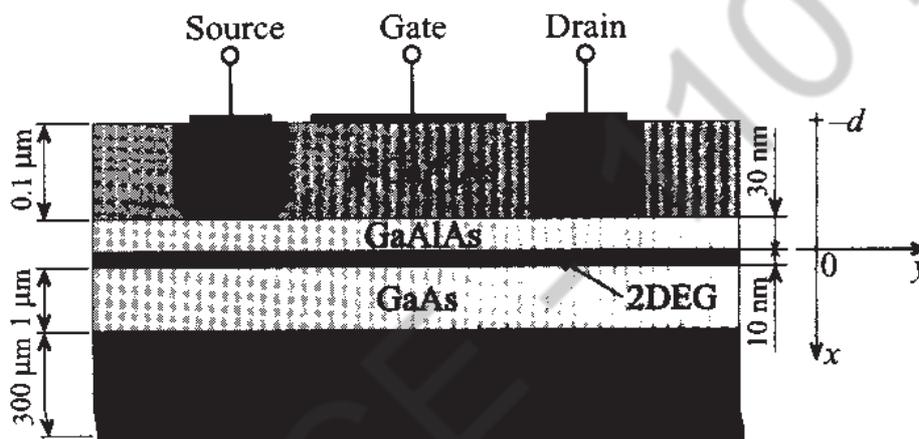
Figure 6-45 Generic heterostructure of a depletion-mode HEMT.

The 2DEG is formed in the undoped GaAs layer for zero gate bias condition because the Fermi level is above the conduction band so that electrons accumulate in this narrow potential well. As discussed later, the electron concentration can be depleted by applying an increasingly negative gate voltage.

HEMTs are primarily constructed of heterostructures with matching lattice constants to avoid mechanical tensions between layers. Specific examples are the GaAlAs-GaAs and InGaAs-InP interfaces. Research is also ongoing with mis-matched lattices whereby, for instance, a larger InGaAs lattice is compressed onto a smaller GaAs lattice. Such device configurations are known as **pseudomorphic HEMTs**, or pHEMTs.

6.5.2 Functionality

The key issue that determines the drain current flow in a HEMT is the narrow interface between the GaAlAs and the GaAs layers. For simplicity, we neglect the spacer layer and concentrate our attention at the energy band model shown in Figure 6-46.

A mathematical model similar to (6.21) can be developed by writing down the one-dimensional Poisson equation in the form
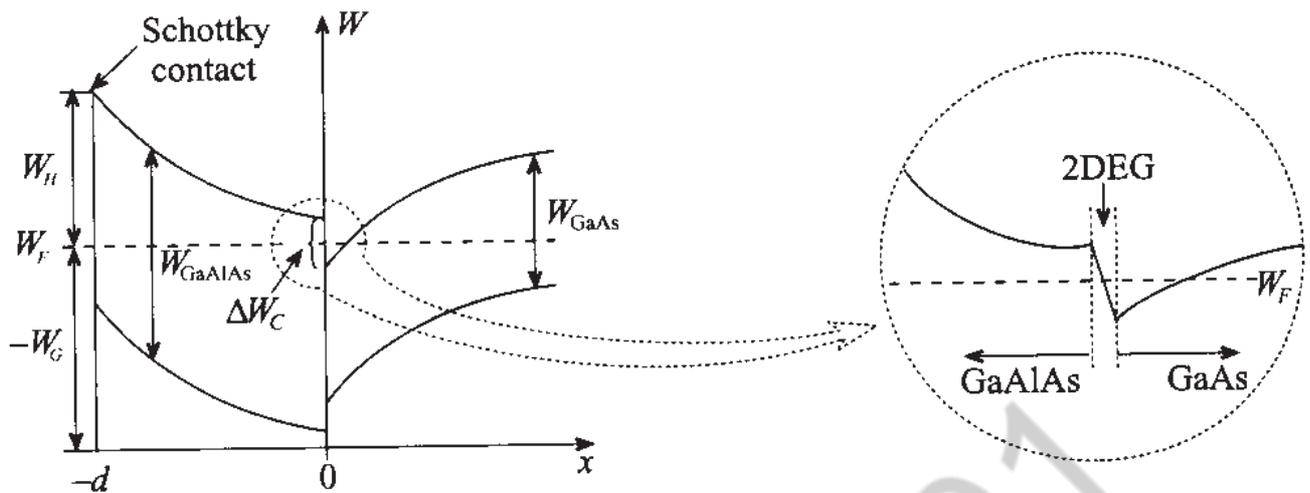
(a) Energy band diagram   (b) Close-up view of conduction band

Figure 6-46   Energy band diagram of GaAlAs-GaAs interface for an HEMT.

$$\frac{d^2V}{dx^2} = -\frac{qN_D}{\varepsilon_H} \tag{6.94}$$

where N_D and ε_H are the donor concentration and dielectric constant in the GaAlAs heterostructure. The boundary conditions for the potential are imposed such that $V(x = 0) = 0$ and at the metal-semiconductor side $V(x = -d) = -V_b + V_G + \Delta W_C/q$. Here V_b is the barrier voltage, see (6.38); ΔW_C is the energy difference in the conduction levels between the n-doped GaAlAs and GaAs; and V_G is comprised of the gate-source voltage as well as the channel voltage drop $V_G = -V_{GS} + V(y)$. To find the potential, (6.94) is integrated twice. At the metal-semiconductor we set

$$V(-d) = \frac{qN_D}{2\varepsilon_H}x^2 - E_y(0)d \tag{6.95}$$

which yields

$$E(0) = \frac{1}{d}(V_{GS} - V(y) - V_{T0}) \tag{6.96}$$

where we defined the HEMT threshold voltage V_{T0} as $V_{T0} = V_b - \Delta W_C/q - V_P$. Here we have used the previously defined pinch-off voltage $V_P = qN_D d^2/(2\varepsilon_H)$. From the known electric field at the interface, we find the electron drain current

$$I_D = \sigma E_y A = -q\mu_n N_D E W d = q\mu_n N_D \left(\frac{dV}{dy}\right) W d \tag{6.97}$$

As mentioned previously, the current flow is restricted to a very thin layer so that it is appropriate to carry out the integration over a surface charge density Q_S at $x = 0$. The

result is $\sigma = -\mu_n Q/(WLd) = -\mu_n Q_S/d$. For the surface charge density we find with Gauss's law $Q_S = \varepsilon_H E(0)$. Inserted in (6.97), we obtain

$$\int_0^L I_D \, dy = \mu_n W \int_0^{V_{DS}} Q_S \, dV \tag{6.98a}$$

Upon using (6.96), it is seen that the drain current can be found

$$I_D L = \mu_n W \int_0^{V_{DS}} \frac{\varepsilon_H}{d}(V_{GS} - V - V_{T0}) \, dV \tag{6.98b}$$

or

$$I_D = \mu_n \frac{W\varepsilon_H}{Ld}\left\{ V_{DS}(V_{GS} - V_{T0}) - \frac{V_{DS}^2}{2} \right\} \tag{6.98c}$$

Pinch-off occurs when the drain-source voltage is equal to or less than the difference of gate-source and threshold voltages (i.e., $V_{DS} \le V_{GS} - V_{T0}$). If the equality of this condition is substituted in (6.98c), it is seen

$$I_D = \mu_n \frac{W\varepsilon_H}{2Ld}(V_{GS} - V_{T0})^2 \tag{6.99}$$

The threshold voltage allows us to determine if the HEMT is operated as an enhancement or depletion type. For the depletion type we require $V_{T0} < 0$, or $V_b - (\Delta W_C/q) - V_P < 0$. Substituting the pinch-off voltage $V_P = qN_D d^2/(2\varepsilon)$ and solving for d, this implies

$$d > \left\{ \frac{2\varepsilon_H}{qN_D}\left(V_b - \frac{\Delta W_C}{q}\right) \right\}^{1/2} \tag{6.100}$$

and if d is less than the preceding expression (i.e., $V_{T0} > 0$), we deal with an enhancement HEMT.
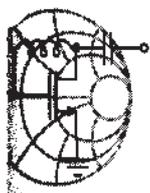
⟜RF&MW→

Example 6-11: Computation of HEMT-related electric characteristics

Determine typical numerical values for a HEMT device such as pinch-off voltage, threshold voltage, and drain current for $V_{GS} = -1, -0.75, -0.5, -0.25,$ and 0 V as a function of drain-

source   voltage V_{DS}.   Assume   the   following   parameters:
$N_D = 10^{18}$ cm^{-3}, $V_b = 0.81$ V, $\varepsilon_H = 12.5\varepsilon_0$, $d = 50$ nm, $\Delta W_C = 3.5 \times 10^{-20}$ W·s, $W = 10$ μm, $L = 0.5$ μm, and $\mu_n = 8500$ cm^2/(V·s).

Solution: The pinch-off voltage of a HEMT is evaluated as

$$V_P = qN_Dd^2/(2\varepsilon_H) = 1.81 \text{ V}$$

Knowing V_P we can find the threshold voltage as

$$V_{T0} = V_b - \Delta W_C/q - V_P = -1.22 \text{ V}$$

Using these values the drain current is computed by relying either on equation (6.98c) for $V_{DS} \leq V_{GS} - V_{T0}$ or equation (6.99) for $V_{DS} \geq V_{GS} - V_{T0}$. The results of these computations are plotted in Figure 6-47. We notice in this graph that unlike the GaAs MESFET in Figure 6-43, a channel length modulation is not taken into account. In practical simulations such a heuristic adjustment can be added.
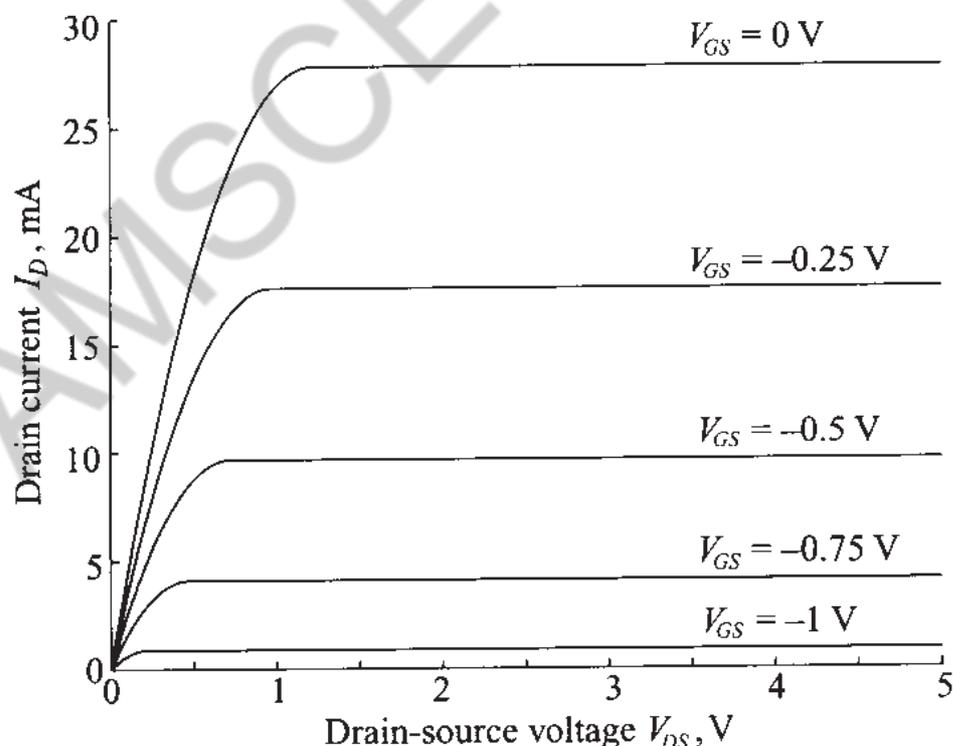
Figure 6-47   Drain current in a GaAs HEMT.

Both GaAs MESFET and HEMT exhibit similar output characteristics and are thus represented by the same electric circuit model.

6.5.3 Frequency Response

The high-frequency performance of the HEMT is determined by the transit time similar to the MESFET. However, the transit time τ is expressed best through the electron mobility μ_n and the electric field E of the drain-source voltage according to

$$\tau = \frac{L}{v_{\text{sat}}} = \frac{L}{\mu_n E_y} = \frac{L^2}{\mu_n V_{DS}} \tag{6.101}$$

We therefore obtain a transit frequency $f_T = 1/(2\pi\tau)$ of approximately 190 GHz for the gate length of 1.0 μm and a mobility of $\mu_n = 8000 \text{ cm}^2/(\text{V·s})$ at a typical drain voltage V_{DS} of 1.5 V.

6.6 Summary

To understand the functionality and limitations of the most widely employed active RF solid-state devices, we commenced this chapter with a review of the key elements of semiconductor physics. The concepts of conduction, valence, and Fermi levels as part of the energy band model are used as the starting point to examine the various solid-state mechanisms.

We next turned our attention to the *pn*-junction, where we derived the barrier voltage

$$V_{\text{diff}} = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right)$$

and the depletion and diffusion capacitances C_d and C_S in the forms

$$C_d = \frac{C_0}{\sqrt{1 - V_A/V_{\text{diff}}}} \quad \text{and} \quad C_S = \frac{\tau I_0}{V_T} e^{V_A/V_T}$$

Both capacitances are of primary importance when dealing with the frequency response of a *pn*-diode whose current is given by the Shockley equation

$$I = I_S\left(e^{V_A/V_T} - 1\right)$$

This equation underscores the nonlinear current-voltage diode characteristics.

Unlike the *pn*-junction, the Schottky contact involves an *n*-type semiconductor and a metal interface. The Schottky barrier potential V_d is now modified and requires the work function of metal, qV_M, semiconductor, $q\chi$, and the conduction band potential V_C, expressed via

Rotating y_C by $l_2 = 3\lambda/8$ we find $y_B = 1 + j3$, which means that we have to make the susceptance of the second stub equal to $jb_{S2} = -j3$ so that $y_{\text{in}} = y_A = 1$. Using the Smith Chart, we find that the length of the second stub is $l_{S2} = 0.051\lambda$.

In some practical realizations the stubs are replaced by varactor diodes. This allows an electronic tuning of the diode capacitances and thus the shunt admittances.

8.3 Amplifier Classes of Operation and Biasing Networks

An indispensable building block in any RF circuit is the active or passive biasing network. The purpose of biasing is to provide the appropriate quiescent point for the active devices under specified operating conditions and maintain a constant setting irrespective of transistor parameter variations and temperature fluctuations.

In the following section we introduce a general analysis of the different classes of amplifier operation. This will enable us to develop an understanding of how BJT and FET need to be appropriately biased.

8.3.1 Classes of Operation and Efficiency of Amplifiers

Depending on the application for which the amplifier is designed, specific bias conditions are required. There are several classes of amplifier operation that describe the biasing of an active device in an RF circuit.

In Figure 8-29 the transfer function characteristic of an ideal transistor is displayed. It is assumed that the transistor does not reach saturation or breakdown regions and in the linear operating region the output current is proportional to the input voltage. The voltage V^* corresponds either to the threshold voltage in case of FETs or the base-emitter built-in potential in case of BJTs.

The distinction between different classes of operation is made based upon the so-called **conduction angle**, which indicates the portion of the signal cycle when the current is flowing through the load. As depicted in Figure 8-29(a), in **Class A** operation the current is present during the entire output signal cycle. This corresponds to a $\Theta_A = 360°$ conduction angle. If the transfer characteristic of the transistor in the linear region is close to that of a linear function, then the output signal is an amplified replica of the input signal without suffering any distortion. In practical circuits, however,
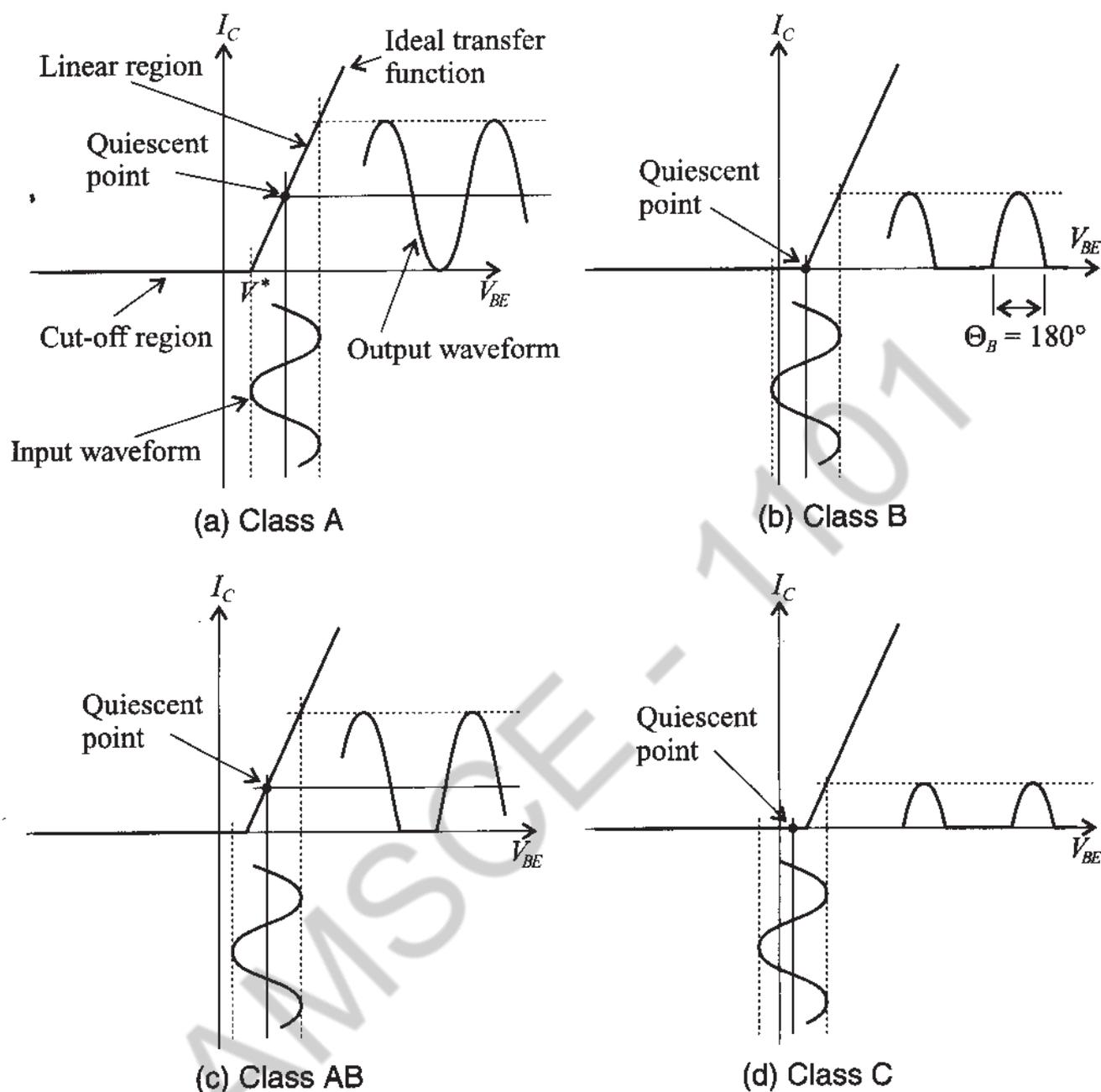
Figure 8-29 Various classes of amplifier operation.

there is always a certain degree of nonlinearity present which results in a distorted output signal of the amplifier.

In **Class B** [Figure 8-29(b)] the current is present during only half of the cycle, corresponding to a $\Theta_B = 180°$ conduction angle. During the second half of the cycle, the transistor is in the cut-off region and no current flows through the device. **Class AB** [Figure 8-29(c)] combines the properties of the classes A and B and has a conduction angle Θ_{AB} ranging from $180°$ to $360°$. This type of amplifier is typically employed when a high-power "linear" amplification of the RF signal is required.

In a **Class C** amplifier [Figure 8-29(d)] we have a nonzero current for less than half of the cycle (i.e., the conduction angle is $0 < \Theta_C < 180°$). This results in maximum distortion of the output signal.

A logical question that arises is why are not all amplifiers operated in Class A since this mode delivers the least signal distortion? The answer is directly linked to the amplifier efficiency. Efficiency, η, is defined as the ratio of the average RF power P_{RF} delivered to the load over the average power P_S supplied by the source, and is usually measured in percent:

$$\eta = \frac{P_{RF}}{P_S} 100\%$$ (8.22)

The theoretical maximum efficiency of the Class A amplifier is only 50%, but the efficiency of Class C can reach values close to 100%. Fifty percent efficiency of Class A amplifiers means that half of the power supplied by the source is dissipated as heat. This situation may not be acceptable in portable communication systems where most devices are battery operated. In practical applications, designers usually choose the class of operation that gives maximum efficiency but still preserves the informational content of the RF signal.

In the following example we derive the maximum theoretical efficiency η of the amplifier as a function of conduction angle.
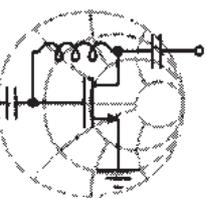
Example 8-11: Amplifier efficiency computation

Derive the general expression for the amplifier efficiency η as a function of conduction angle Θ_0. List the values of η for both Class A and Class B amplifiers.

Solution: The electrical current through the load for a conduction angle of Θ_0 has a waveform shown in Figure 8-30(a), where the cosine current amplitude is given by I_0.

Similarly, the power supply current I_S has a maximum value of I_0 plus the quiescent current I_Q:

$$I_S = I_Q + I_0 \cos \Theta$$ (8.23)

$$P_{\text{inc}} = \frac{|b_1'|^2}{2} = \frac{1}{2}\frac{|b_S|^2}{|1 - \Gamma_{in}\Gamma_S|^2} \tag{9.2}$$

which is the power launched toward the amplifier. The actual input power P_{in} observed at the input terminal of the amplifier is composed of the incident and reflected power waves. With the aid of the input reflection coefficient Γ_{in} we can therefore write:

$$P_{\text{in}} = P_{\text{inc}}(1 - |\Gamma_{\text{in}}|^2) = \frac{1}{2}\frac{|b_S|^2}{|1 - \Gamma_{\text{in}}\Gamma_S|^2}(1 - |\Gamma_{\text{in}}|^2) \tag{9.3}$$

The **maximum power transfer** from the source to the amplifier is achieved if the input impedance is complex conjugate matched ($Z_{\text{in}} = Z_S^*$) or, in terms of the reflection coefficients, if $\Gamma_{\text{in}} = \Gamma_S^*$. Under maximum power transfer condition, we define the **available power** P_A as

$$P_A = P_{\text{in}}\Big|_{\Gamma_{\text{in}} = \Gamma_S^*} = \frac{1}{2}\frac{|b_S|^2}{|1 - \Gamma_{\text{in}}\Gamma_S|^2}\Bigg|_{\Gamma_{\text{in}} = \Gamma_S^*}(1 - |\Gamma_{\text{in}}|^2) = \frac{1}{2}\frac{|b_S|^2}{1 - |\Gamma_S|^2} \tag{9.4}$$

This expression makes clear the dependence on Γ_S. If $\Gamma_{\text{in}} = 0$ and $\Gamma_S \neq 0$ it is seen from (9.2) and (9.4) that $P_{\text{inc}} = |b_S|^2/2$.

9.2.2 Transducer Power Gain

We can next investigate the **transducer power gain** G_T, which quantifies the gain of the amplifier placed between source and load.

$$G_T = \frac{\text{power delivered to the load}}{\text{available power from the source}} = \frac{P_L}{P_A}$$

or with $P_L = \frac{1}{2}|b_2|^2 \cdot (1 - |\Gamma_L|^2)$ we obtain

$$G_T = \frac{P_L}{P_A} = \frac{|b_2|^2}{|b_S|^2}(1 - |\Gamma_L|^2)(1 - |\Gamma_S|^2) \tag{9.5}$$

In this expression, the ratio b_2/b_S has to be determined. With the help of our signal flow discussion in Section 4.4.5 and based on Figure 9-2, we establish

$$b_2 = \frac{S_{21}a_1}{1 - S_{22}\Gamma_L} \tag{9.6a}$$

$$b_S = \left[1 - \left(S_{11} + \frac{S_{21}S_{12}\Gamma_L}{1 - S_{22}\Gamma_L}\right)\Gamma_S\right]a_1 \tag{9.6b}$$

The required ratio is therefore given by

$$\frac{b_2}{b_S} = \frac{S_{21}}{(1 - S_{11}\Gamma_S)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_S} \tag{9.7}$$

Inserting (9.7) into (9.5) results in

$$G_T = \frac{(1 - |\Gamma_L|^2)|S_{21}|^2(1 - |\Gamma_S|^2)}{|(1 - S_{11}\Gamma_S)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_S|^2} \tag{9.8}$$

which can be rearranged by defining the input and output reflection coefficients (see Problem 9.2)

$$\Gamma_{\text{in}} = S_{11} + \frac{S_{21}S_{12}\Gamma_L}{1 - S_{22}\Gamma_L} \tag{9.9a}$$

$$\Gamma_{\text{out}} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S} \tag{9.9b}$$

With these two definitions, two more transducer power gain expressions can be derived. First, by incorporating (9.9a) into (9.8), it is seen that

$$G_T = \frac{(1 - |\Gamma_L|^2)|S_{21}|^2(1 - |\Gamma_S|^2)}{|1 - \Gamma_S\Gamma_{\text{in}}|^2|1 - S_{22}\Gamma_L|^2} \tag{9.10}$$

Second, using (9.9b) in (9.8) results in the expression

$$G_T = \frac{(1 - |\Gamma_L|^2)|S_{21}|^2(1 - |\Gamma_S|^2)}{|1 - \Gamma_L\Gamma_{\text{out}}|^2|1 - S_{11}\Gamma_S|^2} \tag{9.11}$$

An often employed approximation for the transducer power gain is the so-called **unilateral power gain**, G_{TU}, which neglects the feedback effect of the amplifier ($S_{12} = 0$). This simplifies the form (9.11) to

$$G_{TU} = \frac{(1 - |\Gamma_L|^2)|S_{21}|^2(1 - |\Gamma_S|^2)}{|1 - \Gamma_L S_{22}|^2|1 - S_{11}\Gamma_S|^2} \tag{9.12}$$

As discussed in Section 9.4.1, equation (9.12) is often used as a basis to develop approximate designs for an amplifier and its input and output matching networks.

Using (9.2) in conjunction with (9.1) allows us to find the incident power flow into the amplifier:

$$P_{inc} = \frac{1}{2}\frac{|b_S|^2}{|1 - \Gamma_{in}\Gamma_S|^2} = \frac{1}{2}\frac{Z_0}{(Z_S + Z_0)^2}\frac{|V_S|^2}{|1 - \Gamma_{in}\Gamma_S|^2} = 74.7 \text{ mW}$$

Often P_{inc} is expressed in dBm as

$$P_{inc}(\text{dBm}) = 10\log[P_{inc}/(1 \text{ mW})] = 18.73 \text{ dBm}$$

Similarly, from (9.2) we find the available power to be $P_A = 78.1$ mW or $P_A = 18.93$ dBm. Finally, the power delivered to the load is the available power multiplied by the transducer gain. This results in $P_L = P_A G_T = 981.4$ mW, or, expressed in dBm,

$$P_L(\text{dBm}) = P_A(\text{dBm}) + G_T(\text{dB}) = 29.92 \text{ dBm}$$

It is interesting to point out that the unilateral power gain often matches the actual transducer power gain very closely. As discussed further, the use of the unilateral amplifier gain significantly simplifies the amplifier design task.

9.3 Stability Considerations

9.3.1 Stability Circles

One of the first requirements that an amplifier circuit must meet is a stable performance in the frequency range of interest. This is a particular concern when dealing with RF circuits, which tend to oscillate depending on operating frequency and termination. The phenomenon of oscillations can be understood in the context of a voltage wave along a transmission line. If $|\Gamma_0| > 1$, then the return voltage increases in magnitude (positive feedback) causing instability. Conversely, $|\Gamma_0| < 1$ causes a diminished return voltage wave (negative feedback).

Let us regard the amplifier as a two-port network characterized through its S-parameters and external terminations described by Γ_L and Γ_S. Stability then implies that the magnitudes of the reflection coefficients are less than unity. Namely,

$$|\Gamma_L| < 1, |\Gamma_S| < 1 \tag{9.15a}$$

$$|\Gamma_{\text{in}}| = \left|\frac{S_{11} - \Gamma_L \Delta}{1 - S_{22}\Gamma_L}\right| < 1 \tag{9.15b}$$

$$|\Gamma_{\text{out}}| = \left|\frac{S_{22} - \Gamma_S \Delta}{1 - S_{11}\Gamma_S}\right| < 1 \tag{9.15c}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$ has been used to re-express (9.9a) and (9.9b). Since the S-parameters are fixed for a particular frequency, the only factors that have a parametric effect on the stability are Γ_L and Γ_S.

In terms of the amplifier's *output port*, we need to establish the condition for which (9.15b) is satisfied. To this end the complex quantities

$$S_{11} = S_{11}^R + jS_{11}^I, \, S_{22} = S_{22}^R + jS_{22}^I, \, \Delta = \Delta^R + j\Delta^I, \, \Gamma_L = \Gamma_L^R + j\Gamma_L^I \tag{9.16}$$

are substituted into (9.15b), resulting after some algebra in the **output stability circle** equation

$$(\Gamma_L^R - C_{\text{out}}^R)^2 + (\Gamma_L^I - C_{\text{out}}^I)^2 = r_{\text{out}}^2 \tag{9.17}$$

where the circle radius is given by

$$r_{\text{out}} = \frac{|S_{12}S_{21}|}{\left||S_{22}|^2 - |\Delta|^2\right|} \tag{9.18}$$

and the center of this circle is located at

$$C_{\text{out}} = C_{\text{out}}^R + jC_{\text{out}}^I = \frac{(S_{22} - S_{11}^*\Delta)^*}{|S_{22}|^2 - |\Delta|^2} \tag{9.19}$$

as depicted in Figure 9-3(a). In terms of the *input port*, substituting (9.16) into (9.15c) yields the **input stability circle** equation

$$(\Gamma_S^R - C_{\text{in}}^R)^2 + (\Gamma_S^I - C_{\text{in}}^I)^2 = r_{\text{in}}^2 \tag{9.20}$$

where

$$r_{\text{in}} = \frac{|S_{12}S_{21}|}{\left||S_{11}|^2 - |\Delta|^2\right|} \tag{9.21}$$

and

$$C_{\text{in}} = C_{\text{in}}^R + jC_{\text{in}}^I = \frac{(S_{11} - S_{22}^* \Delta)^*}{|S_{11}|^2 - |\Delta|^2} \qquad (9.22)$$
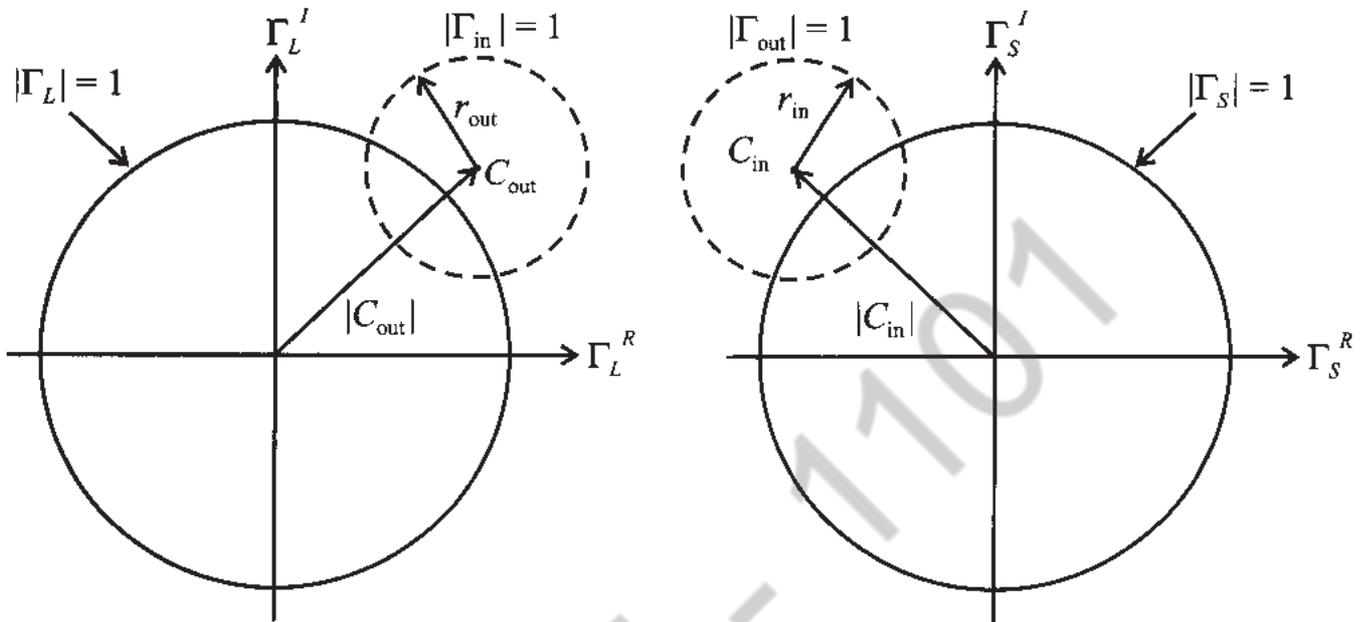
(a) Output stability circle   (b) Input stability circle

Figure 9-3   Stability circle $|\Gamma_{\text{in}}| = 1$ in the complex Γ_L plane and stability circle $|\Gamma_{\text{out}}| = 1$ in the complex Γ_S plane.

When plotted in the Γ_S-plane we obtain a response as schematically shown in Figure 9-3(b).

To interpret the meaning of Figure 9-3 correctly, a critical issue arises that is investigated for the output circle [Figure 9-3(a)], although the same argument holds for the input circle. If $\Gamma_L = 0$, then $|\Gamma_{\text{in}}| = |S_{11}|$ and two cases have to be differentiated depending on $|S_{11}| < 1$ or $|S_{11}| > 1$. For $|S_{11}| < 1$, the origin (the point $\Gamma_L = 0$) is part of the stable region, see Figure 9-4(a). However, for $|S_{11}| > 1$ the matching condition $\Gamma_L = 0$ results in $|\Gamma_{\text{in}}| = |S_{11}| > 1$, i.e. the origin is part of the unstable region. In this case the only stable region is the shaded domain between the output stability circle $|\Gamma_{\text{in}}| = 1$ and the $|\Gamma_L| = 1$ circle, see Figure 9-4(b).

For completeness, Figure 9-5 shows the two stability domains for the input stability circle. The rule-of-thumb is the inspection if $|S_{22}| < 1$, which leads to the conclusion that the center ($\Gamma_S = 0$) must be stable; otherwise the center becomes unstable for $|S_{22}| > 1$.

Care has to be exercised in correctly interpreting the stability circles if the circle radius is larger than $|C_{\text{in}}|$ or $|C_{\text{out}}|$. Figure 9-6 depicts the input stability circles for $|S_{22}| < 1$ and the two possible stability domains depending on $r_{\text{in}} < |C_{\text{in}}|$ or $r_{\text{in}} > |C_{\text{in}}|$.

(a) Shaded region is stable, since $|S_{11}| < 1$

(a) Stable region excludes the origin, $\Gamma_L = 0$, since $|S_{11}| > 1$
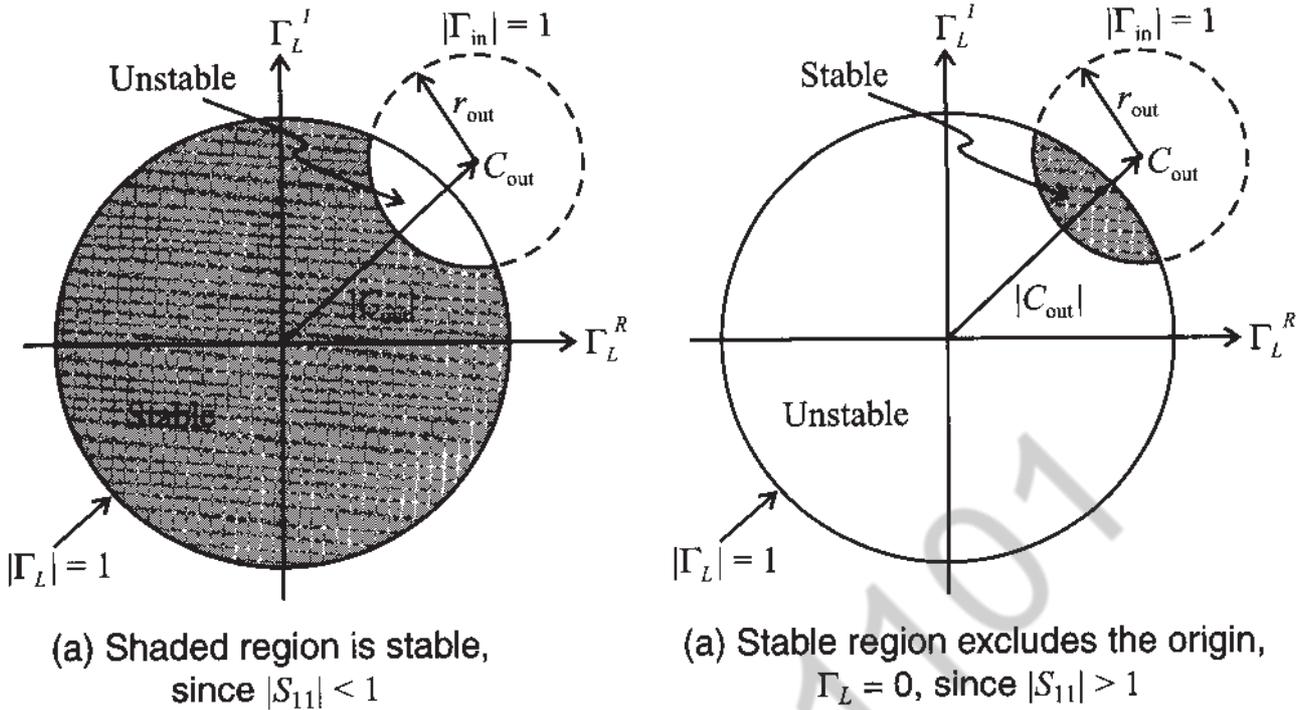
Figure 9-4   Output stability circles denoting stable and unstable regions.

(a) $|S_{22}| < 1$

(b) $|S_{22}| > 1$
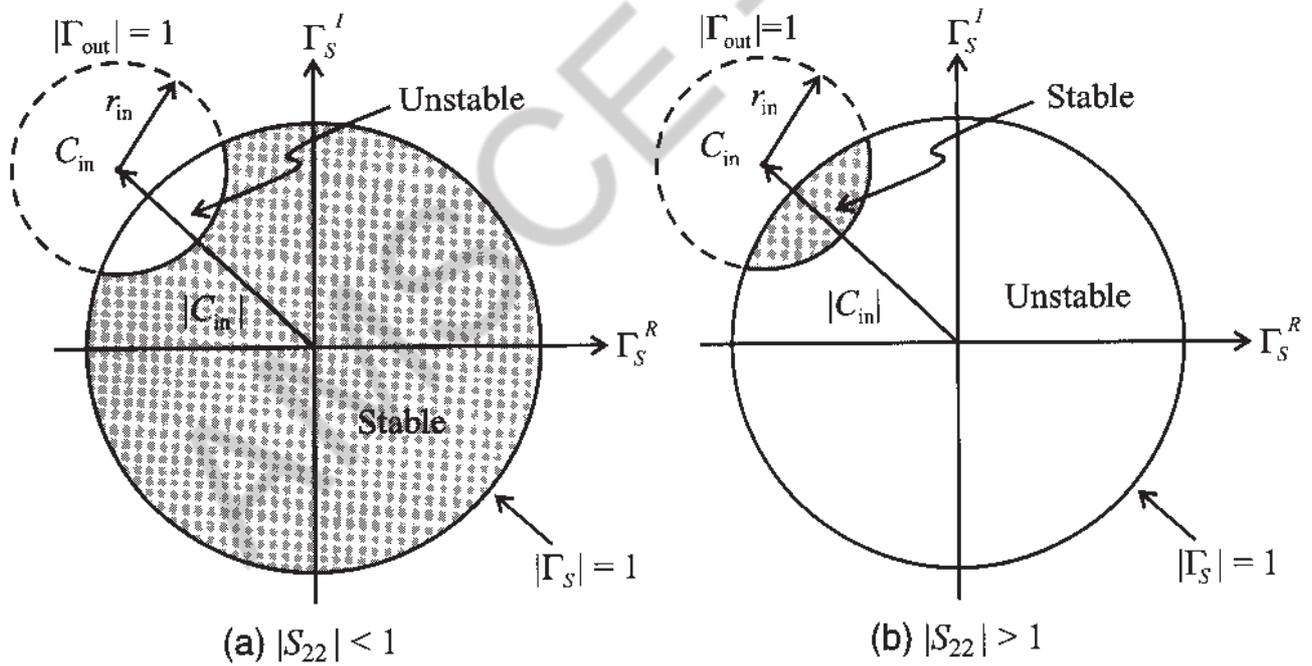
Figure 9-5   Input stability circles denoting stable and unstable regions.

9.3.2   Unconditional Stability

As the name implies, unconditional stability refers to the situation where the amplifier remains stable throughout the entire domain of the Smith Chart at the selected frequency and bias conditions. This applies to both the input and output ports. For $|S_{11}| < 1$ and $|S_{22}| < 1$ it is stated as
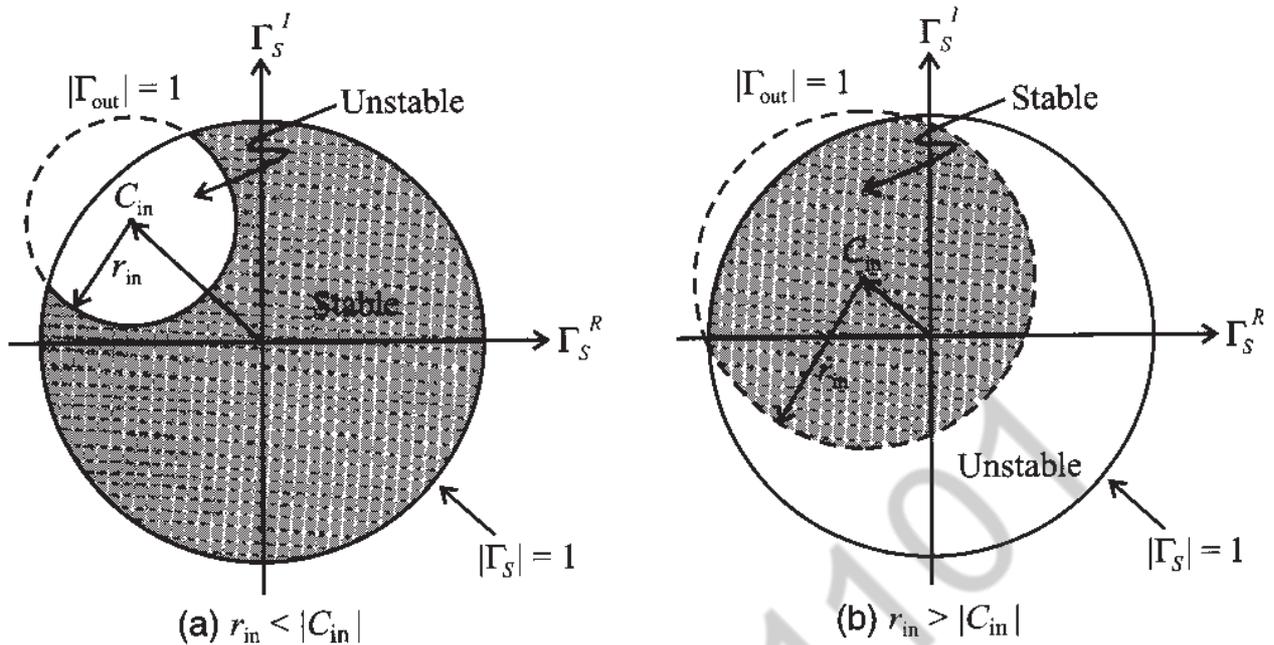
Figure 9-6   Different input stability regions for $|S_{22}| < 1$ depending on ratio between r_S and $|C_{in}|$.

$$||C_{in}| - r_{in}| > 1 \tag{9.23a}$$

$$||C_{out}| - r_{out}| > 1 \tag{9.23b}$$

In other words, the stability circles have to reside completely outside the $|\Gamma_S| = 1$ and $|\Gamma_L| = 1$ circles. In the following discussion we concentrate on the $|\Gamma_S| = 1$ circle shown in Figure 9-7(a). It is shown in Example 9-2 that condition (9.23a) can be reexpressed in terms of the stability or Rollett factor k:

$$k = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}||S_{21}|} > 1 \tag{9.24}$$

Alternatively, unconditional stability can also be viewed in terms of the Γ_S behavior in the complex $\Gamma_{out} = \Gamma_{out}^R + j\Gamma_{out}^I$ plane. Here, the $|\Gamma_S| \leq 1$ domain must reside completely within the $|\Gamma_{out}| = 1$ circle, as depicted in Figure 9-7(b). Plotting $|\Gamma_S| = 1$ in the Γ_{out} plane produces a circle whose center is located at

$$C_S = S_{22} + \frac{S_{12}S_{21}S_{11}^*}{1 - |S_{11}|^2} \tag{9.25}$$

and which possesses a radius of

$$r_S = \frac{|S_{12}S_{21}|}{1 - |S_{11}|^2} \qquad (9.26)$$
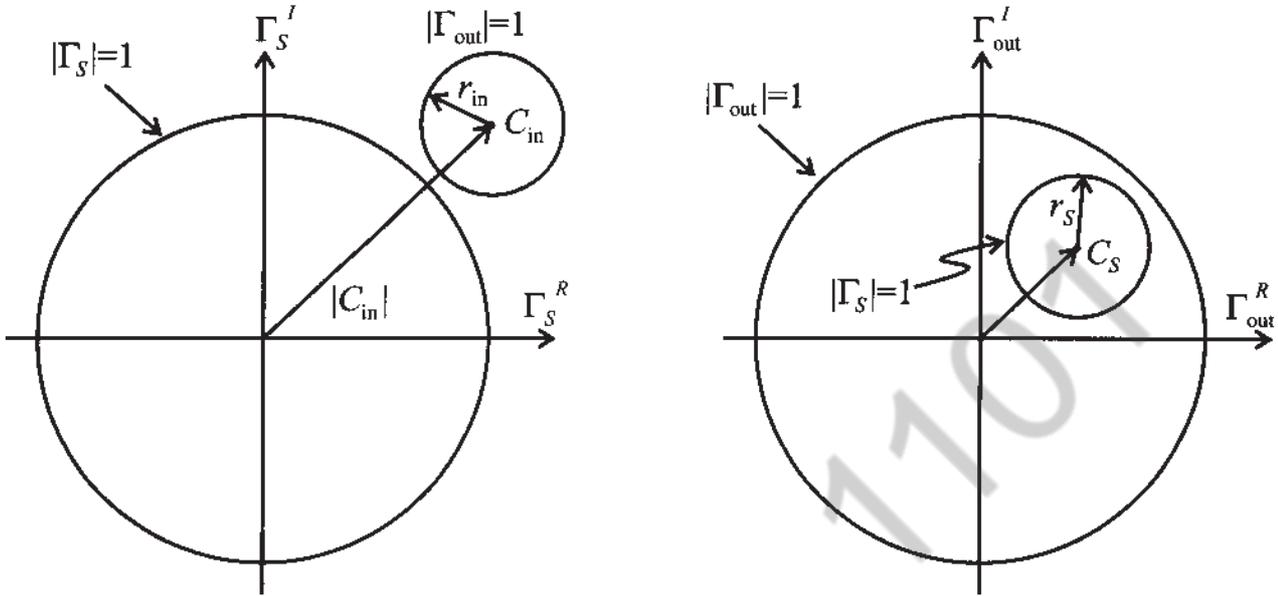
(a) $|\Gamma_{\text{out}}| = 1$ circle must reside outside   (b) $|\Gamma_S| = 1$ circle must reside inside

Figure 9-7   Unconditional stability in the Γ_S and Γ_{out} planes for $|S_{11}| < 1$.

where the condition $|C_S| + r_S < 1$ must hold. We note that (9.25) can be rewritten as $C_S = (S_{22} - \Delta S_{11}^*)/(1 - |S_{11}|^2)$. Employing $|C_S| + r_S < 1$ and (9.26) it is seen that

$$\left|S_{22} - \Delta S_{11}^*\right| + |S_{12}S_{21}| < 1 - |S_{11}|^2 \qquad (9.27a)$$

and since $|S_{12}S_{21}| \leq \left|S_{22} - \Delta S_{11}^*\right| + |S_{12}S_{21}|$ we conclude

$$|S_{12}S_{21}| < 1 - |S_{11}|^2 \qquad (9.27b)$$

A similar analysis can be established for Γ_L in the complex Γ_{in} plane. From the corresponding circle center C_L and radius r_L, we set $|C_L| = 0$ and $r_S < 1$. Thus,

$$|S_{12}S_{21}| < 1 - |S_{22}|^2 \qquad (9.28)$$

However, as long as $|\Delta| < 1$, (9.24) remains the sufficient requirement to ensure unconditional stability. This follows from the fact that when (9.27b) and (9.28) are added, it is seen that

$$2|S_{12}S_{21}| < 2 - |S_{11}|^2 - |S_{22}|^2$$

Introducing the inequality $|\Delta| = |S_{11}S_{22} - S_{12}S_{21}| \leq |S_{11}S_{22}| + |S_{12}S_{21}|$ results in

$$|\Delta| < 1 - \frac{1}{2}(|S_{11}|^2 + |S_{22}|^2 - 2|S_{11}||S_{22}|) = 1 - \frac{1}{2}(|S_{11}| - |S_{22}|)^2$$

Since $(1/2)(|S_{11}| - |S_{22}|)^2 < 1$, it is seen that (9.27b) and (9.28) are equivalent to

$$|\Delta| < 1 \tag{9.29}$$
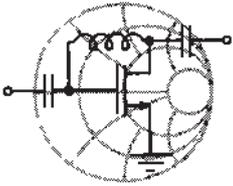
Example 9-2: Stability factor derivation

Derive the stability factor k (Rollett factor) from (9.23a).

Solution: Substituting (9.21) and (9.22) into (9.23a) gives

$$\left| \frac{\left| S_{11} - S_{22}^* \Delta \right| - |S_{12}S_{12}|}{|S_{11}|^2 - |\Delta|^2} \right| > 1 \tag{9.30a}$$

Squaring and rearranging (9.30a) results in

$$2|S_{11} - S_{22}^* \Delta||S_{12}S_{21}| < |S_{11} - S_{22}^* \Delta|^2 + |S_{12}S_{21}|^2 - \left||S_{11}|^2 - |\Delta|^2\right|^2 \tag{9.30b}$$

The term $|S_{11} - S_{22}\Delta|^2$ in (9.30b) can be re-expressed as

$$|S_{11} - S_{22}^* \Delta|^2 = |S_{12}S_{21}|^2 + (1 - |S_{22}|^2)(|S_{11}|^2 - |\Delta|^2) \tag{9.30c}$$

Squaring (9.30b) again and rearranging terms finally gives

$$(|S_{11}|^2 - |\Delta|^2)^2 \left\{ [(1 - |S_{22}|^2) - (|S_{11}|^2 - |\Delta|^2)]^2 - 4|S_{12}S_{21}|^2 \right\} > 0 \tag{9.30d}$$

The terms inside the curly brackets are recognized as the desired stability factor:

$$k = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}||S_{21}|} > 1 \tag{9.30e}$$

A stability analysis starting from (9.23b) would have resulted in exactly the same inequality. Thus, the stability factor k applies for both input and output ports.

It is always prudent to determine that both the $|\Delta| < 1$ and $k > 1$ conditions are fulfilled to ensure an unconditionally stable design. The next example investigates a transistor in common-emitter configuration in terms of its input and output stability behavior.
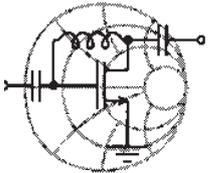
—————————————————————————————RF&MW→

Example 9-3: Stability circles for a BJT at different operating frequencies

Determine the stability regions of the bipolar junction transistor BFG505W (Philips Semiconductors) biased at $V_{CE} = 6$ V and $I_C = 4$ mA. The corresponding S-parameters as a function of frequency are given in Table 9-1.

Table 9-1   BFG505W S-parameters as a function of frequency

Frequency	S_{11}	S_{12}	S_{21}	S_{22}
500 MHz	0.70∠–57°	0.04∠47°	10.5∠136°	0.79∠–33°
750 MHz	0.56∠–78°	0.05∠33°	8.6∠122°	0.66∠–42°
1000 MHz	0.46∠–97°	0.06∠22°	7.1∠112°	0.57∠–48°
1250 MHz	0.38∠–115°	0.06∠14°	6.0∠104°	0.50∠–52°

Solution: Based on the definitions for k, $|\Delta|$, C_{in}, r_{in}, C_{out}, and r_{out}, we compute the values via a MATLAB routine (see m-file ex9_3.m). A summary of the results is given in Table 9-2 for the four frequencies listed in Table 9-1.

Table 9-2   Stability parameters for BFG505W for frequencies listed in Table 9-1

k	$|\Delta|$	C_{in}	r_{in}	C_{out}	r_{out}
0.41	0.69	39.04∠108°	38.62	3.56∠70°	3.03
0.60	0.56	62.21∠119°	61.60	4.12∠70°	3.44
0.81	0.45	206.23∠131°	205.42	4.39∠69°	3.54
1.02	0.37	42.42∠143°	41.40	4.24∠68°	3.22

The example input and output stability circles for the frequencies of $f = 750$ MHz and $f = 1.25$ GHz are shown in Figure 9-8. We notice that $|S_{11}| < 1$ and $|S_{22}| < 1$ in all cases. This implies that the $\Gamma_L = 0$ and $\Gamma_S = 0$ points are stable, indicating that the interior domain of the Smith Chart up to the stability circles denotes the stable region.
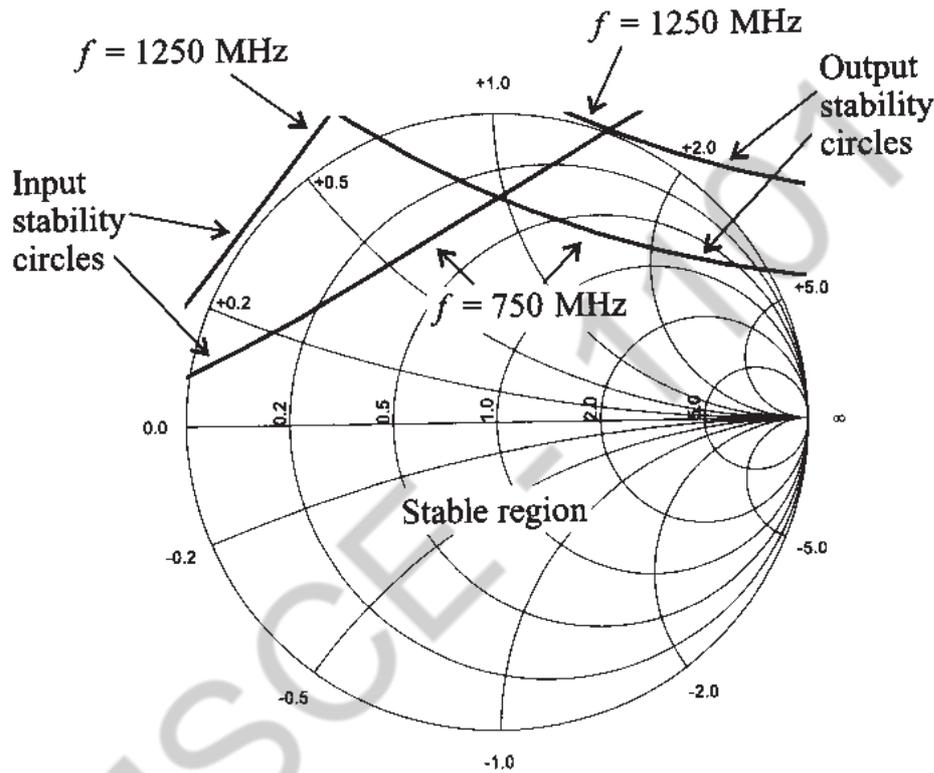
Figure 9-8   Input and output stability circles for BFG505W computed at $f = 750$ MHz and $f = 1.25$ GHz.

Also, as can be seen from Figure 9-8 and Table 9-2, the transistor is unconditionally stable at $f = 1.25$ GHz and both input and output stability circles are located completely outside of the $|\Gamma| = 1$ circle. At all other frequencies transistor is potentially unstable.

The stability circles are not only affected by frequency, but also by the bias conditions. We recall that the S-parameters are given for particular bias conditions. The entire stability analysis must be repeated if biasing, or even temperature, changes.

Even though k can vary widely, most unstable practical designs fall into the range $0 \le k \le 1$. Oscillators, discussed in Chapter 10, target the entire Smith Chart as the unstable domain, resulting in negative values of k. It is also interesting to observe that in the absence of any output to input feedback ($S_{12} = 0$) the transistor is inherently stable, since the stability factor yields $k \to \infty$. In practice, one often examines k alone without paying attention to the $|\Delta| < 1$ condition. This can cause potential problems, as the following example highlights.
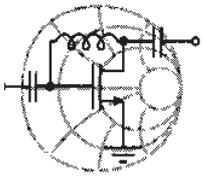
RF&MW→

Example 9-4: Stable versus unstable region of a transistor

Investigate the stability regions of a transistor whose S-parameters are recorded as follows: $S_{11} = 0.7\angle{-70°}$, $S_{12} = 0.2\angle{-10°}$, $S_{21} = 5.5\angle 85°$, and $S_{22} = 0.7\angle{-45°}$

Solution: We again compute the values k, $|\Delta|$, C_{in}, r_{in}, C_{out}, and r_{out}. The results are $k = 1.15$, $|\Delta| = 1.58$, $C_{in} = 0.21\angle 52°$, $r_{in} = 0.54$, $C_{out} = 0.21\angle 27°$, and $r_{out} = 0.54$ (see Figure 9-9). It is seen that even though $k > 1$, the transistor is still potentially unstable because $|\Delta| > 1$. This results in input and output stability
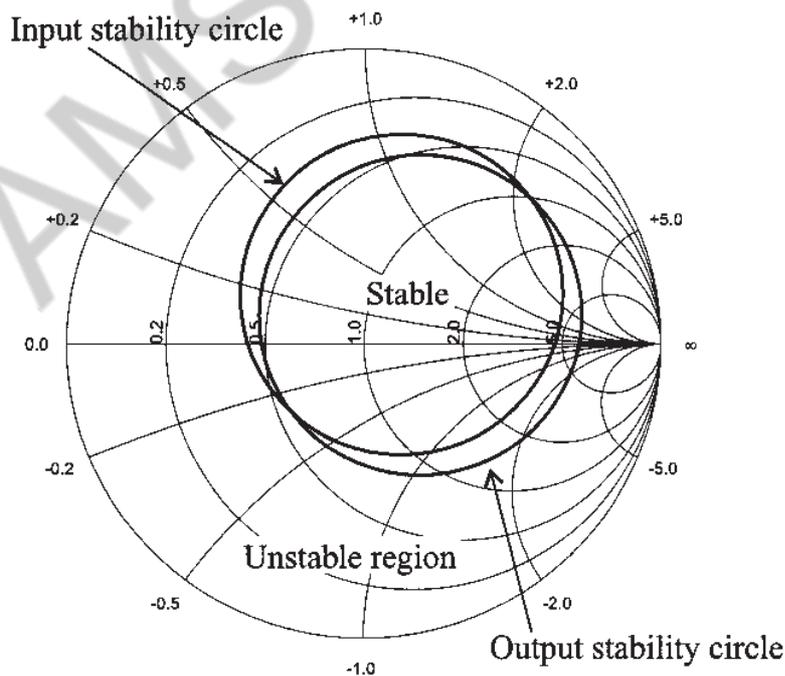
Figure 9-9 Stability circles for $k > 1$ and $|\Delta| > 1$.

circles being located inside of the Smith Chart. Since both $|S_{11}|$ and $|S_{22}|$ are less than unity, the center of the Smith Chart is a stable point. Therefore, since $|C_{in}| < r_{in}$ and $|C_{out}| < r_{out}$, the area inside of the stability circles represents the stable region, as shown in Figure 9-9.

> *Usually manufacturers avoid producing transistors with both $k > 1$ and $|\Delta| > 1$ by incorporating matching networks housed inside the transistor casing.*

9.3.3   Stabilization Methods

If the operation of a FET or BJT is found to be unstable in the desired frequency range, an attempt can be made to stabilize the transistor. We recall that $|\Gamma_{in}| > 1$ and $|\Gamma_{out}| > 1$ can be written in terms of the input and output impedances:

$$|\Gamma_{in}| = \left|\frac{Z_{in} - Z_0}{Z_{in} + Z_0}\right| > 1 \text{ and } |\Gamma_{out}| = \left|\frac{Z_{out} - Z_0}{Z_{out} + Z_0}\right| > 1$$

which imply $\text{Re}\{Z_{in}\} < 0$ and $\text{Re}\{Z_{out}\} < 0$. One way to stabilize the active device is to add a series resistance or a shunt conductance to the port. Figure 9-10 shows the configuration for the input port. This loading in conjunction with $\text{Re}\{Z_S\}$ must compensate the negative contribution of $\text{Re}\{Z_{in}\}$. Thus, we require

$$\text{Re}\{Z_{in} + R_{in}' + Z_S\} > 0 \text{ or } \text{Re}\{Y_{in} + G_{in}' + Y_S\} > 0 \tag{9.31a}$$

(a) Series resistance   (b) Shunt conductance
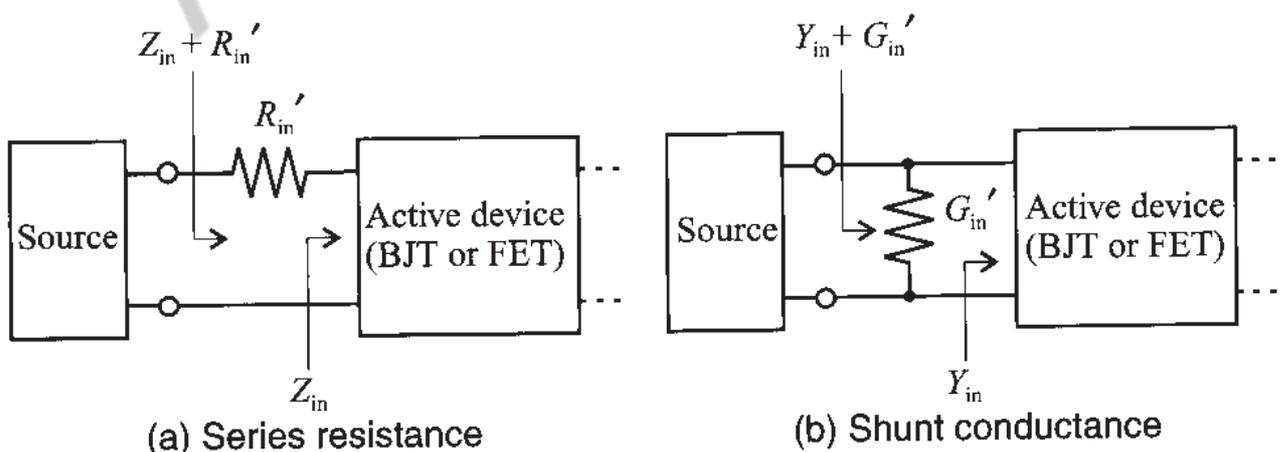
Figure 9-10   Stabilization of input port through series resistance or shunt conductance.

Following an identical argument, Figure 9-11 shows the stabilization of the output port. The corresponding condition is

$$\text{Re}\{Z_{\text{out}} + R_{\text{out}}' + Z_L\} > 0 \ \text{ or } \ \text{Re}\{Y_{\text{out}} + G_{\text{out}}' + Y_L\} > 0 \qquad (9.31b)$$

(a) Series resistance   (b) Shunt conductance
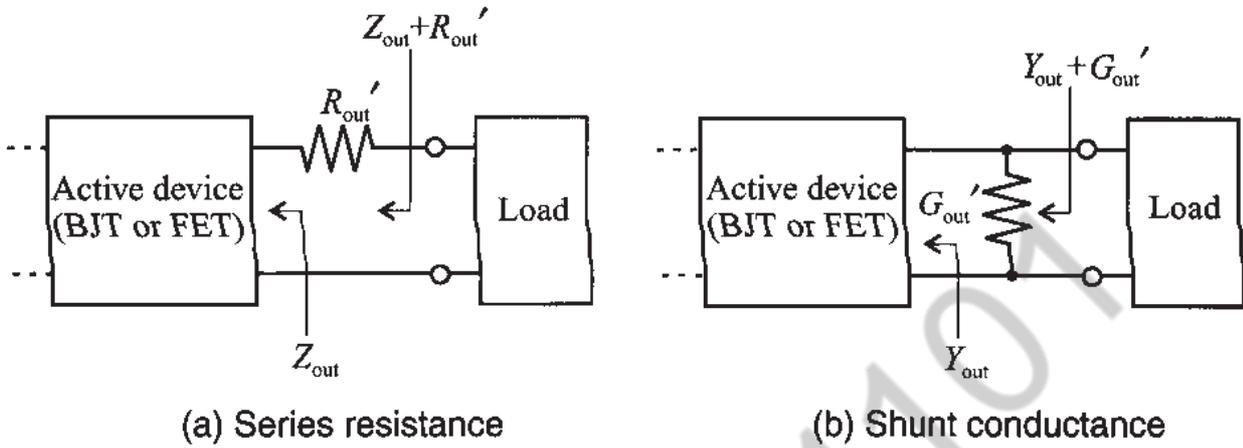
Figure 9-11   Stabilization of output port through series resistance or shunt conductance.

The next example explains the stabilization procedure for transistor.
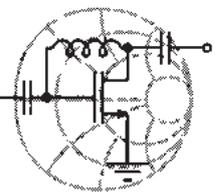
Example 9-5:   Stabilization of a BJT

Using the transistor BFG505W from Example 9-3 operated at $f = 750$ MHz (and with the S-parameters given as follows: $S_{11} = 0.56\angle{-78°}$, $S_{21} = 0.05\angle{33°}$, $S_{12} = 8.64\angle{122°}$, and $S_{22} = 0.66\angle{-42°}$), attempt to stabilize the transistor by finding a series resistor or shunt conductance for the input and output ports.

Solution: With given S-parameters we can identify the input and output stability circles by computing their radii and center positions: $C_{\text{in}} = 62.21\angle{119°}$, $r_{\text{in}} = 61.60$, and $C_{\text{out}} = 4.12\angle{70°}$, $r_{\text{out}} = 3.44$. The corresponding stability circles are shown in Figure 9-12. A constant resistance circle $r' = 0.33$ in the Z-chart indicates the minimal series resistance that has to be connected to the input of the transistor to make this port stable. If a passive network is connected in series to the resistor with the value of $R_{\text{in}}' = r'Z_0 = 16.5 \ \Omega$, then the combined impedance will be located inside of the $r' = 0.33$ circle and therefore in the stable

region. Similarly, by tracing a constant conductance circle $g' = 2.8$ we find the shunt admittance $G_{in}' = g'/Z_0 = 56$ mS that stabilizes the input of the transistor. This time any passive network connected to G_{in}' will have the combined admittance residing inside of the $g' = 2.8$ circle in the Y-chart, which is inside the stable region for the input port of the transistor.
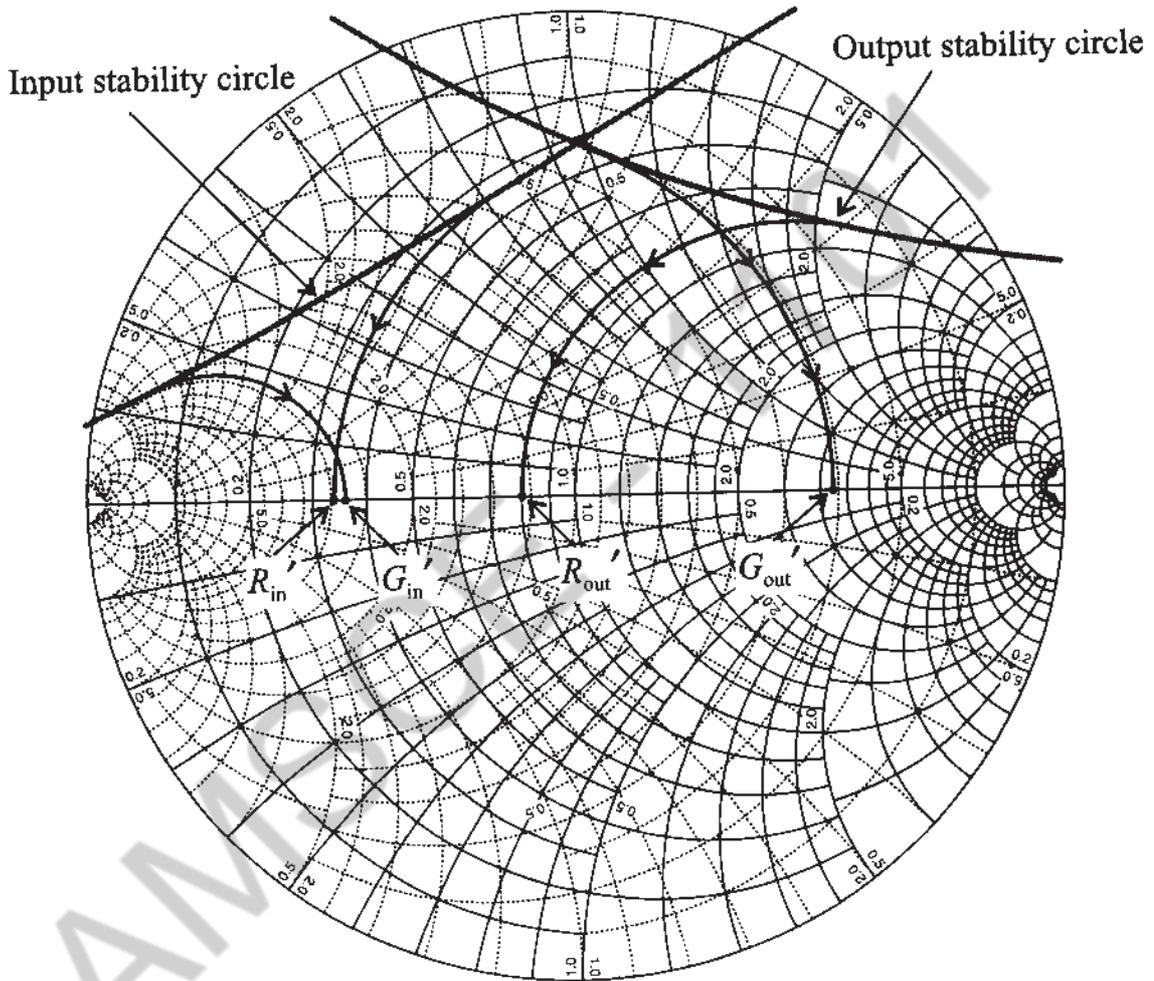
Figure 9-12 Input and output stability circles and circles for finding stabilizing series resistance and shunt conductances.

Following an identical procedure we can find a series resistance of $R_{out}' = 40 \ \Omega$ and a shunt conductance $G_{out}' = 6.2$ mS, which stabilize the output port of the transistor.

Due to the coupling between input and output ports of the transistor it is usually sufficient to stabilize one port. The choice of which port is generally up to the circuit designer. However, one

attempts to avoid resistive elements at the input port since they cause additional noise to be amplified.

Stabilization through the addition of resistors comes at a prize: the impedance matching can suffer, there may be a loss in power flow, and the noise figure typically worsens due to the additional thermal noise sources that the resistors present.

9.4 Constant Gain

9.4.1 Unilateral Design

Besides ensuring stability, the need to obtain a desired gain performance is another important consideration in the amplifier design task. If, as sometimes done in practice, the influence of the transistor's feedback is neglected ($S_{12} \approx 0$), we can employ the unilateral power gain G_{TU} described by (9.12). This equation is rewritten such that the individual contributions of the matching networks become identifiable. With reference to Figure 9-13, we write

$$G_{TU} = \frac{1 - |\Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2} \times |S_{21}|^2 \times \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_L S_{22}|^2} = G_S \times G_0 \times G_L \tag{9.32}$$

where the individual blocks are

$$G_S = \frac{1 - |\Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2}, \ G_0 = |S_{21}|^2, \ G_L = \frac{1 - |\Gamma_L|^2}{|1 - \Gamma_L S_{22}|^2} \tag{9.33}$$
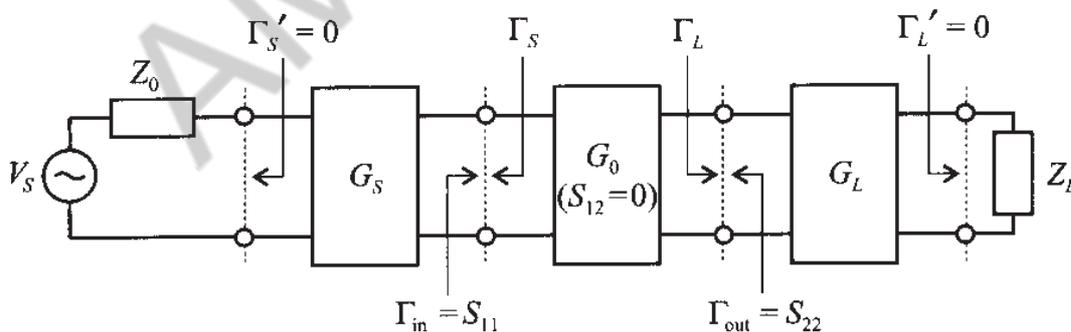
Figure 9-13 Unilateral power gain system arrangement.

Because most gain calculations are done in dB, (9.32) is frequently expressed as

with a being the radius of the YIG sphere. This also determines C_0 from the resonance condition $\omega_0^2 = 1/(L_0 C_0)$; that is,

$$C_0 = L_0 \omega_0^2 \tag{10.42b}$$

Finally, the conductance is

$$G_0 = \frac{d^2}{\mu_0 \omega_m Q_u \left(\frac{4}{3}\pi a^3\right)} \tag{10.42c}$$

In (10.42a)–(10.42c) d is the diameter of the coupling loop.

10.2.4 Voltage-Controlled Oscillator

It is mentioned in Chapter 6 that certain diodes exhibit a large change in capacitance in response to an applied bias voltage. A typical example is the varactor diode, with its variable capacitance $C_V = C_{V0}(1 - V_Q/V_{\text{diff}})^{-1/2}$ that can be affected by the reverse bias V_Q. Figure 10-25 illustrates how the feedback loop for the Clapp oscillator can be modified, by replacing C_3 in Figure 10-25(a) with the varactor diode and an appropriate DC isolation. The modified circuit is shown in Figure 10-25(b). This circuit can readily be analyzed if a simplified BJT model ($R_L \ll h_{22}$) is employed.

(a) Pi-type feedback loop   (b) Redrawn circuit with DC isolation
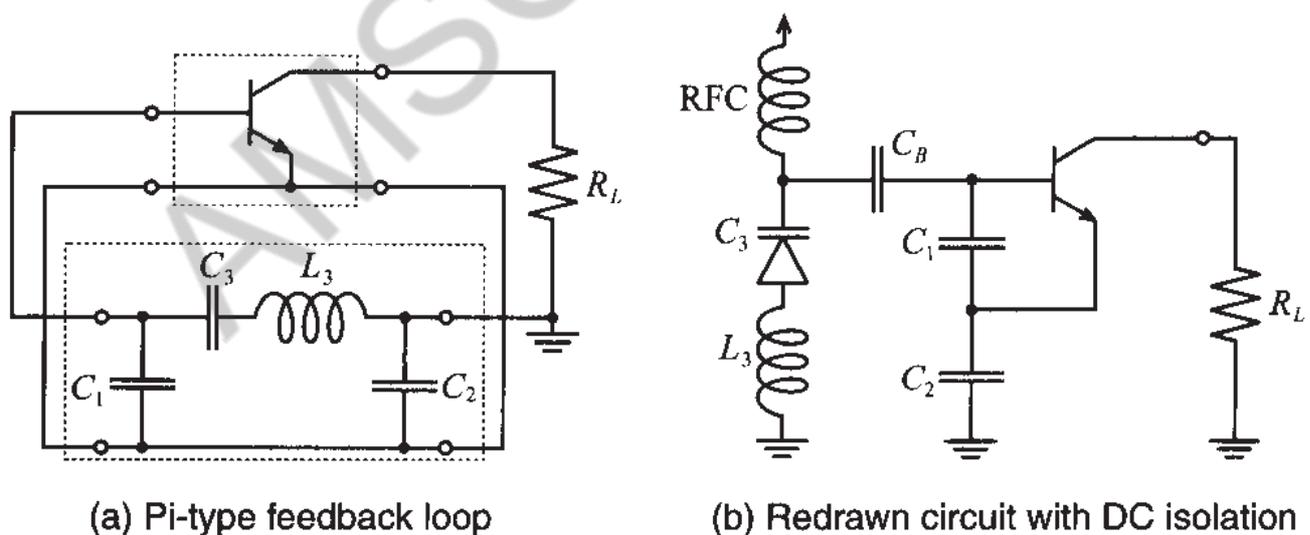
Figure 10-25   Varactor diode oscillator.

In Figure 10-26 the varactor diode and a transmission line element, whose length is adjusted to be inductive, form the termination circuit connected to the input of the oscillator. If the varactor diode and the transmission line segment is disconnected, the input impedance Z_{IN} can be computed from two loop equations:

$$v_{IN} - i_{IN}X_{C1} - i_{IN}X_{C2} + i_B X_{C1} - \beta i_B X_{C2} = 0 \qquad (10.43a)$$

$$h_{11}i_B + i_B X_{C1} - i_{IN}X_{C1} = 0 \qquad (10.43b)$$
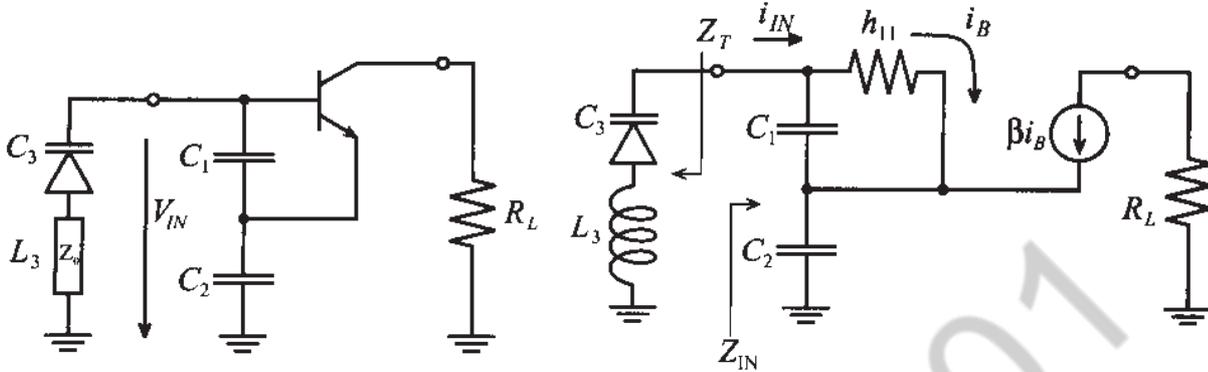
Figure 10-26   Circuit analysis of varactor diode oscillator.

Rearranging leads to

$$Z_{IN} = \frac{1}{h_{11} + X_{C1}}[h_{11}(X_{C1} + X_{C2}) + X_{C1}X_{C2}(1 + \beta)] \qquad (10.44)$$

The equation can be simplified by noting that $(1 + \beta) \approx \beta$ and assuming that $h_{11} \gg X_{C1}$, which results in

$$Z_{IN} = \frac{1}{j\omega}\left[\frac{1}{C_1} + \frac{1}{C_2}\right] - \frac{\beta}{h_{11}}\left(\frac{1}{\omega^2 C_1 C_2}\right) \qquad (10.45)$$

As expected from our previous discussion, the input resistance is negative. Therefore, with $g_m = \beta/h_{11}$,

$$R_{IN} = -\frac{g_m}{\omega^2 C_1 C_2} \qquad (10.46a)$$

and

$$X_{IN} = \frac{1}{j\omega C_{IN}} \qquad (10.46b)$$

where $C_{IN} = C_1 C_2/(C_1 + C_2)$. The resonance frequency follows from the previously established condition $X_1 + X_2 + X_3 = 0$ (see Section 10.1.2), or

$$j\left(\omega_0 L_3 - \frac{1}{\omega_0 C_3}\right) - \frac{1}{j\omega_0}\left[\frac{1}{C_1} + \frac{1}{C_2}\right] = 0 \qquad (10.47)$$

with the result

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{L_3}\left(\frac{1}{C_3}+\frac{1}{C_2}+\frac{1}{C_1}\right)} \tag{10.48}$$

It can be concluded from (10.46a) that the combined resistance of the varactor diode must be equal to or less than $|R_{IN}|$ in order to create sustained oscillations.
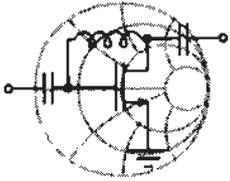
—— $\sim$RF&MW$\rightarrow$

Example 10-7: Design of a varactor-controlled oscillator

A typical varactor diode has an equivalent series resistance of 45 Ω and a capacitance ranging from 10 pF to 30 pF for reverse voltages between 30 V and 2 V. Design a voltage controlled Clapp-type oscillator with center frequency of 300 MHz and ±10% tuning capability. Assume that the transconductance of the transistor is constant and equal to $g_m = 115$ mS.

Solution: To create sustained oscillations, we have to ensure that the series resistance of the varactor diode is smaller or equal to $|R_{IN}|$ over the entire frequency range as computed in (10.46a). From (10.46a) we can conclude that $|R_{IN}|$ achieves its minimum value at maximum frequency of operation. Substituting $\omega_{max} = 2\pi f_{max}$ (with $f_{max} = 1.1f_0 = 330$ MHz being the maximum oscillation frequency) into (10.46a), it is found that the capacitances C_1 and C_2 are related as

$$C_1 = \frac{g_m}{\omega_{max}^2 R_S C_2} = \frac{1}{kC_2} = \frac{1}{1.68\times10^{21}C_2} \tag{10.49}$$

where $R_S = 45$ Ω is the varactor's equivalent series resistance.

Since the maximum oscillation is obtained when the varactor capacitance has its minimum value, and the minimum frequency corresponds to the maximum C_3, we can rewrite (10.48) as

$$f_{min} = \frac{1}{2\pi}\sqrt{\frac{1}{L_3}\left(\frac{1}{C_{3max}}+\frac{1}{C_2}+kC_2\right)} \tag{10.50}$$

$$f_{max} = \frac{1}{2\pi}\sqrt{\frac{1}{L_3}\left(\frac{1}{C_{3max}}+\frac{1}{C_2}+kC_2\right)} \tag{10.51}$$

where relation (10.49) is used to eliminate C_1. Dividing (10.50) by (10.51) and taking the square of the result, the following quadratic equation is obtained for C_2:

$$k(1 - \alpha^2)C_2^2 + \left(\frac{1}{C_{3\max}} - \frac{\alpha^2}{C_{3\min}}\right)C_2 + (1 - \alpha^2) = 0 \qquad (10.52)$$

where $\alpha = f_{\min}/f_{\max}$. Solving (10.52) and substituting the result in (10.49) and (10.50) or (10.51), we find $C_1 = 12.4$ pF, $C_2 = 48$ pF, and $L_3 = 46.9$ nH as our desired values.

> *Unlike a mechanically adjustable dielectric resonator, the varactor diode permits dynamic tuning over a substantial frequency range.*

10.2.5 Gunn Element Oscillator

The **Gunn element** can be employed to create oscillators from 1 to 100 GHz at low power outputs of roughly up to 1 W. It exploits a unique negative resistance phenomenon first discovered by Gunn in 1963. When certain semiconductor structures are subjected to an increasing electric field, they begin to shift, or transfer, electrons from the main valley to side valleys in the energy band structure. The accumulation of up to 90–95% of the electron concentration into these valleys results in a substantial decrease in effective carrier mobility and produces a technologically interesting *I-V* characteristic. Semiconductors with these band structures are primarily GaAs and InP. Figure 10-27 depicts a Gunn element and its current versus applied voltage response.

We notice that in the presence of an applied DC voltage to the Gunn element it behaves like a normal ohmic contact resistor for low field strength. However, if a certain threshold voltage V_0 is exceeded, **dipole domains** begin to be created below the cathode triggered by doping fluctuations. The formation of these domains lowers the current, as indicated in Figure 10-27 (b). The current then remains constant while the domains travel from cathode to anode. After collection, the process repeats itself. The frequency can be estimated from the drift velocity of the domain motion $v_d \approx 10^5$ m/s and the travel length L of the active zone of the Gunn element. For a length of 10 µm, we obtain

$$f = \frac{v_d}{L} = \frac{10^5 \text{ m/s}}{10 \times 10^{-6} \text{ m}} = 10 \text{ GHz} \qquad (10.53)$$

(a) Gunn element structure   (b) Current vs. applied voltage response
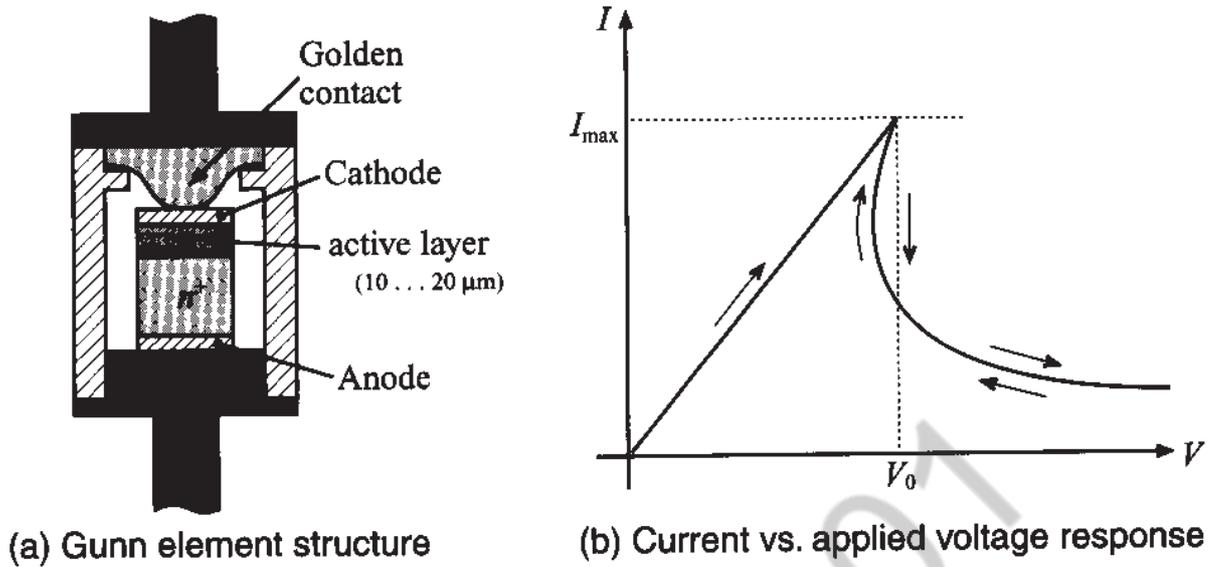
Figure 10-27   Gunn element and current versus voltage response.

If an external DC voltage is applied, the domain motion can be influenced and thus the resonance frequency is varied. The tuning range is approximately within 1% of the resonance frequency.

Figure 10-28 shows a microstrip line implementation of a Gunn element oscillator. Here the Gunn element is connected to a $\lambda/4$ microstrip line, which in turn is coupled to a dielectric resonator. The bias voltage for the Gunn element is fed through an RFC onto the microstrip line.
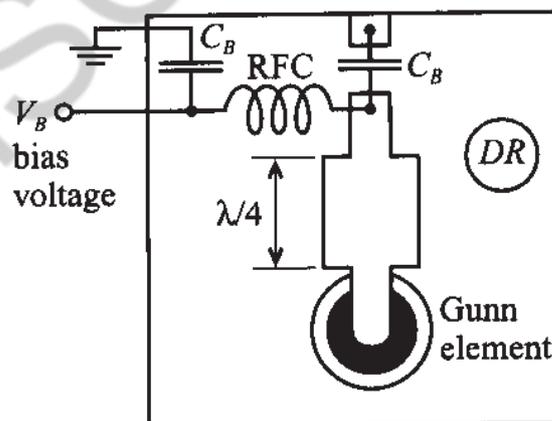
Figure 10-28   Gunn element oscillator circuit with dielectric resonator (DR).

10.3 Basic Characteristics of Mixers

Mixers are commonly used to multiply signals of different frequencies in an effort to achieve frequency translation. The motivation for this translation stems from the fact that filtering out a particular RF signal channel centered among many densely populated, narrowly spaced neighboring channels would require extremely high Q filters.

The task, however, becomes much more manageable if the RF signal carrier frequency can be reduced or downconverted within the communication system. Perhaps one of the best known systems is the downconversion in a **heterodyne receiver**, schematically depicted in Figure 10-29.
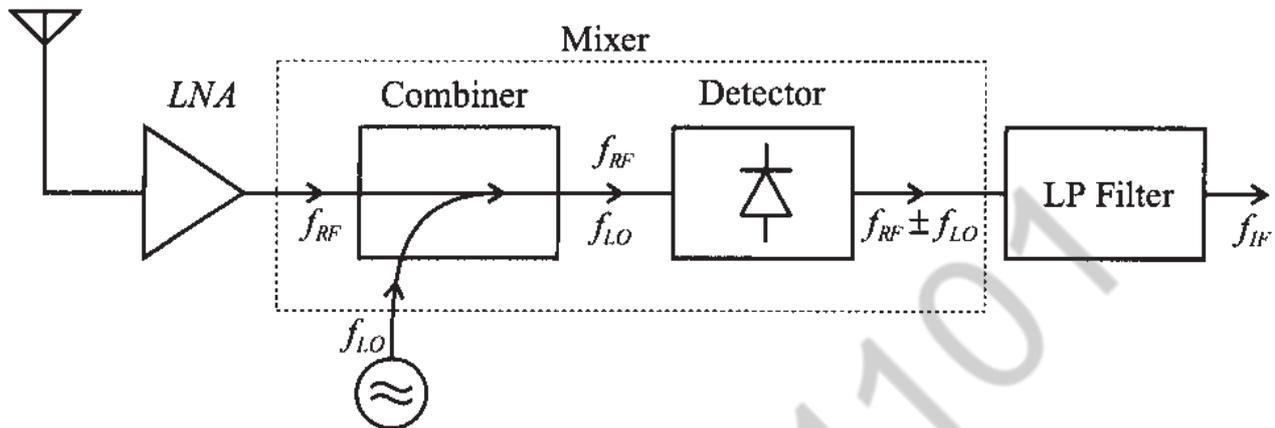
Figure 10-29 Heterodyne receiver system incorporating a mixer.

Here the received RF signal is, after preamplification in a low-noise amplifier (LNA), supplied to a mixer whose task is to multiply the input signal of center frequency f_{RF} with a **local oscillator** (LO) frequency f_{LO}. The signal obtained after the **mixer** contains the frequencies $f_{RF} \pm f_{LO}$, of which, after low-pass (LP) filtering, the lower frequency component $f_{RF} - f_{LO}$, known as the **intermediate frequency** (IF), is selected for further processing.

The two key ingredients constituting a mixer are the **combiner** and **detector**. The combiner can be implemented through the use of a $90°$ (or $180°$) directional coupler. A discussion of couplers and hybrids is found in Appendix G. The detector traditionally employs a single diode as a nonlinear device. However, antiparallel dual diode and double-balanced quadrupole diode configurations are also utilized, as discussed later. In addition to diodes, BJT and MESFET mixers with low noise figure and high conversion gain, have been designed up to the X-band.

10.3.1 Basic Concepts

Before going into details of the circuit design, let us briefly review how a mixer is capable of taking two frequencies at its input and producing multiple frequency components at the output. Clearly a linear system cannot achieve such a task, and we need to select a nonlinear device such as a diode, FET, or BJT that can generate multiple harmonics. Figure 10-30 depicts the basic system arrangement of a mixer connected to an RF signal, $V_{RF}(t)$, and local oscillator signal, $V_{LO}(t)$, which is also known as the **pump signal**.
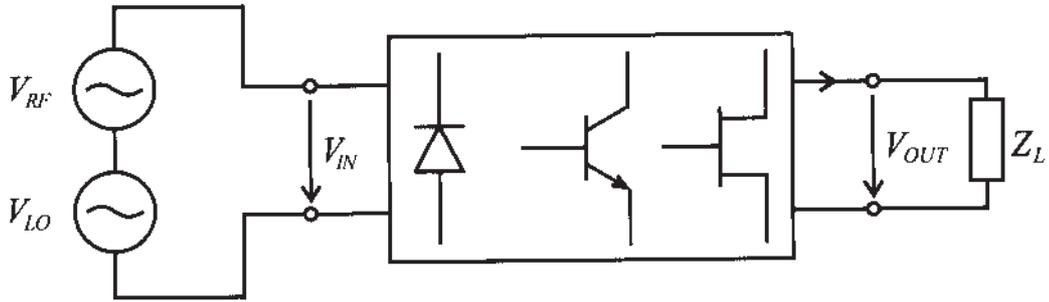
Figure 10-30   Basic mixer concept: two input frequencies are used to create new frequencies at the output of the system.

It is seen that the RF input voltage signal is combined with the LO signal and supplied to a semiconductor device with a nonlinear transfer characteristic at its output side driving a current into the load. Both diode and BJT have an exponential transfer characteristic, as expressed for instance by the Shockley diode equation discussed in Chapter 6:

$$I = I_0(e^{V/V_T} - 1) \tag{10.54a}$$

Alternatively, for a MESFET we have approximately a square behavior:

$$I(V) = I_{DSS}(1 - V/V_{T0})^2 \tag{10.54b}$$

where the subscripts denoting drain current and gate-source voltage are omitted for simplicity. The input voltage is represented as the sum of the RF signal $v_{RF} = V_{RF}\cos(\omega_{RF}t)$ and the LO signal $v_{LO} = V_{LO}\cos(\omega_{LO}t)$ and a bias V_Q; that is,

$$V = V_Q + V_{RF}\cos(\omega_{RF}t) + V_{LO}\cos(\omega_{LO}t) \tag{10.55}$$

This voltage is applied to the nonlinear device whose current output characteristic can be found via a Taylor series expansion around the Q-point:

$$I(V) = I_Q + V\left(\frac{dI}{dV}\right)\bigg|_{V_Q} + \frac{1}{2}V^2\left(\frac{d^2I}{dV^2}\right)\bigg|_{V_Q} + \dots = I_Q + VA + V^2B + \dots \tag{10.56}$$

where the constants A and B refer to $(dI/dV)|_{V_Q}$ and $(1/2)(d^2I/dV^2)|_{V_Q}$, respectively. Neglecting the constant bias V_Q and I_Q, the substitution of (10.55) into (10.56) yields

$$
\begin{aligned}
I(V) = &\ A\{V_{RF}\cos(\omega_{RF}t) + V_{LO}\cos(\omega_{LO}t)\} \\
&+ B\{V_{RF}^2\cos^2(\omega_{RF}t) + V_{LO}^2\cos^2(\omega_{LO}t)\} \\
&+ 2BV_{RF}V_{LO}\cos(\omega_{RF}t)\cos(\omega_{LO}t) + \dots
\end{aligned} \tag{10.57}
$$

The factors containing the cosine square terms can be rewritten, via the trigono-metric identity $\cos^2(\omega t) = (1/2)\{1 - \cos(2\omega t)\}$, into DC terms and terms involving $2\omega_{RF}t$ and $2\omega_{LO}t$. The key lies in the last term of (10.57), which becomes

$$I(V) = \ldots + BV_{RF}V_{LO}\{\cos[(\omega_{RF} + \omega_{LO})t] + \cos[(\omega_{RF} - \omega_{LO})t]\} \qquad (10.58)$$

This expression makes clear that the nonlinear action of a diode or transistor can gener-ate new frequency components of the form $\omega_{RF} \pm \omega_{LO}$. It is also noted that the ampli-tudes are multiplied by $V_{RF}V_{LO}$, and B is a device-dependent factor.

Equation (10.58) is the Taylor series representation up to the third term, and thus up to **second-order intermodular product** (V^2B). Any higher-order products, such as **third-order intermodular product** (V^3C), are neglected. For diodes and BJTs these higher-order harmonic terms can significantly affect the performance of a mixer. How-ever, the second-order intermodular product is the only surviving term if a FET with quadratic transfer characteristic is utilized. Thus, a FET is less prone to generate undes-ired higher-order intermodular products.

The following example discusses the down conversion process from a given RF signal frequency to a desired intermediate frequency.

$\longrightarrow$ RF&MW $\rightarrow$
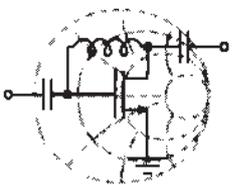
Example 10-8: Local oscillator frequency selection

An RF channel with a center frequency of 1.89 GHz and bandwidth of 20 MHz is to be downconverted to an IF of 200 MHz. Select an appropriate f_{LO}. Find the quality factor Q of a bandpass filter to select this channel if no downconversion is involved, and determine the Q of the bandpass filter after downconversion.

Solution: As seen in (10.58), by mixing RF and LO frequencies through a nonlinear device we produce an IF frequency that is equal to either $f_{IF} = f_{RF} - f_{LO}$ or $f_{IF} = f_{LO} - f_{RF}$, depending on whether f_{RF} or f_{LO} is higher. Thus, to produce a $f_{IF} = 200$ MHz from $f_{RF} = 1.89$ GHz we can use either

$$f_{LO} = f_{RF} - f_{IF} = 1.69 \text{ GHz or } f_{LO} = f_{RF} + f_{IF} = 2.09 \text{ GHz}$$

These two choices are equally valid and are both used in practice. When $f_{RF} > f_{LO}$ is chosen, the mixer is said to have **low-side**

injection, whereas when $f_{RF} < f_{LO}$ the design is called **high-side injection**. The first approach is generally preferred since lower LO frequencies are easier to generate and process.

Before down conversion, the signal has a bandwidth of $BW = 20$ MHz at a center frequency of $f_{RF} = 1.89$ GHz. Therefore, if we attempted to filter out the desired signal we would have to use a filter with $Q = f_{RF}/BW = 94.5$. However, after downconversion, the bandwidth of the signal does not change but the center frequency shifts to $f_{IF} = 200$ MHz, thus requiring a bandpass filter with a quality factor of only $Q = f_{IF}/BW = 10$.

This example shows that less selective filtering is required once the mixer has downconverted the RF signal.

10.3.2 Frequency Domain Considerations

It is important to place the previous section into a frequency domain perspective. To this end it is assumed that the angular RF signal is centered at ω_{RF} with two extra frequency components situated ω_W above and below ω_{RF}. The LO signal contains one single component at ω_{LO}. After performing mixing, according to (10.58), the resulting spectral representation contains both **upconverted** and **downconverted** frequency components. Figure 10-31 graphically explains this process.

Typically the upconversion process is associated with the modulation in a **transmitter**, whereas the downconversion is encountered in a **receiver**. Specifically, when dealing with modulation, the following terminology is common:

- Lower sideband, or LSB ($\omega_{RF} - \omega_{LO}$)

- Upper sideband, or USB ($\omega_{RF} + \omega_{LO}$)

- Double sideband, or DSB ($\omega_{RF} + \omega_{LO}, \omega_{RF} - \omega_{LO}$)

A critical question to answer is the choice of an LO frequency that shifts the RF frequency to a suitable IF level.

An interrelated issue is the problem of **image** frequencies mapping into the same downconverted frequency range. To understand this problem, assume an RF signal is downconverted with a given LO frequency. In addition to the desired signal, we have

(a) RF signal

(b) LO signal

(c) Down- and upconverted spectral products
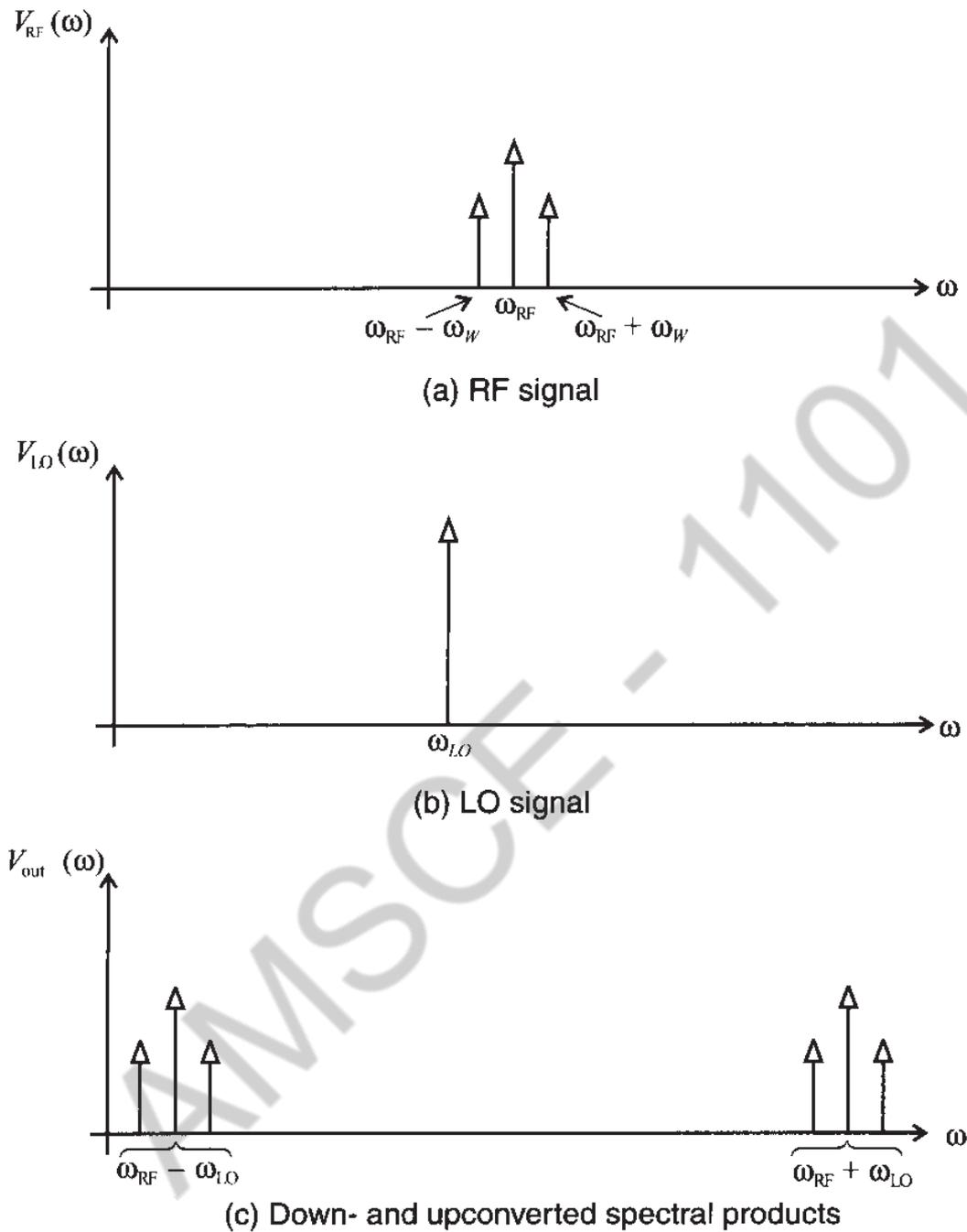
Figure 10-31   Spectral representation of mixing process.

placed symmetrically an interferer about IF (see Figure 10-32). The desired RF signal transforms as expected:

$$\omega_{RF} - \omega_{LO} = \omega_{IF} \tag{10.59a}$$

However, the image frequency ω_{IM} transforms as

$$\omega_{IM} - \omega_{LO} = (\omega_{LO} - \omega_{IF}) - \omega_{LO} = -\omega_{IF} \tag{10.59b}$$

Since $\cos(-\omega_{IF}t) = \cos(\omega_{IF}t)$, we see that both frequency spectra are shifted to the same frequency location, as Figure 10-32 illustrates.
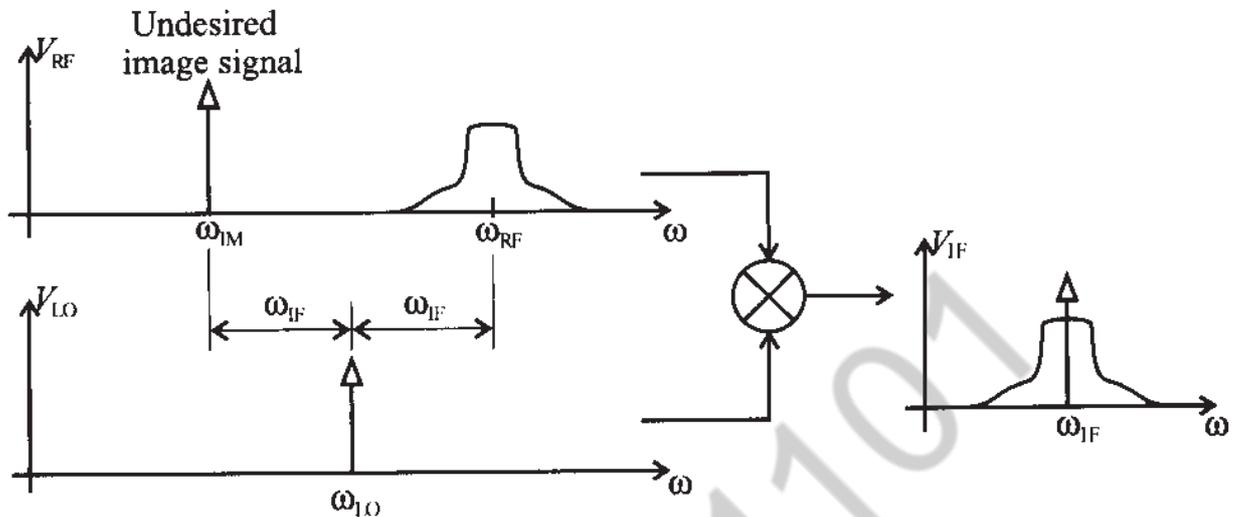
Figure 10-32   Problem of image frequency mapping.

To avoid the presence of undesired image signals that can be greater in magnitude than the RF signal, a so-called **image filter** is placed before the mixer circuit to suppress this influence, provided sufficient spectral separation is assured. More sophisticated measures involve an image rejection mixer.

10.3.3 Single-Ended Mixer Design

The simplest and least efficient mixer is the single-ended design involving a Schottky diode, as shown in Figure 10-33(a). The RF and LO sources are supplied to an appropriately biased diode followed by a resonator circuit tuned to the desired IF. In contrast, Figure 10-33(b) shows an improved design involving a FET, which, unlike the diode, is able to provide a gain to the incoming RF and LO signals.

In both cases the combined RF and LO signal is subjected to a nonlinear device with exponential (diode) or nearly quadratic (FET) transfer characteristic followed by a bandpass filter whose task is to isolate the IF signal. The two very different mixer realizations allow us to contrast a number of parameters important when developing suitable designs:

- Conversion loss or gain between the RF and IF signal powers
- Noise figure
- Isolation between LO and RF signal ports
- Nonlinearity

(a) Diode mixer

(b) FET mixer
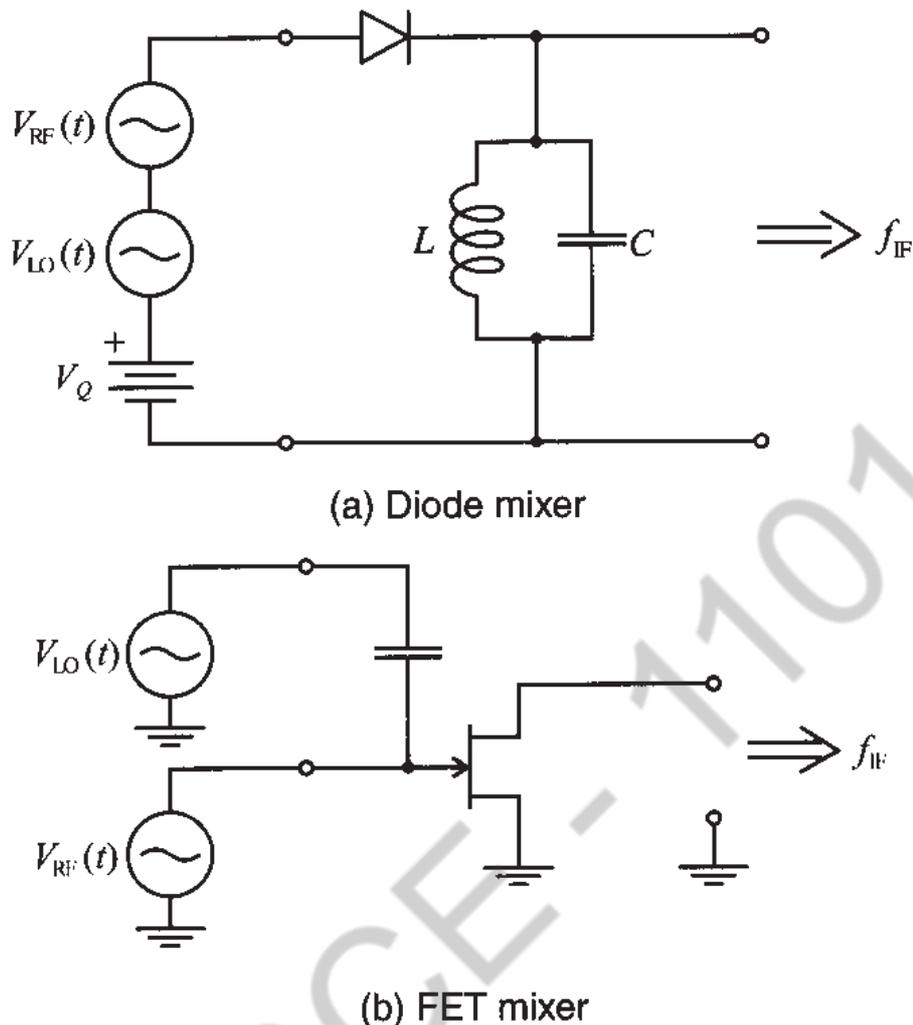
Figure 10-33 Two single-ended mixer types.

Since LO and RF signals are not electrically separated in Figure 10-33(a), there is the potential danger that the LO signal can interfere with the RF reception, possibly even reradiating portions of the LO energy through the receiving antenna. The FET realization in Figure 10-33(b) allows not only for LO and RF isolation but also provides signal gain and thus minimizes conversion loss. The **conversion loss** (CL) of a mixer is generally defined in dB as the ratio of supplied input power P_{RF} over the obtained IF power P_{IF}:

$$CL = 10\log\left(\frac{P_{RF}}{P_{IF}}\right) \tag{10.60}$$

When dealing with BJTs and FETs, it is preferable to specify a **conversion gain** (CG) defined as the inverse of the power ratio.

Additionally, the noise figure of a mixer is generically defined as

$$F = \frac{P_{n_{out}}}{CGP_{n_{in}}} \tag{10.61}$$

with CG again being the conversion gain, and $P_{n_{out}}$, $P_{n_{in}}$ the noise power at the output due to the RF signal input (at RF) and the total noise power at the output (at IF). The FET generally has a lower noise figure than a BJT, and because of a nearly quadratic transfer characteristic (see Section 7.2) the influence of higher-order nonlinear terms is minimized. Instead of the FET design, a BJT finds application when high conversion gain and low voltage bias conditions are needed (for instance, for systems relying on battery operation).

Nonlinearities are customarily quantified in terms of **conversion compression** and **intermodular distortion** (IMD). Conversion compression relates to the fact that the IF output power as a function of RF input power begins to deviate from the linear curve at a certain point. The point where the deviation reaches 1 dB is a typical mixer performance specification. As already encountered in the amplifier discussion, the intermodulation distortion is related to the influence of a second frequency component in the RF input signal, giving rise to distortion. To quantify this influence, a two-tone test is typically employed. If f_{RF} is the desired signal and f_2 is a second input frequency, then the mixing process produces a frequency component at $2f_2 - f_{RF} \pm f_{LO}$, where the +/– sign denotes up- or downconversion. The influence of this intermodulation product can be plotted in the same graph as the conversion compression (see Figure 10-34).
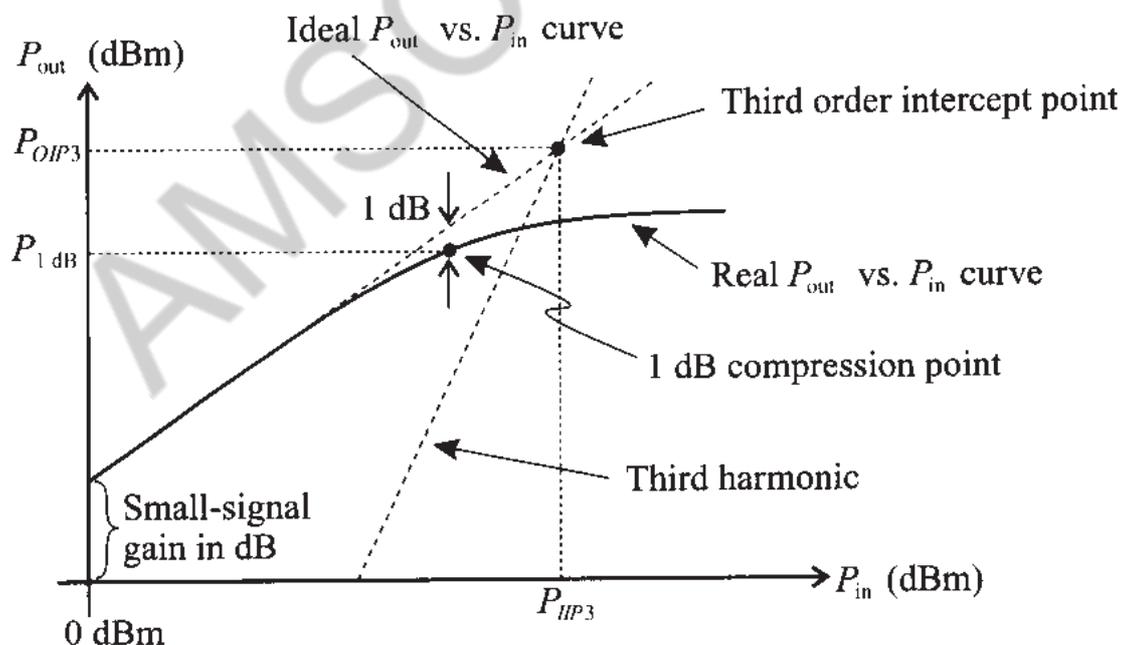
Figure 10-34   Conversion compression and intermodulation product of a mixer.

The intercept point between the desired linear output response and the undesirable third-order IMD response is a common figure of merit, indicating the ability of a mixer to suppress this influence.

Additional mixer definitions include distortion generated inside the mixer which is defined as **harmonic IMD**; **isolation** between RF and IF ports, which is directly linked to the influence of the combiner (hybrid coupler; see Appendix G); and **dynamic range**, which specifies the amplitude range over which no performance degradation occurs.

The circuit design of an RF mixer follows a similar approach as discussed when dealing with an RF amplifier. The RF and LO signals are supplied to the input of an appropriately biased transistor or diode. The matching techniques of the input and output side are presented in Chapter 8 and directly apply for mixers as well. However, one has to pay special attention to the fact that there is a large difference in frequencies between RF, LO on the input side, and IF on the output side. Since both sides have to be matched to the typical 50 Ω line impedance, the transistor port impedances (or S-parameter representation) at these two different frequencies have to be specified. Furthermore, to minimize interference at the output side of the device, it is important to short circuit the input to IF, and conversely short circuit the output to RF (see Figure 10-35). Including these requirements as part of the matching networks is not always an easy task.
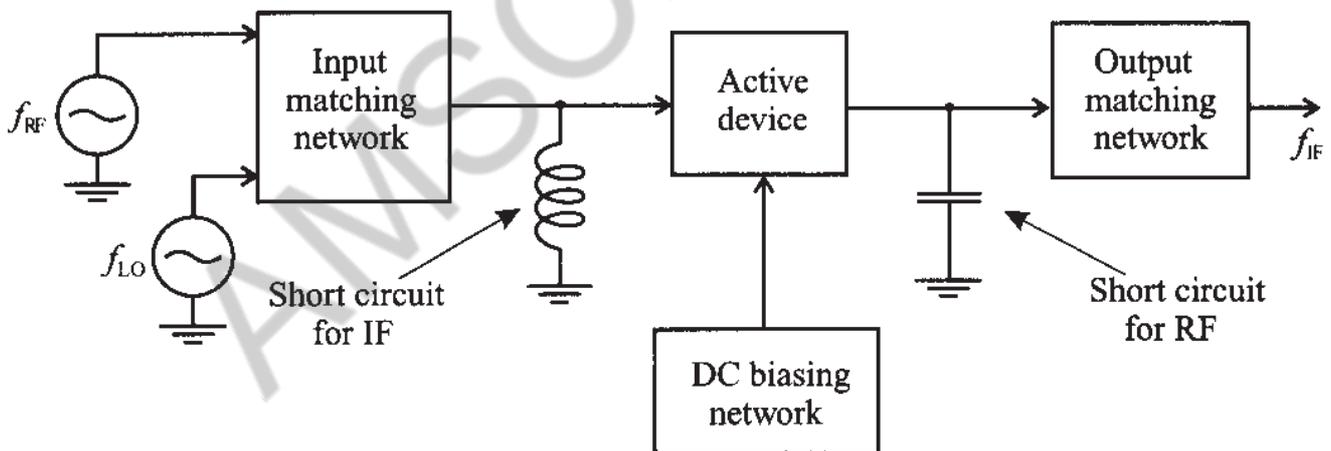
Figure 10-35   General single-ended mixer design approach.

These short-circuit conditions in general affect the transistor's behavior through internal feedback mechanism. Ideally, $\Gamma_{in}(\omega_{RF})$ should be known based on the short-circuit output condition and similarly $\Gamma_{out}(\omega_{IF})$ requires a short-circuit input condition. Typically, an additional load resistance is added to the output port to adjust the conversion gain. In the following example, the salient design steps are explained.
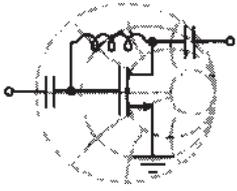
Example 10-9: Design of a single-ended BJT mixer

For the DC-biasing topology shown in Figure 10-36, compute the values of the resistors R_1 and R_2 such that biasing conditions are satisfied. Using this network as a starting point, design a low-side injection mixer for $f_{RF} = 1900$ MHz and $f_{IF} = 200$ MHz. The BJT is measured at IF to have an output impedance of $Z_{out} = (677.7 - j2324)\Omega$ for short-circuit input, and an input impedance of $Z_{in} = (77.9 - j130.6)\Omega$ for short-circuit output at RF frequency. Attempt to minimize the component count in this design.
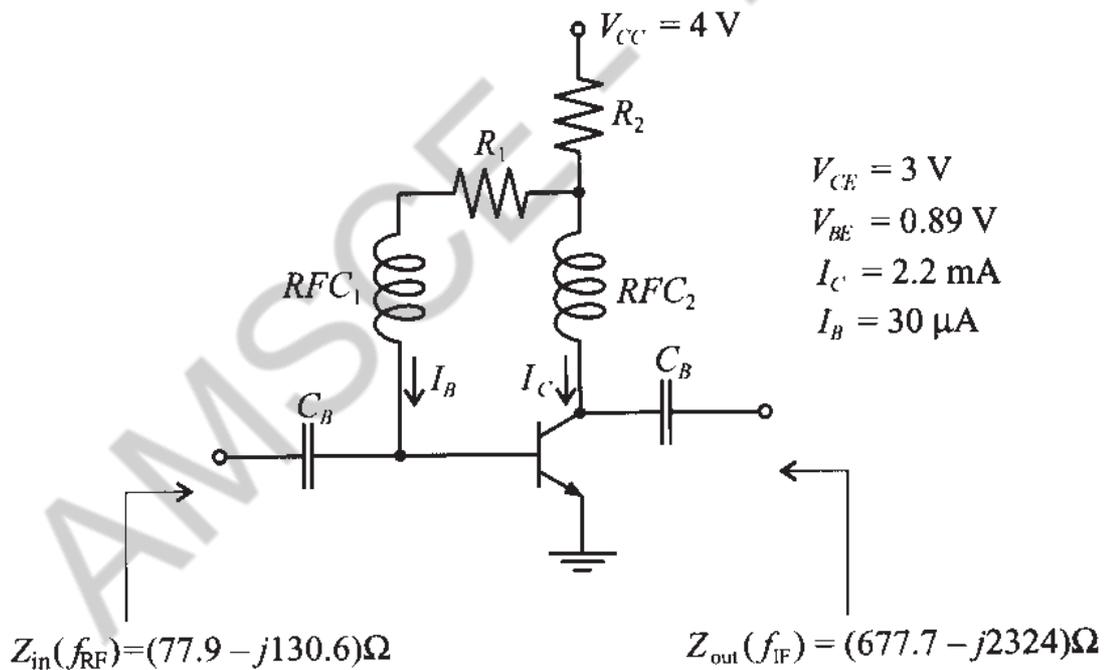
Figure 10-36   DC-biasing network for BJT mixer design.

Solution: Since the voltage drop across resistor R_2 is equal to the difference between V_{CC} and V_{CE} and the current is the sum of the base and collector currents, R_2 is computed as

$$R_2 = \frac{V_{CC} - V_{CE}}{I_C + I_B} = 448 \ \Omega$$

Similarly, the base resistor R_1 is computed as a ratio of $V_{CE} - V_{BE}$ over the base current:

$$R_1 = \frac{V_{CE} - V_{BE}}{I_B} = 70.3 \text{ k}\Omega$$

Before beginning the design of an input matching network, we have to decide on how to supply the LO signal. The simplest arrangement is to connect the LO source directly to the base of the transistor via a decoupling capacitor, as shown in Figure 10-37.
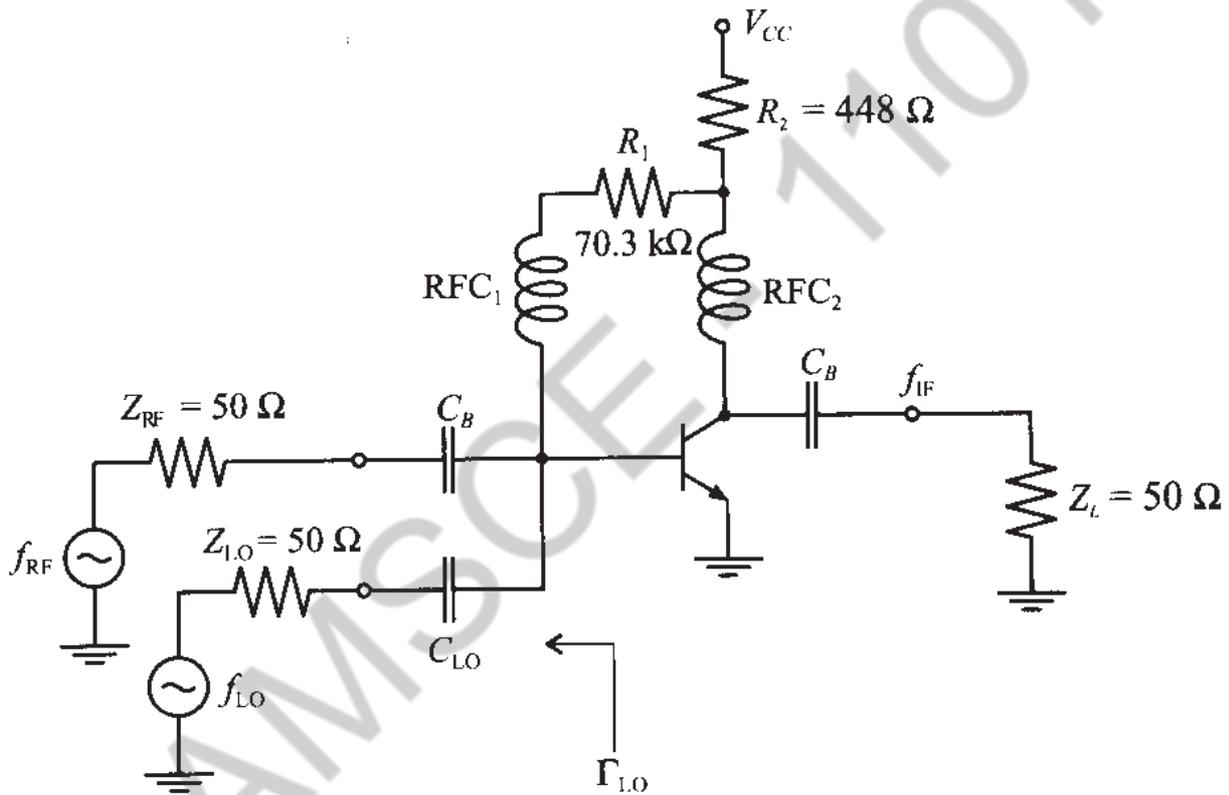
Figure 10-37 Connection of RF and LO sources to the BJT.

The value of this capacitor C_{LO} has to be chosen small enough so as to prevent RF signal coupling into the LO source. We arbitrary pick $C_{LO} = 0.2$ pF. In this case the series combination of C_{LO} and Z_{LO} creates a return loss RL_{RF} of only 0.24 dB, since

$$RL_{RF} = -20\log\left|\Gamma_{LO}\right|_{f_{RF}}\Big| = -20\log(0.9727) = 0.24 \text{ dB}$$

Unfortunately, the LO frequency is very close to f_{RF} so that the same capacitance will attenuate not only the RF signal but the LO as well. We can compute the insertion loss IL_{RF} due to this capacitor at $f_{LO} = f_{RF} - f_{IF}$

$$IL_{\text{RF}} = -10\log\left(1 - |\Gamma_{\text{LO}}|^2\big|_{f_{\text{LO}}}\right) = 13.6 \text{ dB}$$

Thus, if the LO source pumps at –20 dBm, only –33.6 dBm reaches the transistor. This seemingly high power loss is still tolerable since we can adjust the power provided by the local oscillator.

The presence of C_{LO} and Z_{LO} modifies the value of the input impedance. A new total input impedance Z'_{in} can be computed as a parallel combination of C_{LO} and Z_{LO}, and the input impedance of the transistor connected to the LO source is

$$Z'_{\text{in}} = \left(Z_{\text{LO}} + \frac{1}{j\omega_{\text{RF}}C_{\text{LO}}}\right) \| Z_{\text{in}} = (47.2 - j103.5)\Omega$$

The output impedance does not change since the input is shorted during the measurement of Z_{out}.

Knowing Z'_{in}, we can next design an input matching network using any of the methods described in Chapter 8. One of the possible topologies consists of a shunt inductor followed by a series capacitor, as shown in Figure 10-38, where we added the blocking capacitor C_{B1} to prevent DC short circuit to ground.
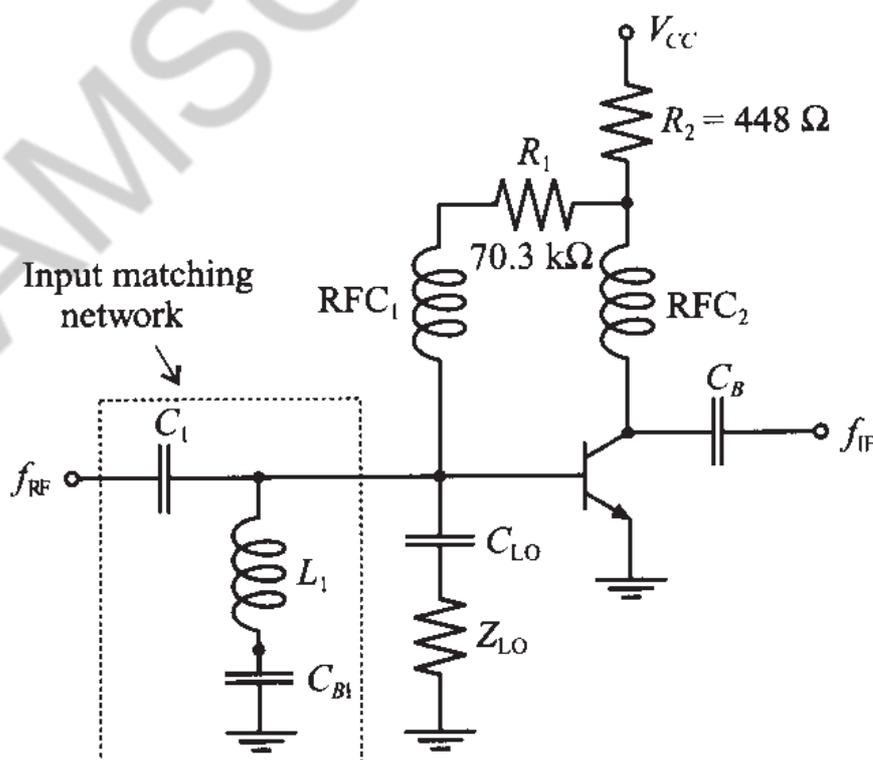
Figure 10-38   Input matching network for a single-ended BJT mixer.

There are several modifications that can be made to the circuit in Figure 10-38. First we notice that instead of biasing the base of the transistor through an RFC, we can connect R_1 directly to the contact between L_1 and C_{B1}. In this case we still bias the base of the transistor through L_1 and maintain isolation of the RF signal from the DC supply by grounding the RF through C_{B1}. One more task of this matching network is to provide a short-circuit condition for the IF signal. Even though the impedance of the inductor L_1 is rather small at f_{IF}, we still can lower it by choosing the value of C_{B1} such that L_1 and C_{B1} exhibit a series resonance at IF. For example, if we choose $C_{B1} = 120$ pF, we still maintain a solid short circuit for the RF signal and we improve the path to ground for the f_{IF} signal. The modified input matching network is shown in Figure 10-39.
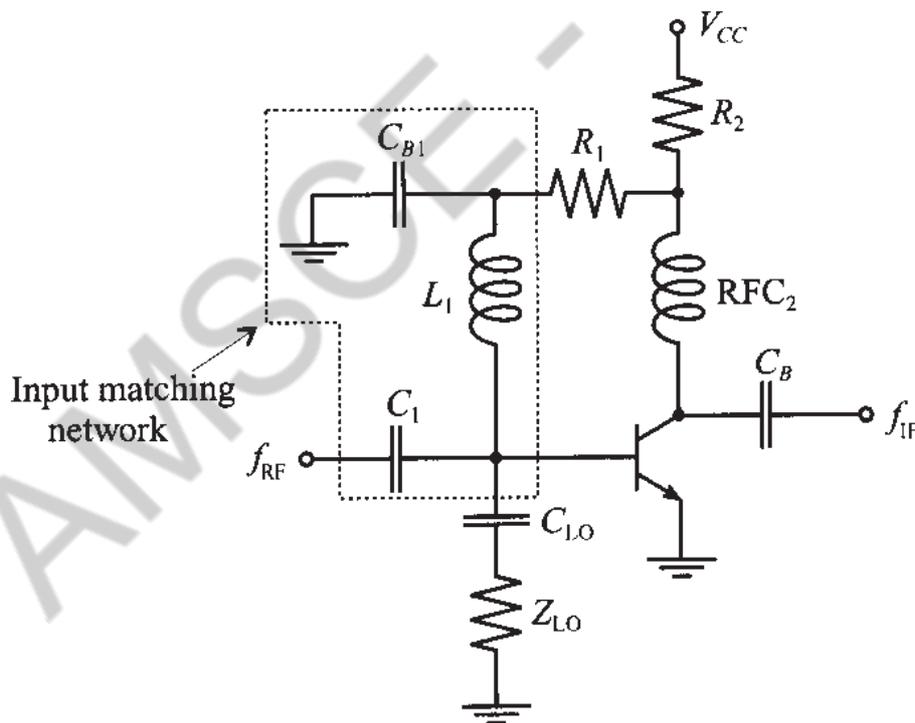
Figure 10-39 Modified input matching network.

The output matching network is developed using a similar approach. The original matching network again consisted of a shunt inductance L_2 followed by a series capacitance C_2. The values are $L_2 = 416$ nH and $C_2 = 1.21$ pF. This topology allows us to eliminate the RFC at the collector terminal of the transistor. However, the problem with this topology is that it does not provide a short circuit

to ground for the RF signal that may interfere with the output. To remedy this drawback we replace L_2 with an equivalent LC combination where the additional capacitance $C_3 = 120$ pF is chosen to provide solid ground condition for the f_{RF} signal and L_2 is adjusted to $L_2 = 5.2$ nH. The complete circuit of the designed single-ended BJT mixer is shown in Figure 10-40.
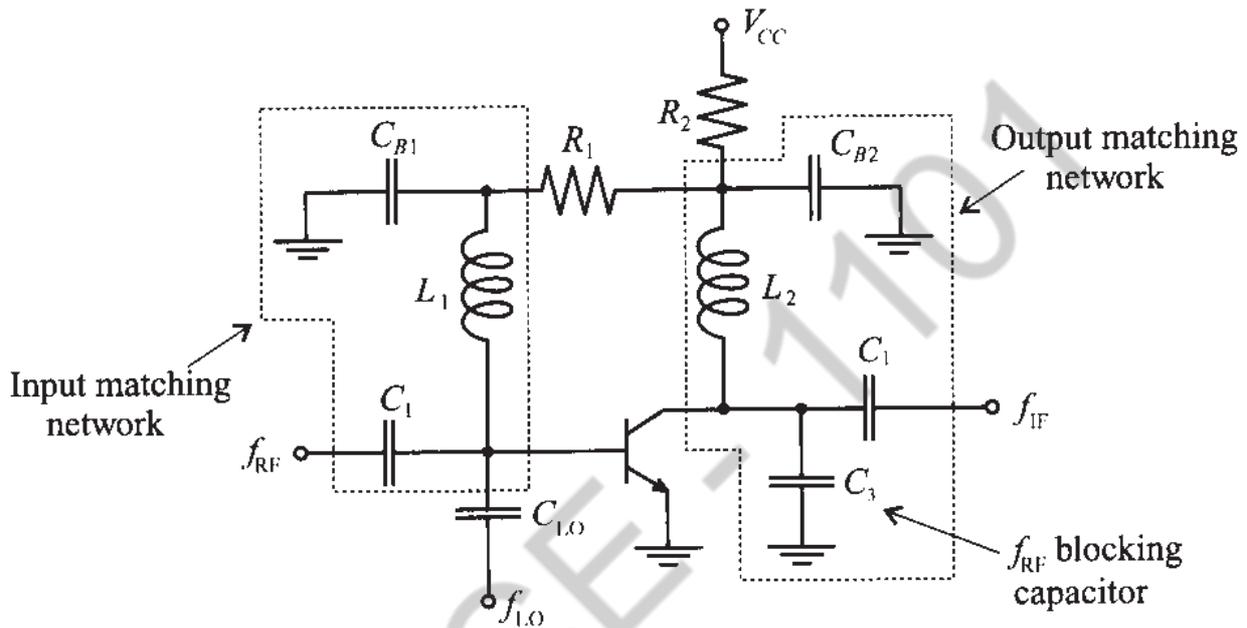
Figure 10-40   Complete electrical circuit of the low-side injection, single-ended BJT mixer with $f_{RF} = 1900$ MHz and $f_{IF} = 200$ MHz.

This design shows the multiple purposes that a matching network can perform. At first glance they are often difficult to understand. Specifically, the dual network purposes of matching and isolation provide challenges for the circuit designer.

10.3.4 Single-Balanced Mixer

From the previous section it is seen that the single-ended mixers are rather easy to construct circuits. The main disadvantage of these designs is the difficulty associated with providing LO energy while maintaining separation between LO, RF, and IF signals for broadband applications. The balanced dual-diode or dual-transistor mixer in conjunction with a hybrid coupler offers the ability to conduct such broadband operations. Moreover, it provides further advantages related to noise suppression and spurious mode rejection. Spurs arise in oscillators and amplifiers due to parasitic resonances

and nonlinearities and are only partially suppressed by the front end. Thermal noise can critically raise the noise floor in the receiver. Figure 10-41 shows the basic mixer design featuring a quadrature coupler and a dual-diode detector followed by a capacitor acting as summation point.
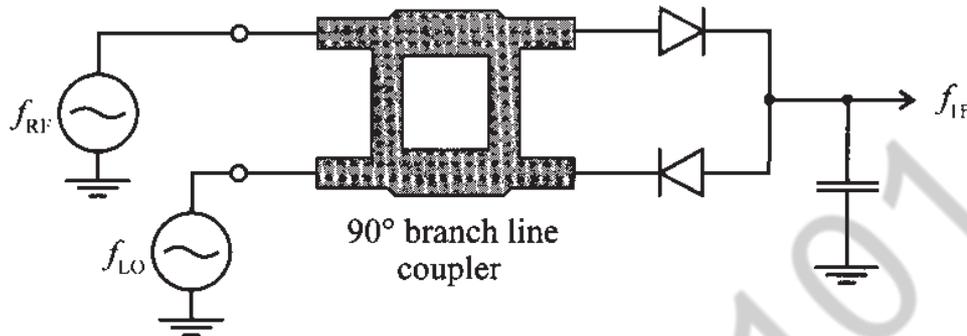
Figure 10-41   Balanced mixer involving a hybrid coupler.

Besides an excellent VSWR (see Appendix G), it can be shown that this design is capable of suppressing a considerable amount of noise because the opposite diode arrangement in conjunction with the 90° phase shift provides a good degree of noise cancellation. The proof is left as an exercise, see Problem 10.22.

A more sophisticated design, involving two MESFETs and 90° and 180° hybrid couplers is shown in Figure 10-42. The 180° phase shift is needed since the second MESFET cannot easily be reversed as done in the anti-parallel diode configuration seen in Figure 10-41. It is also important to point out that this circuit exhibits LO to RF as well as LO to IF signal isolation, but no RF to IF signal isolation. For this reason, a low-pass filter is typically incorporated into the output matching networks of each of the transistors in Figure 10-42.
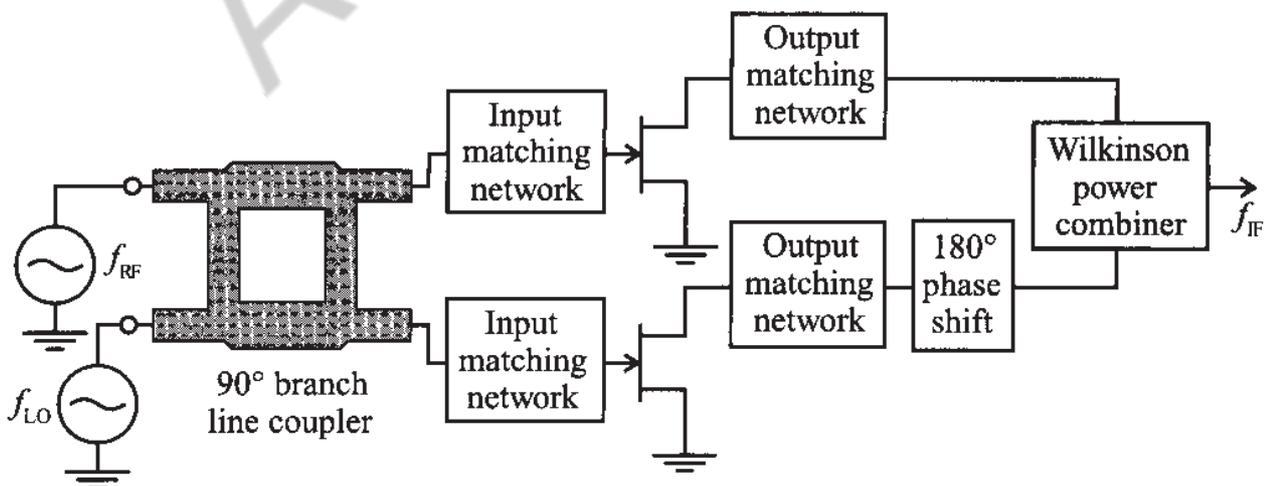
Figure 10-42   Single-balanced MESFET mixer with coupler and power combiner.

10.3.5 Double-Balanced Mixer

The double-balanced mixer can be constructed by using four diodes arranged in a rectifier configuration. The additional diodes provided better isolation and an improved suppression of spurious modes. Unlike the single-balanced approach, the double-balanced design eliminates all even harmonics of both the LO and RF signals. However, the disadvantages are a considerably higher LO drive power and increased conversion loss. Figure 10-43 depicts a typical circuit of the double-balanced design. All three signal paths are decoupled, and the input and output transformers enable a symmetric mixing with the LO signal.
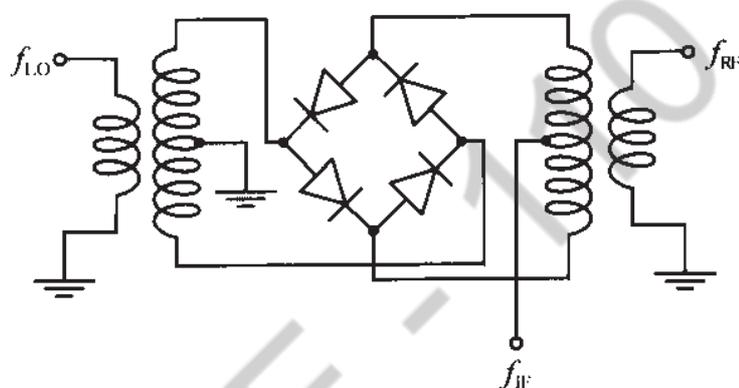
Figure 10-43   Double-balanced mixer design.

For design details of double-balanced mixers the reader is referred to the books by Vendelin and Mass listed at the end of this chapter.

10.4 Summary

Oscillators and mixers require a nonlinear transfer characteristic and are therefore more difficult to design than standard linear amplifiers. It is not uncommon to encounter circuits that perform as desired, but the design engineer does not understand exactly why they behave this way. Today's extensive reliance on CAD tools has often reduced our thinking to trial-and-error approaches. This certainly applies both to oscillators and mixer RF circuits.

One of the key design requirements of an oscillator is the negative resistance condition as a result of the feedback loop equation, which can be formulated as the Barkhausen criterion:

$$H_F(\omega)H_A(\omega) = 1$$

For instance, the feedback Pi-type network results in a host of different oscillator types, of which we discussed the Hartley, Colpitts, and Clapp designs. At frequencies up to approximately 250 MHz one of the passive feedback elements can be replaced by a