4.14 IMPEDANCE TRANSFORMATION EQUATION

One of the most common tasks in microwave engineering is the determination of how a load impedance Z_L is transformed to a new input impedance Z_{IN} by a length of uniform transmission line of characteristic impedance Z_0 and electrical length θ (Fig. 4.14-1).

To simplify this derivation, we assume that the line length is lossless. With the choice of $x = 0$ at the load, the input to the line is at $x = -l$, and the input impedance there is

$$Z_{\text{IN}} = Z(x = -l) = \frac{V(x = -l)}{I(x = -l)} = Z_0 \left[\frac{e^{j\beta l} + \Gamma_L e^{-j\beta l}}{e^{j\beta l} - \Gamma_L e^{-j\beta l}} \right] \tag{4.14-1}$$

Now substitute $\Gamma_L = (Z_L - Z_0)/(Z_L + Z_0)$, $\beta l = \theta$, the identities $e^{j\beta l} = \cos \beta l + j \sin \beta l$ and $e^{-j\beta l} = \cos \beta l - j \sin \beta l$ into (4.14-1), and remove canceling terms to get

$$\begin{aligned} Z_{\text{IN}} &= Z_0 \left[\frac{2Z_L \cos \theta + j2Z_0 \sin \theta}{2Z_0 \cos \theta + j2Z_L \sin \theta} \right] \\ Z_{\text{IN}} &= Z_0 \left[\frac{Z_L + jZ_0 \tan \theta}{Z_0 + jZ_L \tan \theta} \right] \end{aligned} \tag{4.14-2}$$

Similar reasoning can be used to evaluate the input impedance when the transmission line has finite losses. The result is

$$Z_{\text{IN}} = Z_0 \left[\frac{Z_L + Z_0 \tanh \gamma l}{Z_0 + Z_L \tanh \gamma l} \right] \tag{4.14-3}$$

This is one of the most important equations in microwave engineering and is called *the impedance transformation equation*. Remarkably, (4.14-2) and (4.14-3) have exactly the same format when derived in terms of admittance. For the lossless line

$$Y_{\text{IN}} = Y_0 \left[\frac{Y_L + jY_0 \tan \theta}{Y_0 + jY_L \tan \theta} \right] \tag{4.14-4}$$

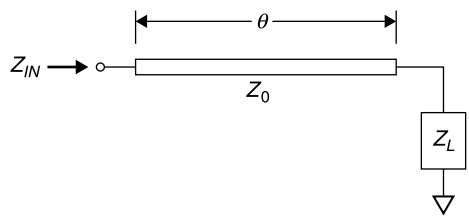


Figure 4.14-1 Equivalent circuit of lossless uniform transmission line of electrical length θ and characteristic impedance Z_0 terminated in impedance Z_L.

and for the line with loss

$$Y_{\text{IN}} = Y_0 \left[\frac{Y_L + Y_0 \tanh \gamma l}{Y_0 + Y_L \tanh \gamma l} \right] \tag{4.14-5}$$

where $Y_0 = 1/Z_0$, $Y_L = 1/Z_L$, and $Y_{\text{IN}} = 1/Z_{\text{IN}}$. Expressions (4.14-4) and (4.14-5) can be verified by substituting these equivalences, respectively, into (4.14-2) and (4.14-3).

In general, even the lossless expressions of (4.14-2) and (4.14-4) are complex to apply, as will be seen by a subsequent example. Beside the fact that when $Z_L = Z_0$ the line is always matched and $Z_{\text{IN}} = Z_0$, there are four additional cases that are both easy to apply and worthy of note.

Case 1 *Lossless lines of zero length or multiples of a half wavelength present the load impedance unchanged at their input terminals:*

$$Z_{\text{IN}} = Z_L \tag{4.14-6}$$

If the transmission line's length is zero or an integer number of half wavelengths $(0°, 180°, 360°, 540° \ldots)$, then $\tan \theta = 0$, and (4.14-2) reduces to $Z_{\text{IN}} = Z_L$. That is, whatever impedance terminates, the line will be presented unchanged at the input.

Case 2 *A short-circuit-terminated lossless transmission line is inductive for* θ *up to* $90°$*, becomes an open circuit when* $\theta = 90°$*, and is capacitive for* $90° < \theta < 180°$*. Thereafter the behavior is repeated every half wavelength. For the short-circuit termination (4.14-2) reduces to*

$$Z_{\text{IN}} = jZ_0 \tan \theta \tag{4.14-7a}$$

A plot of the input reactance $(Z_0 \tan \theta)$ for $Z_0 = 1$ is shown in Figure 4.14-2. The line length is chosen to be $90°$ at 1000 MHz. Notice that when the electrical length of the line is $45°$ (at 500 MHz) the reactance is inductive and numerically equal to Z_0, while at $135°$ (1500 MHz) it is capacitive and the reactance is numerically equal to $-Z_0$.

The fact that a shorted quarter wavelength line $(\theta = 90°)$ becomes an open circuit (at 1000 MHz in Fig. 4.14-2) is useful in rotary joints for antennas, bias injection and filter networks, and a variety of other applications.

Notice also that if the shorted line section is short compared to a wavelength, $\tan \theta \approx \theta$ (in radians) and (4.14-2) can be approximated as

$$Z_{\text{IN}} \approx jZ_0\theta = j\omega \left(\frac{Z_0 l}{v} \right) \tag{4.14-7b}$$

since $\theta = \beta l = (2\pi/\lambda)l = 2\pi(f/v)l = \omega l/v$. For most transmission lines θ is

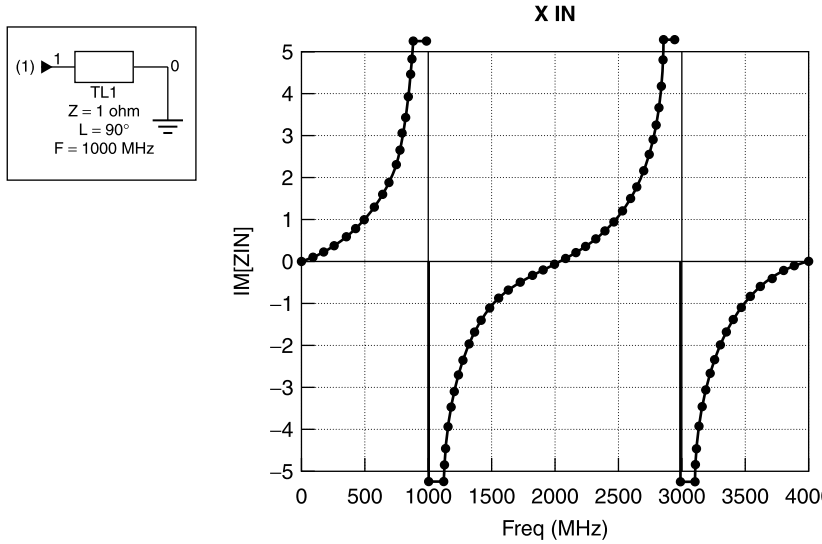


Figure 4.14-2 Reactance of lossless short-circuited transmission line versus frequency ($\theta = 90°$ at 1 GHz).

linearly proportional to frequency, as is the impedance of an inductor, $Z = j\omega L$. Therefore *a short length of short-circuited terminated transmission line behaves as an inductor of value*

$$L_{\text{EFF}} \approx \frac{\theta Z_0}{\omega} = \frac{Z_0 l}{v} \quad (\theta \leq 0.53 \text{ rad}) \tag{4.14-7c}$$

For example, if a 1-cm length of $Z_0 = 120\ \Omega$, air coax line is terminated in a short:

$$L_{\text{EFF}} \approx \frac{Z_0 l}{v} = \frac{120(1 \text{ cm})}{3 \times 10^{10} \text{ cm/s}} = 4 \text{ nH}$$

Since 1 cm $\approx$ 0.4 in., the inductance of a high impedance air-dielectric line is approximately equal to 10 nH/in., the approximation presented earlier in (2.12-6).

The approximation (4.14-7c) also applies to a short length of Z_0 transmission line terminated in a load Z_L provided that its characteristic impedance $Z_0 \gg |Z_L|$, as can be verified by applying this condition and $\tan\theta \approx \theta$ in (4.14-2).

The approximation $\tan\theta \approx \theta$ occurs frequently and has an error magnitude of less than 10% for values of θ up to 0.524 rad (30°). A radian is $360°/2\pi \approx 57.3°$. Thus 0.5 rad is nearly $\frac{1}{12}$ of a wavelength. The values of θ and

TABLE 4.14-1 Approximating tan θ by Its Argument θ (in radians)

θ (deg)	0	10	20	25	30
θ (rad)	0	0.175	0.349	0.436	0.524
tan θ	0	0.176	0.364	0.466	0.577
Error in tan θ ≈ θ	0	−0.6%	−4.1%	−6.4%	−9.1%

tan θ are compared in Table 4.14-1 along with the size of the error in using this approximation, showing that the error magnitude is less than 10% at $\theta = 30°$ and diminishes rapidly for lesser values of θ.

Case 3 *An open-circuit terminated lossless transmission line is capacitive for θ up to 90°, becomes a short circuit when $\theta = 90°$, and is inductive for $90° < \theta < 180°$. Thereafter the behavior is repeated every half wavelength. Thus, from (4.14-4) with $Y_L = 0$*

$$Y_{IN} = jY_0 \tan \theta \quad \text{or} \quad Z_{IN} = -jZ_0 \cot \theta \tag{4.14-8a}$$

This is plotted in Figure 4.14-3. Notice that the input reactance magnitude is capacitive and equal to $-Z_0$ when the line is 45° long (500 MHz). For $\theta = 90°$ (1000 MHz in Fig. 4.14-3) the open-circuit load is transformed to a short circuit at the input. It is inductive and equal to Z_0 when 135° long (1500 MHz). The input repeats every half wavelength after 180°. The facility for transforming an open circuit to a short circuit is equally useful in a variety of microwave circuits.

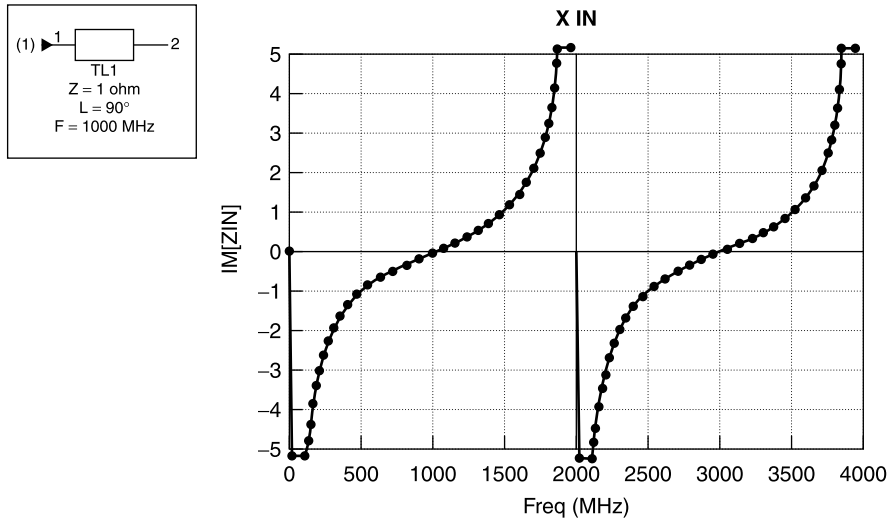


Figure 4.14-3 Normalized reactance of lossless open-circuited transmission line versus frequency (electrical length = 90° at 1 GHz).

By using similar reasoning to that applied to a short section of a shorted transmission line, and again applying the approximation $\tan\theta \approx \theta$ for small values of θ, a short section of open-circuited transmission line has a capacitive admittance given by

$$Y_{\text{IN}} \approx jY_0\theta \tag{4.14-8b}$$

Since $B_C = \omega C$ and $\theta = \omega l/v$, the equivalent capacitance is

$$C_{\text{EFF}} \approx \frac{\theta}{\omega Z_0} = \frac{l}{Z_0 v} \quad (\theta \leq 0.53 \text{ rad}) \tag{4.14-8c}$$

and a reactance of

$$\frac{1}{\omega C} \approx \frac{Z_0}{\theta} \tag{4.14-8d}$$

For example, a section of 25-Ω line that is 20° long at 1 GHz has an equivalent capacitance of

$$C_{\text{EFF}} \approx \frac{0.35 \text{ rad}}{6.28 \times 1 \times 10^9 \text{ Hz} \times 25\ \Omega} = 2.2 \text{ pF}$$

Its reactance X_C at 1 GHz is

$$\frac{1}{\omega C_{\text{EFF}}} \approx \frac{25\ \Omega}{0.35} = 71\ \Omega$$

If a short length of Y_0 line is connected to a load of admittance Y_L, it behaves as a shunt capacitance approximated by (4.14-8c) provided that $|Y_L| \ll Y_0$ (i.e., $|Z_L| \gg Z_0$) and $\tan\theta \approx \theta$.

Case 4 *The normalized input impedance to a quarter wavelength long transmission line is the reciprocal of the normalized load impedance.*

When the transmission line is a quarter wavelength long, $\theta = 90°$, and the input impedance to the line is

$$Z_{\text{IN}} = Z_0 \frac{Z_0}{Z_L} \tag{4.14-9}$$

Normalizing to Z_0 (dividing all impedances by Z_0),

$$z_{\text{IN}} = \frac{1}{z_L} \tag{4.14-10}$$

For this reason a quarter wavelength line is often called an *impedance inverter*. Notice that the normalized input impedance is equal to the normalized load admittance, that is,

$$z_{\text{IN}} = \frac{1}{z_L} = y_L \tag{4.14-11}$$

where the normalized admittance $y_L = Y_L/Y_0$ and $Y_0 = 1/Z_0$.

Later we will see that this property allows us to convert between admittance and impedance using the Smith chart. Notice also that this impedance inverting property occurs not only for quarter wavelength length lines but for line lengths that are any odd multiple of a quarter wavelength $(\lambda/4, 3\lambda/4, 5\lambda/4, \ldots)$.

In general, to transform any resistive load R_L to a desired input resistance R_{IN}, one need only select a quarter wave line section having a Z_0 that satisfies

$$Z_0 = \sqrt{R_{\text{IN}} R_L} \tag{4.14-12}$$

As an example of the resistive transformation of a 90° line length, consider a 50-Ω, quarter wavelength, transmission line terminated in 25 Ω. From (4.14-9) the input resistance is

$$Z_{\text{IN}} = Z_0 \frac{Z_0}{Z_L} = 50 \frac{50}{25} = 100\ \Omega \tag{4.14-13}$$

Thus, the 25-Ω resistive load has been "transformed" to an input resistance of 100 Ω at 1 GHz, and the variation of the input impedance with frequency for this case is shown in Figure 4.14-4.

It is a common but ill-advised practice to refer to the quarter wavelength transmission line as a *quarter wave transformer* because the inversion property can be used to transform a resistive load of one value to an input resistance of another value *at a single frequency*. However, this practice should be avoided. It is better to visualize it as a *quarter wave impedance inverter*, recognizing that even that property is limited to a single frequency of operation.

The inversion changes the real parts of the load, but in addition the sign of the imaginary part is also reversed, turning an impedance with an inductive part into one with a capacitive part or vice versa. For example, if in the prior case, rather than a purely resistive load of 25 Ω, we had instead a load of $(25 + j25)$ Ω, the input impedance becomes

$$\begin{aligned} Z_{\text{IN}} &= Z_0 \frac{Z_0}{Z_L} = 50 \frac{50}{25 + j25} = 50 \frac{50}{25 + j25} \frac{25 - j25}{25 - j25} \\ &= \frac{(2500)(25 - j25)}{625 + 625} = (50 - j50)\ \Omega \end{aligned} \tag{4.14-14}$$

Notice that the transformation ratio of the real part is now 2 instead of the 4 obtained with a purely resistive load. Equally noteworthy, notice that the imaginary part of the input impedance is capacitive whereas the load's imagi-

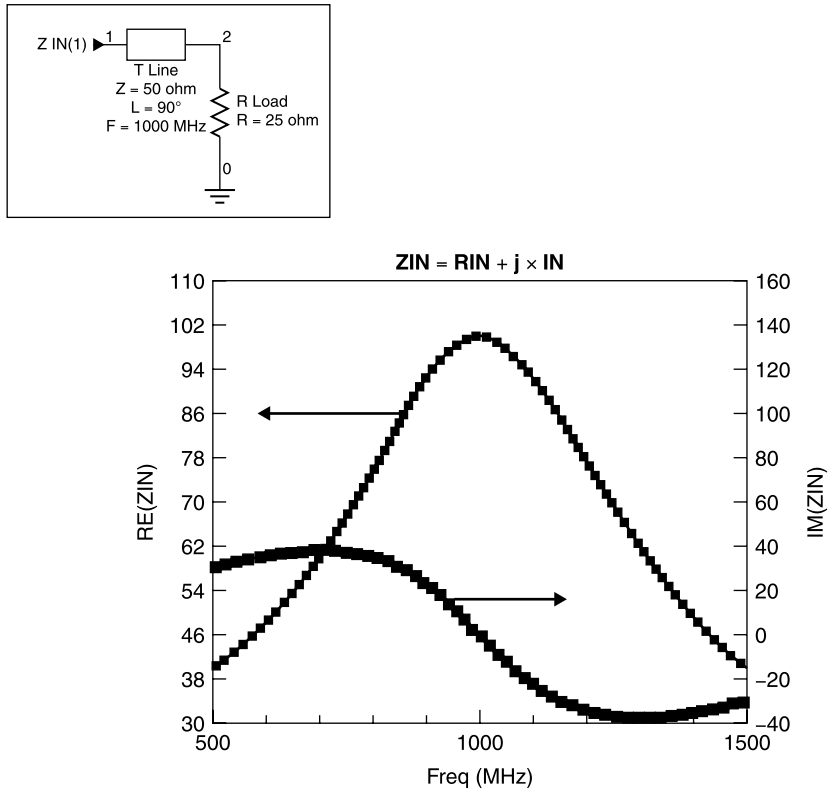


Figure 4.14-4 Variation of input impedance with frequency when 50-Ω "impedance inverter" is used to match load of 25 Ω to input resistance of 100 Ω at 1 GHz.

nary part was inductive! Thus *the action of the quarter wavelength transmission line is quite different from that of an ideal impedance transformer*, which simply multiplies any impedance by a constant factor.

Also, an *ideal transformer*, which is well approximated at low frequencies by a pair of magnetically coupled wire windings, *has a fixed impedance transformation ratio that is frequency independent*. On the other hand, the impedance transforming property of a quarter wave line section only approximates a constant transformation ratio for resistive loads and only over a narrow bandwidth centered at the design frequency as shown in Figure 4.14-4. Despite these limitations, the quarter wavelength line section has considerable practical utility for its transforming properties.

The preceding four special transmission line cases demonstrate the insight to be gained by application of the input impedance formula of (4.14-2). This formula is key to transmission line analyses. However, for complex load values and arbitrary line lengths, the formula can be quite tedious to apply. As an example, suppose that we wish to find the input impedance to a 50-Ω line that is 50° long and terminated in a load impedance of $(25 + j30)$ Ω (Fig. 4.14-5).

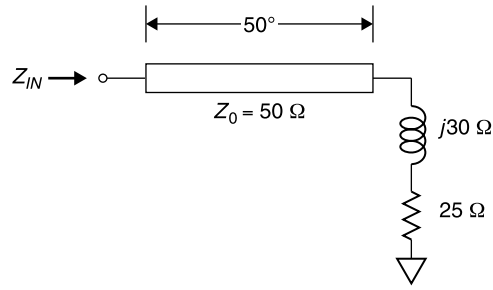


Figure 4.14-5 Example of determination of input impedance to length of 50-Ω transmission line terminated in a complex impedance of $(25 + j30)\ \Omega$.

Applying (4.14-2) gives

$$\begin{aligned} Z_{\text{IN}} &= Z_0\left[\frac{Z_L + jZ_0\tan\theta}{Z_0 + jZ_L\tan\theta}\right] = 50\frac{25 + j30 + j50\tan 50^\circ}{50 + j(25 + j30)\tan 50^\circ} \\ &= 50\frac{25 + j(30 + 59.6)}{50 - 35.8 + j29.8} = 50\frac{25 + j89.6}{-14.2 + j29.8} \\ &= 50\frac{\sqrt{625 + 8027}\angle 74.4^\circ}{\sqrt{202 + 888}\angle 64.5^\circ} = 50\frac{93.0\angle 74.4^\circ}{33.0\angle 64.5^\circ} \\ &= 141\angle 10^\circ = (139 + j24)\ \Omega \end{aligned} \tag{4.14-15}$$

Notice that the 25-Ω real part of the load has been transformed to 139 Ω and that the $+j30$-Ω imaginary part of the load has been transformed to $j24$-Ω. From this it can be seen that, *in general, the impedance transformation properties of a simple transmission line can be quite complex*.

It is tedious to perform this calculation and very easy to make a mathematical error while doing so, thereby obtaining an erroneous result. Also, the calculation must be re-performed for each frequency of evaluation.

It would be a considerable challenge to predict, using only (4.14-2), how Z_{IN} varies with θ as it is varied over, say, 0° to 180°. This evaluation would require charting Z_{IN} for numerous θ values using the laborious calculation method described in (4.14-15). It was for this reason that Phillip Smith found an elegant graphical solution to determine input impedance, the subject of the next chapter.

4.15 IMPEDANCE MATCHING WITH ONE TRANSMISSION LINE

In the previous section it was shown that a length of transmission line can transform a real load resistance to one of a different value. Accordingly, one might ask whether any complex load impedance $Z_L = R_L + jX_L$ can be con-

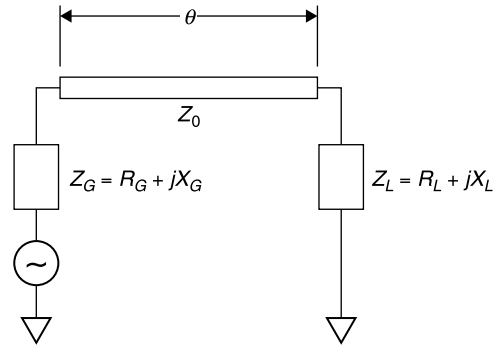


Figure 4.15-1 Conjugately matching complex load impedance Z_L to a generator impedance Z_G, using only transmission line of characteristic impedance Z_0 and electrical length θ.

jugately matched to any other complex source impedance, $Z_G = R_G + jX_G$, merely by proper selection of a line length of the appropriate Z_0 and electrical length θ. The following derivation (refer to Figure 4.15-1) yields the conditions under which this can be achieved.

From (4.14-2) with Z_{IN} set equal to the complex conjugate of Z_G,

$$Z_{IN} = R_G - jX_G = Z_0 \frac{Z_L + jZ_0 \tan \theta}{Z_0 + jZ_L \tan \theta} \tag{4.15-1}$$

Next, set the real and imaginary parts of this equation separately equal to each other and solve for Z_0 and θ. Note that $-jX_G$ was used in the equation in order that Z_{IN} will be the complex conjugate of Z_G:

$$Z_0 = \sqrt{\frac{R_L(R_G^2 + X_G^2) - R_G(R_L^2 + X_L^2)}{R_G - R_L}} \tag{4.15-2}$$

$$\theta = \tan^{-1}\left[\frac{Z_0(R_L - R_G)}{R_L X_G - R_G X_L}\right] \tag{4.15-3}$$

The above equations are always mathematically valid, however, a practical solution exists only when the solution for Z_0 is a real and finite number.

4.16 FANO'S (AND BODE'S) LIMIT

Type A Mismatched Loads

Over a given bandwidth there is a definite limit to how well one can tune a mismatched termination. The most well known are called *Fano's limits*. Fano derived these interesting and useful results within a thesis project [5].

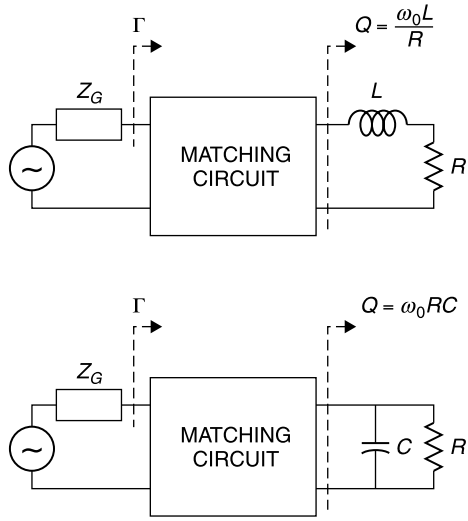


Figure 4.16-1 Fano's type A mismatched loads [5, 6].

One of Fano's two integral equation limits applies to what he described as type A mismatched loads (Fig. 4.16-1). It is

$$\int_0^\infty \ln\left|\frac{1}{\Gamma}\right| d\omega \leq \frac{\pi\omega_0}{Q} \tag{4.16-1}$$

where the reflection coefficient Γ is referenced to Z_G and the real part of Z_G is R, the same as the load resistance R in Figure 4.16-1.

Basically, (4.16-1) states that no matter what matching circuit is used, the area under the $\ln(1/\Gamma)$ curve cannot exceed the value $\pi\omega_0/Q$. Given this limit, we would usually prefer that Γ_{MIN} would be obtained over our specified bandwidth, $\omega_1 - \omega_2$, and then that Γ be unity for all other frequencies. This ideal matching requirement is shown in Figure 4.16-2.

In practice, of course, the matching behavior described in Figure 4.16-2 can only be approximated using real tuning circuits, but the ideal limit serves as an insightful restriction on how well a circuit can be tuned over a given bandwidth.

If we assume that the reflection coefficient is a constant equal to its minimum value in the passband, then the integral for the type A mismatched loads reduces to a constant multiplied by the frequency interval:

$$\int_0^\infty \ln\left|\frac{1}{\Gamma}\right| d\omega = (\omega_2 - \omega_1) \ln\left|\frac{1}{\Gamma}\right| \tag{4.16-2}$$

where ω_1 is the frequency (in radians/second) at the lower band edge and ω_2 is

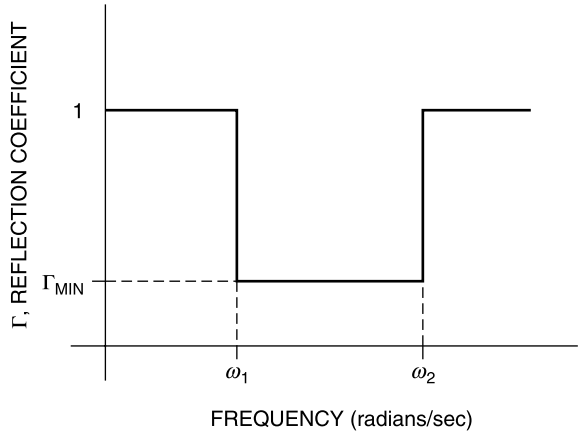


Figure 4.16-2 Idealized Γ_{MIN} match in the passband and total reflection ($\Gamma_{MIN} = 1$) at all other frequencies [5, 6].

the upper band edge. We also define

$$\text{Center frequency} = \omega_0 = \sqrt{\omega_1 \omega_2} \tag{4.16-3}$$

and

$$\text{Fractional bandwidth} = \Delta\omega = \frac{\omega_2 - \omega_1}{\omega_0} \tag{4.16-4}$$

Then for the type A mismatched load, assuming that a minimum and constant value of reflection coefficient, Γ_{MIN}, could be obtained throughout the passband $\omega_1 - \omega_2$

$$(\omega_2 - \omega_1) \ln \left| \frac{1}{\Gamma_{MIN}} \right| \leq \frac{\pi \omega_0}{Q_L} \tag{4.16-5}$$

Equivalently,

$$|\Gamma_{MIN}| \geq e^{-\pi\omega_0/Q(\omega_2 - \omega_1)} = e^{-(\pi/Q)(f_0/\Delta f)} \tag{4.16-6}$$

where $f_0 = \omega_0/2\pi$ and $\Delta f = (\omega_2 - \omega_1)/2\pi$.

For example, assume there is a type A circuit consisting of a 50-Ω load resistor in parallel with a 10-pF capacitor to be matched to a 50-Ω generator over the 700- to 1100-MHz band. Notice that 10 pF presents a large susceptance in parallel with the 50-Ω load at the geometric center of this frequency band, 877.5 MHz. In reactance terms, at 877.5 MHz this is a capacitive reactance of only 18.1 Ω!

The minimum reflection coefficient magnitude, $|\Gamma_{MIN}|$, that we can expect to average over the band is found as follows.

First, calculate Q, f_0, and Δf:

$$f_0 = \sqrt{(700)(1100)}\ \text{MHz} = 877.5\ \text{MHz}$$

$$X_C = \frac{159}{(10\ \text{pF})(0.8775\ \text{GHz})} = 18.1\ \Omega$$

$$Q = \frac{R}{X_C} = \frac{50\ \Omega}{18.1\ \Omega} = 2.76$$

$$\Delta f = 700 - 1100 = 400\ \text{MHz}$$

Then

$$|\Gamma_{MIN}| \geq e^{(-\pi/Q)(f_0/\Delta f)} = e^{(-3.14/2.76)(877.5/400)} = e^{-2.5} = 0.082 \quad (4.16\text{-}7)$$

and the corresponding minimum average VSWR is

$$\text{VSWR}_{MIN} = \frac{1 + |\Gamma_{MIN}|}{1 - |\Gamma_{MIN}|} = \frac{1 + 0.082}{1 - 0.082} = 1.18 \quad (4.16\text{-}8)$$

To examine this case, a 10-element ladder network was used as a matching network [10, Sec. 5.3]. The element values were selected arbitrarily, 10 nH for inductors and 10 pF for capacitors, and the circuit optimized using a network simulator. The simulation goal was set equal to a 1.2 maximum VSWR over the 700 to 1100 MHz bandwidth. The result, obtained once the optimization appeared to produce little further improvement, is shown in Figure 4.16-3. An average VSWR of about 1.3 was obtained over the 700 to 1100 MHz bandwidth, comparing moderately well with the theoretical limit of 1.18.

Type B Mismatched Loads

Similar reasoning can be used with Fano's type B mismatched loads (Fig. 4.16-4), for which Fano's integral limit is

$$\int_0^\infty \frac{1}{\omega^2} \ln\left|\frac{1}{\Gamma}\right| d\omega \leq \frac{Q}{\pi\omega_0} \quad (4.16\text{-}9)$$

The limits for both type A and type B mismatched loads are commonly attributed to Fano [5]; however, Fano, himself, credits the limit for the type A circuits to Bode [7]. Whatever their origin, the limits give a useful insight into the problem of broadband matching.

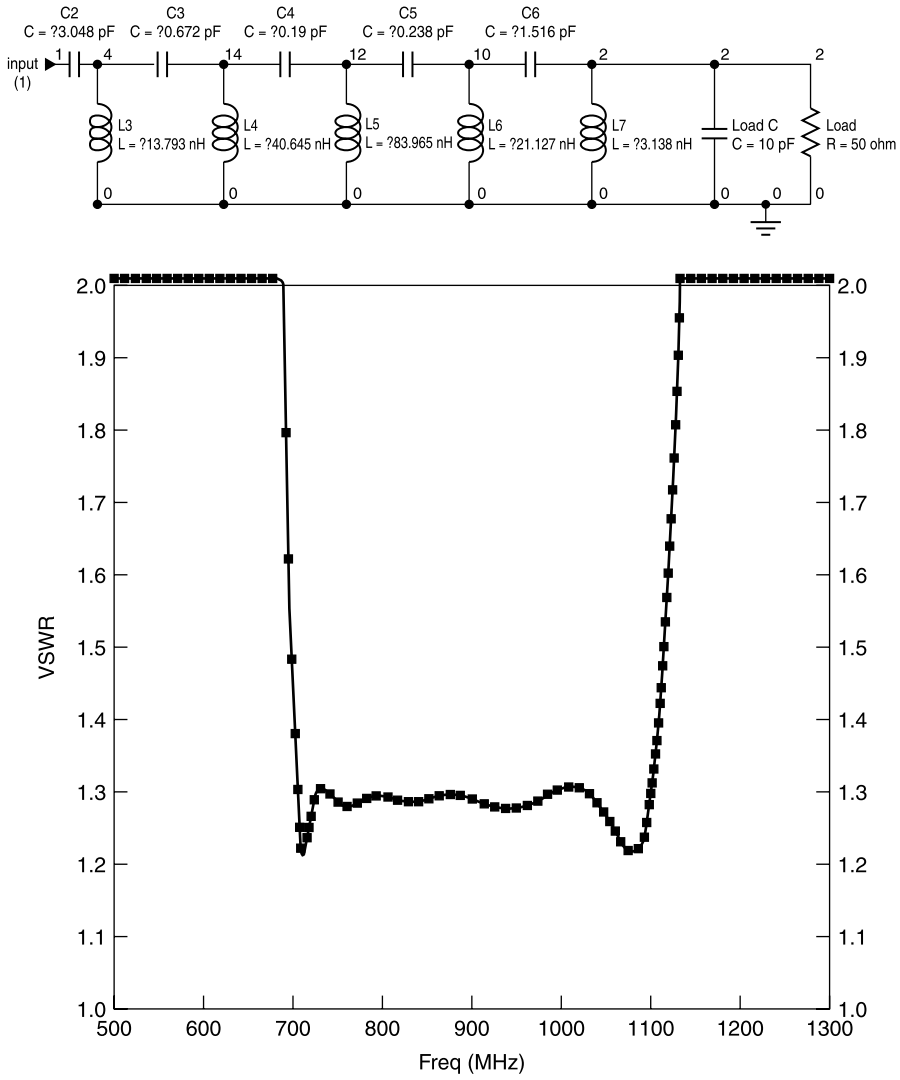


Figure 4.16-3 A 10-element matching network and its VSWR.

Impedance Transformation Not Included

When Fano and Bode performed their research, presumably for the telephone industry, the frequencies of their concern were relatively low, a few megahertz. For this reason they probably considered that a simple change in resistive level could always be accommodated using an "ideal" transformer, which has no bandwidth limitations. Probably for this reason, they only considered the reactive part of the load as a mismatch with inherent bandwidth limitations. At

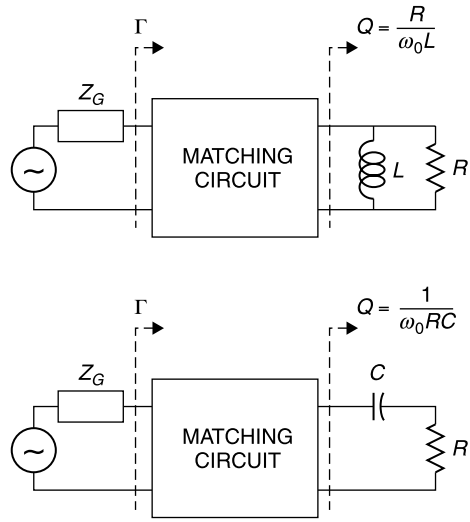


Figure 4.16-4 Fano's type B mismatched loads [5, 6].

microwave frequencies the construction of a transformer that approximates frequency-independent behavior is not as practical. Thus, the usefulness of Fano's limits at microwave frequencies are mainly confined to those cases for which the load has the correct resistive value ($R = Z_G$) but has some series or shunt reactance that must be tuned using a matching network.

REFERENCES

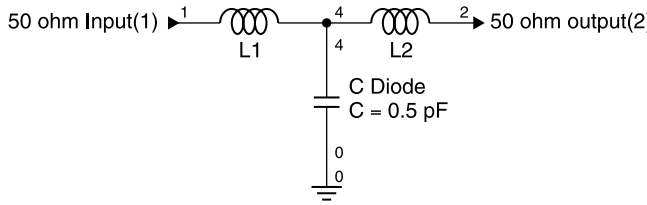
1. Joseph F. White, *Microwave Semiconductor Engineering*, Noble Publishing, Norcross GA, 1995. *Extensive treatment of PIN diodes and their switching and phase shifting applications. Also includes fundamentals of microwave circuits.*
2. L. B. W. Jolley, *Summation of Series*, 2nd rev. ed., Dover, New York, 1961.
3. Peter A. Rizzi, *Microwave Engineering, Passive Circuits*, Prentice-Hall, Englewood Cliffs, NJ, 1988. *Excellent microwave engineering textbook covering theory and design of transmission lines, couplers, filters, and numerous other passive devices.*
4. Simon Ramo and John R. Whinnery, *Fields and Waves in Modern Radio*, Wiley, New York, 1944, and 1953. *This is a classic introductory text describing electromagnetic fields and Maxwell's equations.*
5. R. M. Fano, "Theoretical limitations on the broad-band matching of arbitrary impedances," *Journal of the Franklin Institute, Vol. 249, 1950*, p. 57. *Also see his doctoral thesis from the Massachusetts Institute of Technology (1947) and Technical Report No. 41, Massachusetts Institute of Technology, Research Laboratory of Electronics.*
6. Theodore Grosch, *Small Signal Microwave Amplifier Design*, Noble Publishing, Norcross, GA, 1999. *A thorough derivation of all of the important formulas for tran-*

sistor amplifier design and evaluation, based on the S parameters of the transistor and the circuit that surrounds it.

7. H. W. Bode, "A method of impedance correction," *Bell System Technical Journal*, Vol. IX, October 1930, Sections 10.7 and 10.8, pp. 794–835.
8. Theodore Saad (Ed.), *The Microwave Engineer's Handbook, Vols. I and II*, Horizon House Microwave (now Artech Books, Canton, Massachusetts) 1971. *An excellent collection of graphs and charts useful in microwave engineering.*
9. Simon Ramo, John R. Whinnery, and Theodore Van Duzer, *Fields and Waves in Communication Electronics*, 3rd ed., Wiley, New York, 1994. *This is the updated version of [4].*
10. Devendrak K. Misra, *Radio Frequency and Microwave Communication Circuits*, Wiley, New York, 2001. *This is a well-written communication text and reference.*

EXERCISES

E4.1-1 A manufacturer of PIN switching diodes proposes to make a broadband switch module consisting of a PIN diode in shunt with a 50-Ω transmission line. When the diode is reverse biased, it appears as a 0.5-pF capacitor in shunt with the transmission line and is to pass signals with minimum reflection.



He suggests to you that if the hookup wires (from the top of the diode to the center conductor of the line and shown as L1 and L2 above) are designed to have inductance values to satisfy

$$Z_0 = \sqrt{\frac{L}{C}} = 50\ \Omega$$

the module will be electrically indistinguishable from a length of 50-Ω line and consequently will provide matched transmission at all frequencies.

a. What should the hookup wire inductances ($L/2$ each) be made to satisfy the above relation?

b. Is it correct that this will provide an infinitely broadband transmission match?

c. Test your answer on a network simulator over the bandwidth 0 to 10 GHz.

E4.1-2 A 50-Ω air dielectric coaxial transmission line has an inner diameter of the outer conductor equal to 0.500 in. and a center conductor diameter of 0.217 in. What is its distributed capacitance per inch?

E4.4-1 A voltage wave with 100-V peak amplitude is incident on a transmission line extending in the z direction. When it reaches the load, one quarter of its power is reflected back toward the generator which is matched to the line.

a. What is the peak voltage on the line between generator and load?

b. What is the minimum voltage on the line between generator and load?

c. Sketch the manner in which the incident and reflected waves add and subtract along the line as z changes for a fixed time t.

E4.5-1 **a.** What is the magnitude of the reflection coefficient in E4.4-1?

b. What is the VSWR?

E4.5-2 A friend has just purchased a boat containing a marine radio. The radio has a built-in VSWR meter that indicates that the antenna has a VSWR of 3 to 1.

a. How much does this reduce the radiated power of the transmitter relative to that which would be radiated with a perfectly matched antenna?

b. What would you advise to tune the radio to the antenna?

E4.8-1 Show that for a susceptance jB connected in shunt with an otherwise matched transmission line of characteristic impedance Z_0, the magnitude of the transmission coefficient T is less than unity for all values of B.

E4.8-2 You wish to send the output of a 500-MHz signal source to another building via a 50-Ω coaxial cable. The cable loss is 10 dB. which must be compensated by adding an amplifier at the output of the source. When connected to the 50-Ω cable, the source sees the cable as a reflection with VSWR = 10 (equivalently, the source output VSWR is 10), and you have a 20-dB amplifier with input VSWR of 20 to 1 referenced to 50 Ω. Neglecting other system interactions:

a. What will be the mismatch error in connecting this amplifier to the source?

b. Will it overcome the 10-dB cable loss in all cases?

E4.11-1 Show that the real part of the propagation constant, α, can be approximated for a low-loss transmission line having R ohms/unit length and no shunt conductance by

$$\alpha \approx \frac{R}{2Z_0}$$

E4.11-2 Use the result you derived in Exercise 4.11-1 to estimate the loss in decibels at 1 GHz for 100 ft of JAN Type RG 224/U coaxial cable having a copper inner conductor of 0.106 in. diameter (d_1) and a copper outer conductor with 0.370 in. diameter (d_2). Assume the resistivity for copper is 1.72×10^{-8} Ω-m (then conductivity, $\sigma = 5.8 \times 10^7$ ℧/m) and neglect the conduction losses through the dielectric. *Hint: Assume an equivalent current flow within one skin depth on the conductor surfaces.*

E4.14-1 Beginning with (14.4-2) verify that

$$Y_{\text{IN}} = Y_0\left[\frac{Y_L + jY_0 \tan\theta}{Y_0 + jY_L \tan\theta}\right]\left(\right.$$

E4.14-2 Show that a *short length* of Z_0 transmission line terminated by a load Z_L can be approximated as a series inductance $L_{\text{EFF}} \approx Z_0\theta/\omega$ when $Z_0 \gg |Z_L|$.

E4.14-3 A ham radio friend has purchased a 52-MHz transmitter with matched impedance antenna, both 50 Ω. He wishes to install the antenna on his roof and the transmitter in his study, a separation distance of about 25 ft. Unfortunately, he only has a 75-Ω cable to make the hookup. The relative dielectric constant of the cable's dielectric is 2.8. Is there any way that he can retain the very good match between transmitter and antenna using this cable?

E4.14-4 A resourceful engineer found a way to connect a 12.5-Ω load to a 50-Ω generator without reflection when provided only with a spool of 50-Ω cable. How did he do it?

E4.14-5 An error was made in the design of an integrated circuit narrow-band amplifier with the result that its input impedance is $(50 - j50)$ Ω instead of 50 Ω. Your company mounts the integrated chip on a "motherboard" printed circuit inside a housing with 50-Ω connectors. Vibration specifications preclude installing a lumped coil at the input to tune the amplifier. You propose to print a section of 150-Ω transmission line on the motherboard in cascade between the chip and input connector, reasoning that it will appear inductive and help to tune the mismatch. Specifications require an input VSWR of no more than 1.2.

a. What is the VSWR before this tuning?

b. How long should you make the 150-Ω line (in degrees or wavelengths)?

c. What is the VSWR after the proposed tuning? Does it meet the objective of a VSWR = 1.2 maximum?

E4.14-6 An engineer has a coil of coaxial cable and needs to determine its dielectric material for a thermal analysis. He has been told that the dielectric material is either Teflon ($\varepsilon_R = 2.03$) or polyethelene ($\varepsilon_R = 2.26$). He forms a line section $D = 1$ m long, and he places a short circuit at its load end. Then he connects the open end to a slotted line and examines the input impedance using a variable frequency generator. He finds that it has its first resonance (a short circuit at the input) at 105.28 MHz. Can you say what the dielectric material is?

E4.15-1 Can you find a 50-Ω perfectly matched input solution to E4.14-6 using a properly chosen Z_0 and electrical length? If so, what are the values?

E4.15-2 A load impedance of $10 + j10\ \Omega$ is to be matched to a 50-Ω generator. Can this be done using only a section of transmission line in cascade with the load? What is its Z_0 and electrical length θ?

E4.16-1 An available power amplifier design is being considered for an application requiring that the input VSWR not exceed 1.5 over the 1700 to 2100 MHz band. The amplifier, however, has an input resistance of 50 Ω that is shunted by 10 pF capacitance and presently does not meet this specification.

a. What is f_0 for the 1700 to 2100 MHz bandwidth?

b. What is the Q of this capacitively loaded input circuit?

c. What is Fano's limit for the least VSWR obtainable over the 1700 to 2100 MHz bandwidth?

d. What value of input shunting inductance will parallel resonate the 10 pF. Use a network simulator to determine over what bandwidth the 1.5 VSWR can be achieved with this single shunt L element. Does this meet the amplifier requirement?

e. Using as many as six tuning elements and the network simulator optimizer, find a matching network that provides as low a VSWR over the 1700 to 2100 MHz bandwidth as you can achieve. Does it meet the maximum 1.5 VSWR over the 1.7 to 2.1 GHz bandwidth?

CHAPTER 5

The Smith Chart

5.1 BASIS OF THE SMITH CHART

Knowledge of the basis and use of the *Smith chart*, a graphical presentation of the reflection coefficient with normalized impedance as a parameter overlay, is a sine qua non for the microwave engineer. Originally created as an aid to determining the input impedance to a transmission line, the Smith chart has become a universal aid to the design of matching circuits and to the display of measured data.

In the 1930s, when Phillip Smith [1–3] invented the graphical solution that facilitated the determination of how impedances were transformed by lengths of transmission lines, digital computers were unavailable to engineers. Engineers employed graphing strategies to gain insight into the variations produced by the independent variables of complex formulas. However, such insight was not forthcoming by attempting directly to graph the input impedance formula (4.14-2), repeated below, for a transmission line.

$$Z_{\text{IN}} = Z_0 \left[\frac{Z_L + jZ_0 \tan \theta}{Z_0 + jZ_L \tan \theta} \right] \tag{5.1-1}$$

The key to obtaining insight into this transformation was to recognize that, although impedance varies in a complex manner as one moves away from the load along a lossless transmission line, *reflection coefficient variation is quite simple*! *In traveling from the load to the generator along a lossless line, only the angle of the reflection coefficient changes, not its magnitude* (Fig. 5.1-1).

The reflection coefficient is the complex ratio of the reflected to the incident voltage waves. The total voltage on the line consists of the incident, V_I, and the reflected, V_R, voltages:

$$\begin{aligned} V(x) &= V_I e^{-j\beta x} + V_R e^{+j\beta x} \\ V(x) &= V_I e^{-j\theta} + V_R e^{+j\theta} \end{aligned} \tag{5.1-2}$$

$$\Gamma(x) = \frac{V_R(x)}{V_I(x)} = \rho \angle \varphi^\circ \tag{5.1-3}$$

High Frequency Techniques: An Introduction to RF and Microwave Engineering, By Joseph F. White.
ISBN 0-471-45591-1

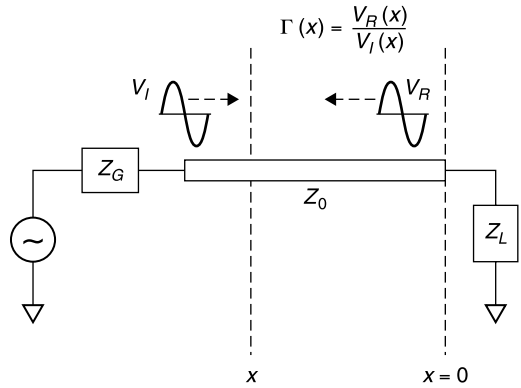


Figure 5.1-1 Moving from the load to generator on lossless transmission line, reflection coefficient changes in phase only because the amplitudes of incident voltage V_I and the reflected voltage V_R do not change.

where $V_R(x) = V_R e^{+j\theta}$, $V_I(x) = V_I e^{-j\theta}$, and $V_R(0)/V_I(0) = \Gamma_{\text{Load}}$. Choosing $x = 0$ at the load and moving an electrical distance θ toward the generator from the load causes the reflected wave V_R to have a negative change of phase, equal to θ, relative to its phase at the load. The incident wave V_I has a positive change of phase equal to θ, but the incident voltage is in the denominator, hence

$$\Gamma = \Gamma_L e^{-2j\theta} \tag{5.1-4}$$

Moving an electrical distance θ *from the load, the reflection coefficient argument is reduced by* -2θ.

This is a key relationship. We will make use of it frequently as we use reflection coefficient change to determine the input impedance of a transmission line.

If the relationships of (4.13-4) and (4.13-5) are normalized to Z_0 (all impedances divided by Z_0), $\Gamma(x)$ and $z(x) = Z(x)/Z_0$ are related by

$$z(x) = \frac{1 + \Gamma(x)}{1 - \Gamma(x)} \tag{5.1-5}$$

$$\Gamma(x) = \frac{z(x) - 1}{z(x) + 1} \tag{5.1-6}$$

Both of these equations are *bilinear transformations* [4, p. 202]. That is, they are of the form

$$w = \frac{Az + B}{Cz + D} \tag{5.1-7}$$

where w and z are complex variables and A, B, C, and D are complex constants satisfying

$$\Delta = AD - BC \neq 0 \tag{5.1-8}$$

The bilinear transformation is a *one-to-one mapping that* has been well studied in the theory of complex functions. In the case of (5.1-5) and (5.1-6) this means that: *Every point in the $z(x)$ plane corresponds to a unique point in the $\Gamma(x)$ plane, and vice versa.*

To illustrate this z and Γ correspondence, consider the example of the load impedance $(25 + j30)\ \Omega$ terminating the 50° long, 50-Ω line described in Figure 4.14-5 and evaluated in (4.14-13). Normalized, this impedance and the corresponding reflection coefficient at the load are

$$z_L = 0.5 + j0.6 \tag{5.1-9}$$

$$\Gamma_L = \frac{z_L - 1}{z_L + 1} = \frac{0.5 + j0.6 - 1}{0.5 + j0.6 + 1} = 0.48\angle 108^\circ \tag{5.1-10}$$

Let us now determine the impedance at a point 50 electrical degrees from the load. To do so we decrease the angle of the reflection coefficient by twice the electrical length (50°) of the transmission line, or 100°, to obtain the reflection coefficient at an electrical distance $\theta = 50^\circ$ from the load:

$$\Gamma_\theta = \Gamma_L e^{-2j\theta} = 0.48\angle(108^\circ - 100^\circ) = 0.48\angle 8^\circ \tag{5.1-11}$$

The corresponding normalized impedance is

$$z_\theta = \frac{1 + \Gamma_\theta}{1 - \Gamma_\theta} = \frac{1 + 0.475 + j0.07}{1 - 0.475 - j0.07} = 2.76 + j0.48 \tag{5.1-12}$$

$$Z_\theta = z_\theta(50\ \Omega) = (138 + j24)\ \Omega \tag{5.1-13}$$

These results are plotted in the Γ plane diagram in Figure 5.1-2.

Our strategy has succeeded. The load impedance $25 + j30\ \Omega$ was normalized to the 50-Ω transmission line and the corresponding reflection coefficient determined ($\Gamma_L = 0.48\angle 108^\circ$). This reflection coefficient was transformed along the 50° long line by preserving its magnitude (0.48) and reducing its angle (108°) by twice the electrical length of the line (100°). The resulting reflection coefficient ($\Gamma_L = 0.48\angle 8^\circ$) was used to calculate the normalized impedance ($z_\theta = 2.76 + j0.48$) at the input to the line, and this value, when unnormalized to 50 Ω, gives the input impedance $Z_\theta = (138 + j24)\ \Omega$. This is close to the value of $(139 + j24)\ \Omega$ obtained for the same example in (4.14-13) by using the impedance transformation formula. The small difference in the real part is likely due to round-off error. Notice that we are able to treat the electrical dis-

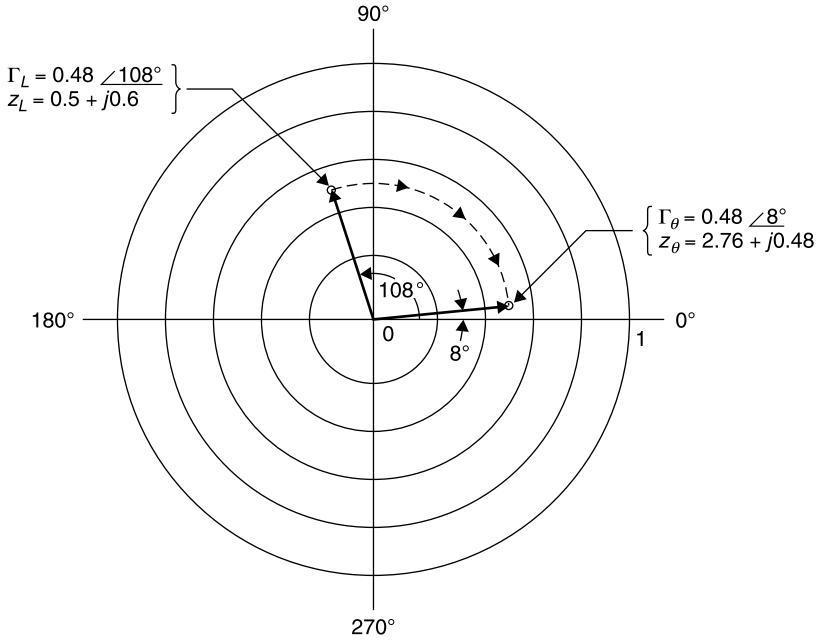


Figure 5.1-2 The Γ plane representation of load impedance z_L normalized to 50 Ω and transformed through a 50° length of 50-Ω transmission line.

tance of the line to the load, θ, as a positive quantity, since the clockwise rotation of Γ takes its sign into account.

While successful, our strategy of using the simple rotation of Γ to calculate the impedance transformation of a load by a line length has not resulted in any real economy of computational effort due to the need to perform the complex transformations from z to Γ (5.1-6) and then from Γ to z (5.1-5).

However, we note that *once a z plane point is mapped into the* Γ *plane it is done once and for all*, since the z and Γ points have a one-to-one correspondence (Fig. 5.1-3). Thus, plotting normalized impedance values in the Γ plane in advance would permit entering the Γ unit circle directly with a normalized load impedance value and, after performing the required 2θ clockwise rotation, reading the corresponding normalized input impedance to the line. *This is the basis of the Smith chart*.

However, performing this translation of the z to the Γ plane poses some obvious problems. First, the flags used to identify the corresponding z and Γ points take up so much room on the Γ chart that it would be impractical to show a sufficient number of them to permit accurate plotting of z. Second, the z plane is semi-infinite for passive loads, corresponding to all impedances for which $r \geq 0$; hence an unduly large sheet of paper would be required to ac-

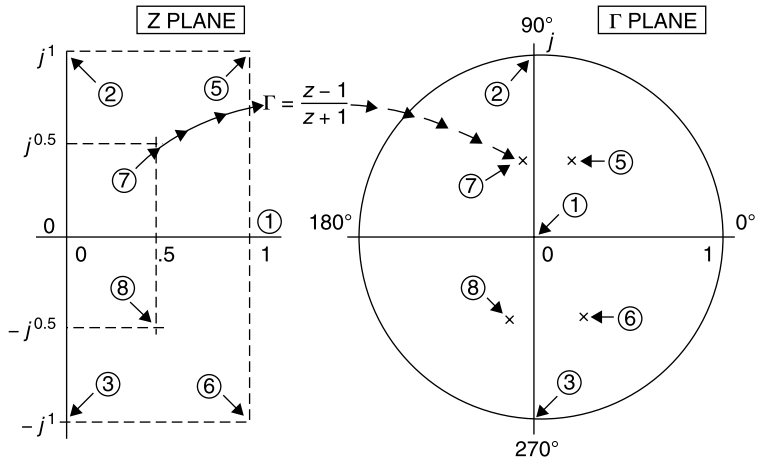


Figure 5.1-3 Mapping discrete impedance points from normalized z plane to Γ plane under the transformation $\Gamma = (z-1)/(z+1)$.

commodate large impedances in the z plane. We note, however, that no such problem exists in the Γ plane, wherein all passive loads must yield a $|\Gamma| \leq 1$, thereby fitting neatly within the unit circle.

This latter observation raises an interesting philosophical question. How can the points of a semi-infinite area (the right half of the z plane) be mapped into a finite area within the $|\Gamma| \leq 1$ circle? The answer is similar to that of the question: "How many angels can dance on the head of a pin?" The answer is: "All the angels can, because angels do not require any space." Similarly, impedance points, no matter how many of them, take up no area and therefore can be mapped from any area into any other area.

The solution to the two problems stated above lies in finding a means of *mapping* not individual points but contours from the z plane to the Γ plane. When we use the z plane, we find the intersection of the $r = \text{const}$ and $x = \text{const}$ lines for the particular impedance at hand. Once these contours are mapped onto the Γ plane (the Smith chart), we can enter the Γ plane directly, using the normalized load impedance on a line; and, after the required clockwise rotation, can read directly the corresponding normalized input impedance at that point on the line. No complex calculations would be required to perform the input impedance calculation of (5.1-1).

In summary, Philip Smith's first important idea in developing the Smith chart was recognizing that reflection coefficient rather than impedance should be used to track movement on a transmission line. The second important idea was the mapping of $r = \text{const}$ and $x = \text{const}$ contours in the z plane to the Γ plane. The Γ plane with this mapping of normalized impedance is called the *Smith chart*.

5.2 DRAWING THE SMITH CHART

The normalized, passive impedance Smith chart is the polar plot of $|\Gamma| \leq 1$ *with an overlay of* $r = const$ *and* $x = const$ *contours that relate* Γ *to* z. Normalizing the Smith chart allows its use with any characteristic impedance transmission line. Alternatively, one can draw a Smith chart for use with a given absolute value of Z_0, such as 50 Ω.

The normalized, passive impedance Smith chart consists of constant resistance and reactance contours mapped from the z to the Γ plane according to the function

$$\Gamma = \frac{z-1}{z+1} \quad \text{where } z = r + jx,\ r \geq 0,\ -\infty \leq x \leq +\infty \tag{5.2-1}$$

The mapping function of (5.2-1) is a bilinear transformation, which has four properties important to the drawing of the Smith chart [4]. These are:

1. *The semi-infinite right half of the z plane (for* $r \geq 0$*) is mapped into the* $|\Gamma| \leq 1$ *unit circle.* (Note that the left half of the z plane for which $r < 0$ would map into the semi-infinite area outside of the $|\Gamma| > 1$ circle, the negative resistance domain.)
2. *The transformation maps circles into circles*, with straight lines being considered circles of infinite radius and points being circles of zero radius.
3. *The mapping is conformal (angle preserving)*. Contours in the z plane that are orthogonal to each other will map into contours in the Γ plane that likewise are orthogonal to each other.
4. *The mapping is analytic*. Therefore, continuous contours in the z plane are mapped into continuous contours in the Γ plane.

Usually the reflection coefficient Γ is expressed in polar form as $\rho e^{j\varphi}$ or $\rho \angle \varphi$; however, it is difficult to recognize the equation of a circle in this format unless the circle's center lies at the origin. Therefore, to demonstrate this mapping, we express Γ in its Cartesian format, $\Gamma = e^{j\varphi} = u + jv$, retaining the same coordinate center for both polar and Cartesian representations.

The relationships between polar and rectangular coordinates are, from polar to rectangular

$$u = \rho \cos \varphi \tag{5.2-2a}$$

$$v = \rho \sin \varphi \tag{5.2-2b}$$

and from rectangular to polar

$$\rho = \sqrt{u^2 + v^2} \tag{5.2-3a}$$

$$\varphi = \tan^{-1}\left(\frac{v}{u}\right) \tag{5.2-3b}$$

Because a complex function such as $\Gamma = u + jv$ has two parts, real and imaginary, as does its complex independent variable, $z = r + jx$, it is not possible to graph $\Gamma(z)$ versus z in a single plane as we would the real variable $f(x)$ versus x. Instead we will show how a contour in the z *plane*, such as $r = 2$, is *mapped* into a corresponding contour in the Γ *plane*.

In this mapping task, we wish to set $r = \text{const}$ for one set of contours and then $x = \text{const}$ for the other, orthogonal, set of contours to be drawn in the Γ plane. To do this it proves convenient to use the relation that gives z as a function of Γ. In this way z can be set either to a constant r or jx and the expression solved for the Γ rectangular coordinates. Thus,

$$z = \frac{1 + \Gamma}{1 - \Gamma} \tag{5.2-4}$$

where z and Γ continue to be functions of position x on the transmission line (not to be confused with normalized reactance, x).

The relation (5.2-4) can be rewritten as

$$r + jx = \left(\frac{1 + \rho e^{j\varphi}}{1 - \rho e^{j\varphi}}\right)\left(\frac{1 - \rho e^{-j\varphi}}{1 - \rho e^{-j\varphi}}\right)\left(\tag{5.2-5}$$

Carrying out the indicated multiplication to remove the complex quantities from the denominator and noting that $e^{j\varphi} - e^{-j\varphi} = 2j \sin \varphi$ and $e^{j\varphi} + e^{-j\varphi} = 2 \cos \varphi$ we obtain

$$r + jx = \frac{1 - \rho^2 + 2j\rho \sin \varphi}{1 + \rho^2 - 2\rho \cos \varphi} = \frac{1 - u^2 - v^2 + 2jv}{1 + u^2 + v^2 - 2u} \tag{5.2-6}$$

This complex equation requires that separately both the real and imaginary parts of each side of the equation be equal to each other. Equating the real parts,

$$r = \frac{1 - u^2 - v^2}{1 + u^2 + v^2 - 2u} \tag{5.2-7}$$

We know from the properties of the bilinear transformation that (5.2-7) must be the equation of a circle in the $\Gamma = u + jv$ plane for $r = \text{const}$. But it requires some algebraic manipulation [5, p. 75] to cast (5.2-7) into the easily recognizable expression for a circle with center at $u_1 + jv_1$ and radius a, namely the format

$$(u - u_1)^2 + (v - v_1)^2 = a^2 \tag{5.2-8}$$

To do so, rewrite (5.2-7) as

$$r(1 + u^2 + v^2 - 2u) = 1 - u^2 - v^2$$

$$u^2(r + 1) - 2ur + (r - 1) + v^2(r + 1) = 0$$

$$u^2 - \frac{2ur}{r+1} + \frac{r-1}{r+1}\frac{r+1}{r+1} + v^2 = 0$$

$$\left(u - \frac{r}{r+1}\right)^2 + v^2 = \left(\frac{1}{r+1}\right)^2 \tag{5.2-9}$$

The last form of (5.2-9) is the recognizable expression for a circle in the Γ plane with center at

$$u = \frac{r}{r+1} \qquad jv = 0$$

and radius equal to $1/(r + 1)$. For example, the contour $r = 1$ in the z plane is mapped into a circle having radius 0.5 and center at $u = 0.5$, $jv = 0$, as shown in Figure 5.2-1.

At this point in the mapping, we note that all impedance values having $r = 1$ as a real part must lie on the $r = 1$ circle in the Γ plane. We do not yet know how to place specific impedances, on this circle, since we have not yet developed the orthogonal x contours for the Γ plane. In a limited sense, we can deduce that $\Gamma = 0$ when $z = 1 + j0$, and that accordingly this impedance point must be the one at which the $r = 1$ circle passes through the origin, since the

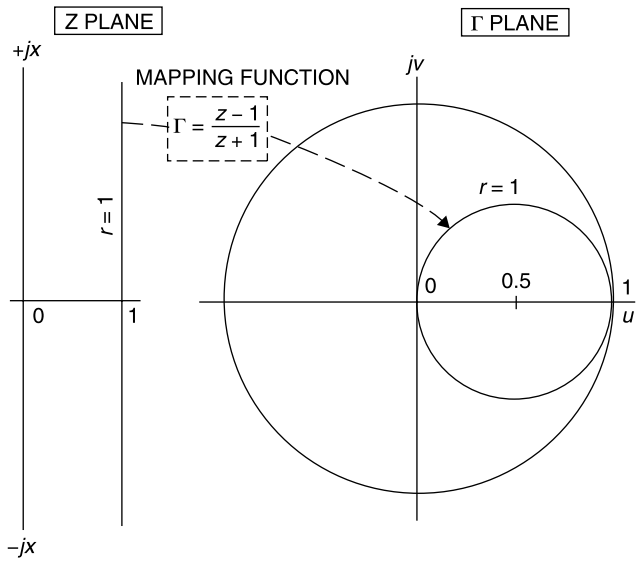


Figure 5.2-1 Contour $r = 1$ in z plane is mapped into corresponding contour in Γ plane.

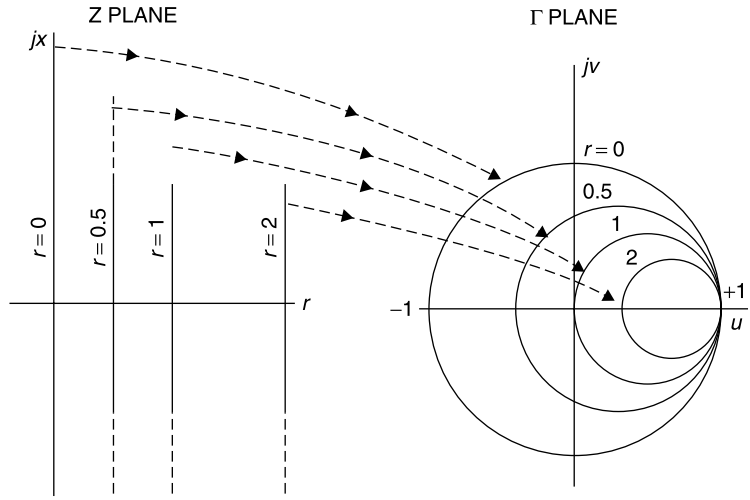


Figure 5.2-2 Constant-resistance lines of z plane map into circles tangent at $u = 1$, $jv = 0$ in Γ plane and centered on real axis.

corresponding reflection coefficient magnitude must be zero. We can also say that at $u = 1$, $jv = 0$, $\Gamma = +1$ (an open circuit), and so this must correspond to the $z = 1 \pm j\infty$ point.

As additional contours are plotted, the constant r lines in the z plane map into circles that are tangent to each other at the $u = 1$, $jv = 0$ point in the Γ plane, as shown in Figure 5.2-2.

Next, we map the orthogonal $x = \text{const}$ contours into the Γ plane. To do so, equate the imaginary parts on both sides of (5.2-6) to obtain

$$x = \frac{2v}{1 + u^2 + v^2 - 2u} \tag{5.2-10}$$

This is manipulated as

$$u^2 - 2u + 1 + v^2 - \frac{2v}{x} + \frac{1}{x^2} = \frac{1}{x^2}$$

$$(u - 1)^2 + \left(v - \frac{1}{x}\right)^2 = \left(\frac{1}{x}\right)^2 \tag{5.2-11}$$

For x equal to a constant, this is the equation of a circle with center at $u = 1$, $jv = j/x$ and radius equal to $1/x$. When contours are mapped for $x = -2$, $-1, -0.5, 0, +0.5, 1$, and 2, the results give partial circles within the $|\Gamma| \leq 1$ unity circle of the Γ plane. The circles can be extended beyond this region, but the corresponding impedances have negative real parts, and our interest for

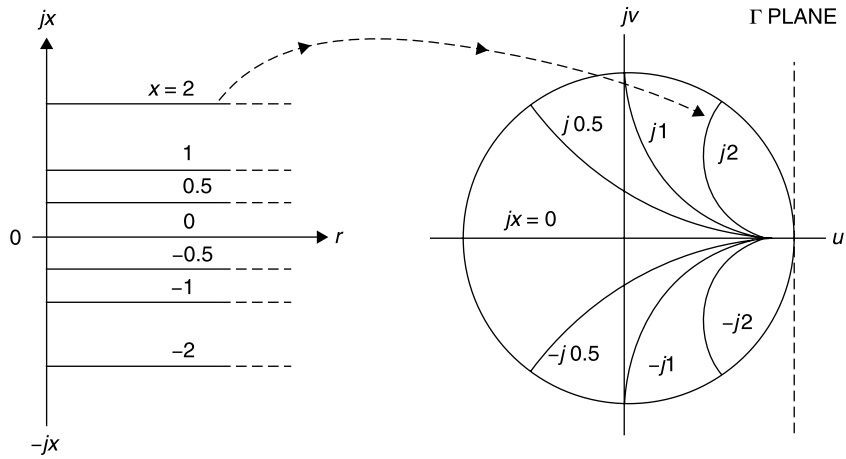


Figure 5.2-3 Constant-reactance x lines of z plane map into circles in Γ plane, but only portions of circles fall within the unit circle, corresponding to passive impedances.

now is the Smith chart for use with passive impedances only, that is, for reflection coefficients no greater than unity. More about negative resistance Smith charts later.

The contours in the Γ plane shown in Figures 5.2-2 and 5.2-3 form orthogonal sets, mapping the $z = r + jx$ values into the reflection coefficient plane, as we desired. Notice that, while the contours are circular, at their intersections their lines are at right angles (orthogonal) to each other. When both sets of contours ($r = \text{const}$ and $x = \text{const}$) are combined, the result is the skeletal form of the Smith chart shown in Figure 5.2-4.

Notice that *positive reactances (inductances) lie in the upper half and negative reactances (capacitances) in the lower half of the Smith chart* (as was true in the z plane).

The diagram of Figure 5.2-4 is very comprehensible because it shows the evolution of the Smith chart by presenting all of the necessary coordinates and scales. However, if the Smith chart is to be used graphically, as was the original intent, it must contain many more impedance contours in order that impedance can be read more precisely. For this reason the conventional format for the Smith chart is as shown in Figure 5.2-5.

Because so many r and x contours are required for a Smith chart that is to be used graphically to determine impedance transformations on transmission lines, there is insufficient space within the $|\Gamma| = 1$ circle to print explicit scales for the magnitude of reflection coefficient, even though *reflection coefficient is the basis of the Smith chart*. Also, the j is not used ahead of reactive impedances or susceptive admittances.

That *the Smith chart fundamentally is a plot of reflection coefficient* is emphasized here because it is not uncommon for individuals to become quite

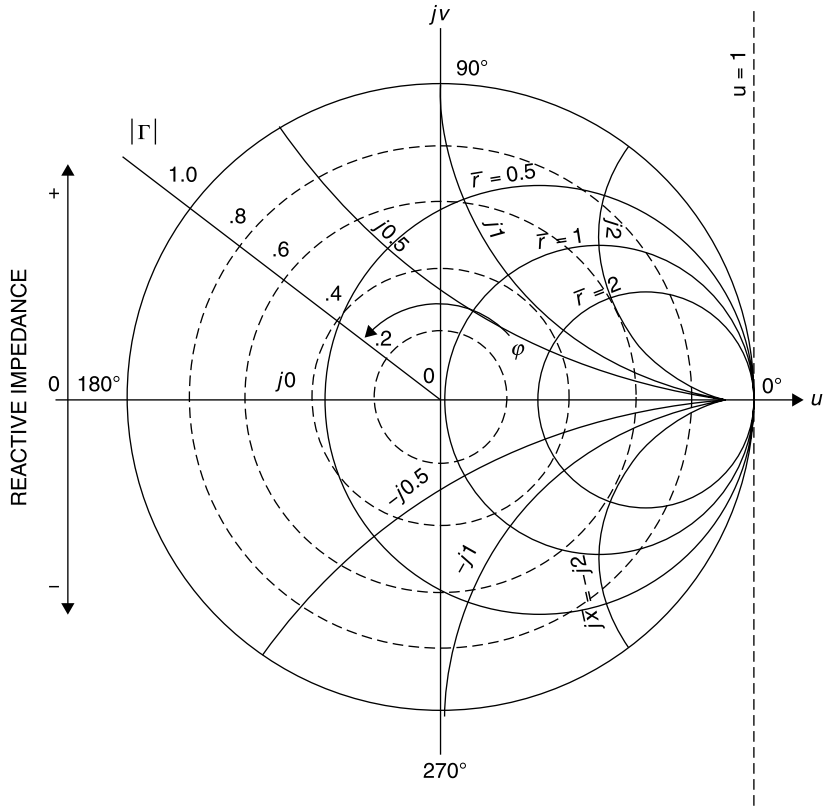


Figure 5.2-4 Smith chart showing both impedance contours r and jx as well as $|\Gamma|$ contours and Γ's angle φ in 90° segments.

adept at using the Smith chart to transform impedances yet forget, or never have been aware of, the fact that the Smith chart is just a polar plot of reflection coefficient with impedance and/or admittance overlays.

Knowing this fact, it is easy to answer many other questions about transmission lines, such as what is the voltage magnitude at a given position x on the line. It is simply $V_I|1 + \Gamma(x)|$ where Γ is readily read for any specified point on the chart.

To facilitate the determination of reflection coefficient and other important transmission line parameters, there are radial scales printed at the bottom of the chart. These are used in combination with a pair of dividers to determine the magnitude of the reflection coefficient. Only the linear reflection coefficient magnitude scale, ρ, is necessary, since the angle of Γ can be read from the degrees scale on the periphery of the Smith chart. The other parameters, return loss, mismatch loss, and other values are readily determined, given ρ; but for convenience separate radial scales for these values are also printed below the chart.

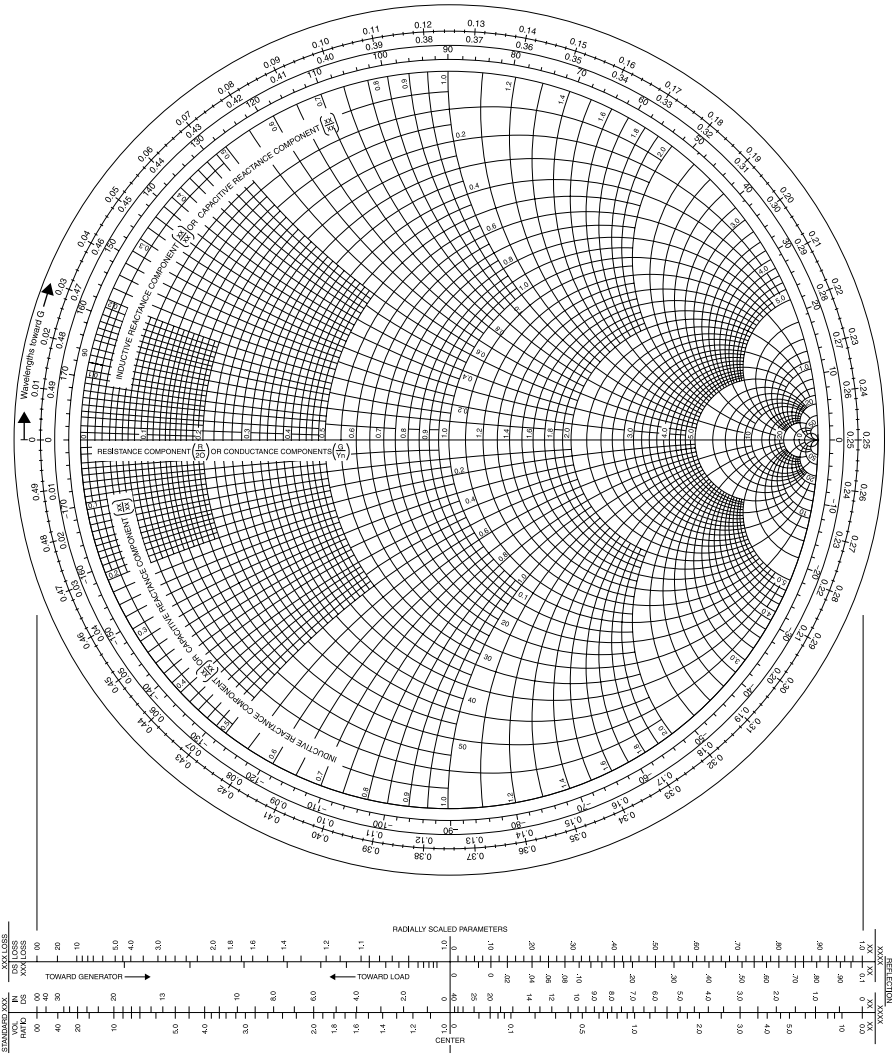


Figure 5.2-5 Smith chart as conventionally printed for use as a graphical calculation tool. (*Smith is a registered trademark of Analog Instrument Co., Box 950, New Providence, NJ 07974. Smith charts are reprinted within this book with permission.*)

5.3 ADMITTANCE ON THE SMITH CHART

At any location on the transmission line the reflection coefficient is related to the normalized impedance by

$$\Gamma = \frac{z-1}{z+1} \tag{5.3-1}$$

Since normalized admittance y is just the reciprocal of normalized impedance z,

$$\Gamma = \frac{1/y - 1}{1/y + 1} = \frac{1 - y}{1 + y} \tag{5.3-2}$$

where the bilinear transformation in (5.3-2) is simply the negative of the transformation in (5.3-1). That is, *the reflection coefficient plot as a function of y is identical to that of z except for a rotation within the* Γ *plane of 180°*. When the admittance coordinates for constant conductance g and constant susceptance b are plotted in the Γ plane, the result is the admittance version of the Smith chart, shown in Figure 5.3-1.

This plot is identical to that of the impedance plot in Figure 5.2-4 except for a 180° rotation about the center of the chart. All of the rules for transforming impedance along a line length apply equally to admittances. With normalized admittances, *the rotation moving from the load toward the generator remains clockwise*, as it was for impedances.

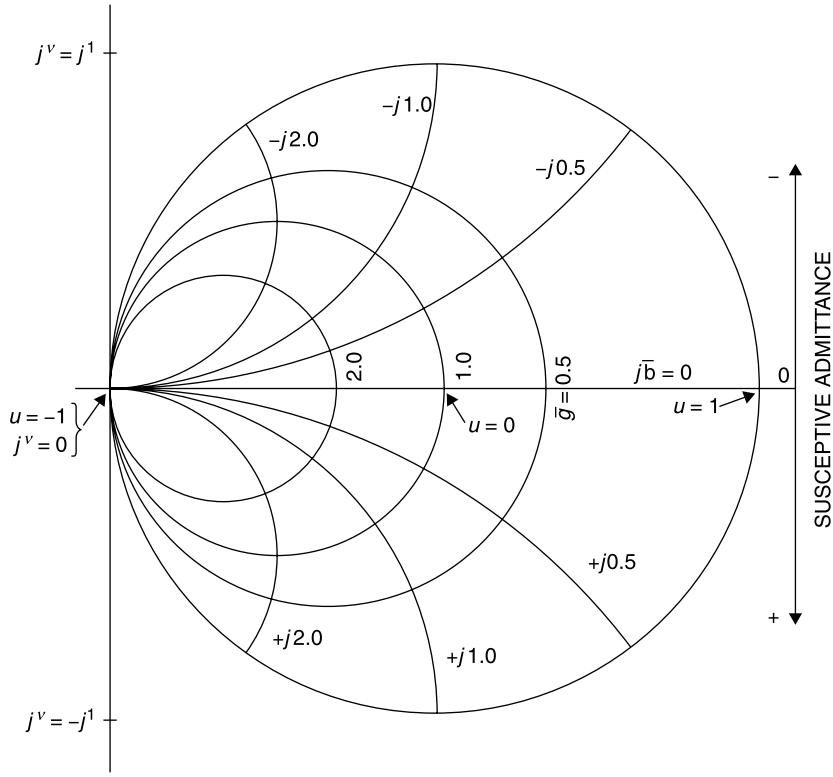


Figure 5.3-1 Smith chart with normalized admittance coordinates.

Inductors have positive reactance but negative susceptance. Capacitors have negative susceptance but positive susceptance with the result that *the upper half of the Smith chart is always inductive and the lower half is always capacitive*.

Due to the mirror symmetry of the impedance and admittance Smith charts, one could create an admittance Smith chart merely by rotating an impedance chart through 180°. In fact, for many years the practice was to consider that the normalized coordinates in Figure 5.2-5 could be defined to be either admittance or impedance, and the "standard" chart shown in Figure 5.2-5 was, and often currently is, labeled "impedance or admittance coordinates."

The problem with defining these coordinates as admittance is that the angle of the reflection coefficient is in error by 180°, and this must be corrected when interpreting the "standard" Smith chart as an admittance Smith chart. Another problem is that this practice is more prone to errors, since it is easy to forget how the coordinates are presently defined, particularly since conversions between admittance and impedance frequently must be made in the solution of a problem.

The modern and preferred procedure is to print both normalized impedance and admittance contours on the same Smith chart, overlaying the coordinates derived in Figures 5.2-4 and 5.3-1, so that one can read directly at any given point either value without resort to crossing or rotating the chart. The resulting Smith chart, having both impedance and admittance contours, is shown in Figure 5.3-2.

To aid the graphical use of this chart, it is usually printed in two colors, for example, red for normalized impedance and blue for normalized admittance. Even so, as a graphical tool it is difficult to use. Fortunately, today there are software versions of the Smith chart for use on a personal computer. Fewer contours need be drawn because the computer allows entry onto the chart at the exact value of load impedance or admittance, creates the traces of transmission line and other transformations, and displays the exact value of input impedance or admittance, as will be seen. Furthermore, either impedance or admittance contours or both can be displayed as desired. This will be demonstrated in the examples to follow.

5.4 TUNING A MISMATCHED LOAD

Originally, the Smith chart was developed to facilitate the determination of load mismatch and the requisite tuning that would match it to the characteristic impedance of the transmission line. While the chart has other useful applications, this load tuning remains a principal application of the Smith chart.

Movement along a lossless line corresponds to a contour having a constant reflection coefficient magnitude, ρ, and thus a constant radial distance from the center of the Smith chart, that is, a circle with center at the center of the chart. The angle, φ, of the reflection coefficient, $\Gamma = \rho e^{j\varphi} = \rho\angle\varphi$, changes at twice the rate of movement along the line because it is the ratio of an incident and a

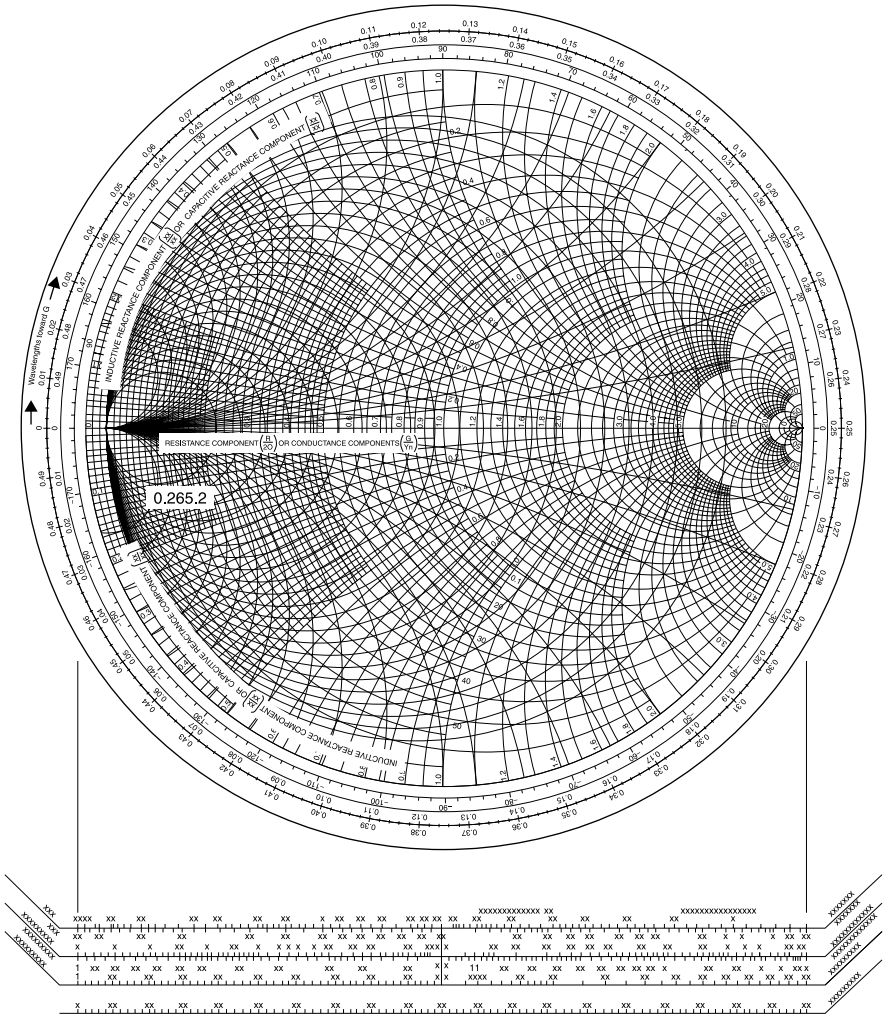


Figure 5.3-2 Smith chart with simultaneous impedance and admittance contours.

reflected wave, each of whose arguments change by an amount equal to the movement and their angles are additive in taking their ratio. *Movement from the load toward the generator on the Smith chart is clockwise.*

To facilitate the computation of movements, the Smith chart has three peripheral scales. The innermost scale gives degrees of the angle of Γ, a total of 360° for the full circle. The outer two scales are labeled ''wavelengths toward the generator'' and ''wavelengths toward the load.'' These two scales go through one-half wavelength for the full circle, thereby taking into account that the reflection angle changes by twice the angle of movement along the transmission line. This movement is demonstrated in the following example.

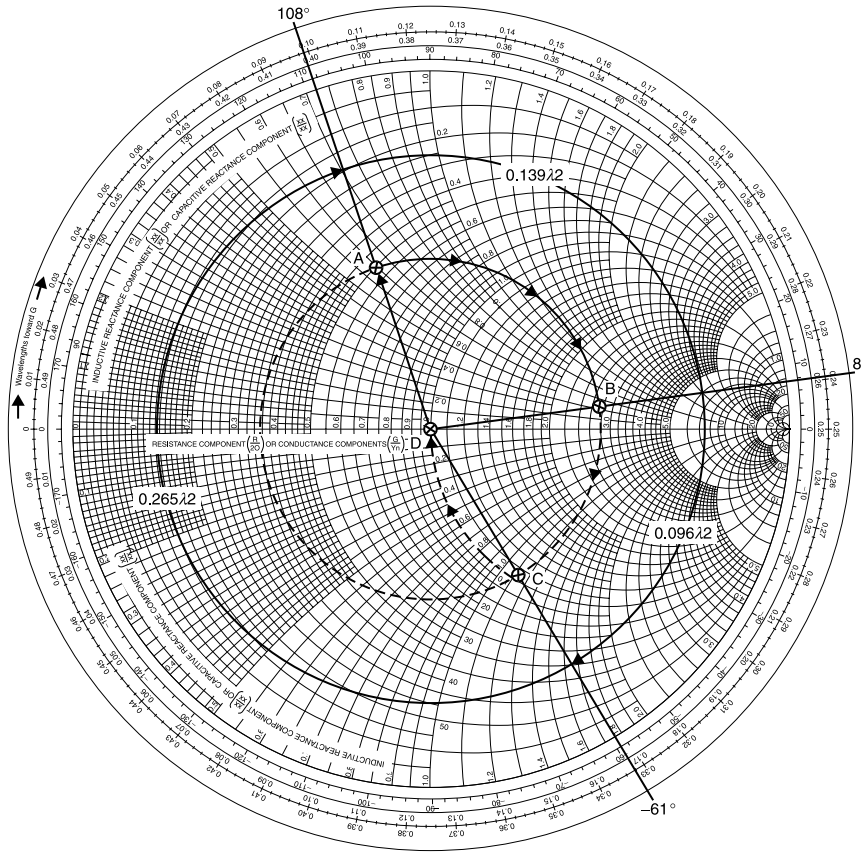


Figure 5.4-1 A $(25 + j30)$-Ω load (point A) on 50-Ω line with tuning to match it to 50 Ω (point D).

For this example, we return to the case of Figure 4.14-5 consisting of a $(25 + j30)$-Ω load at the end of a 50° long section of 50-Ω transmission line. Normalizing the load impedance to 50 Ω, it becomes $0.5 + j0.6$. This is shown as point A in Figure 5.4-1.

To determine the arc length movement on the Smith chart, we note that traveling 50° along the line *toward the generator* results in a reflection coefficient angle change of $-100°$ (*clockwise rotation*) to point B, at which we read a normalized impedance of $2.8 + j0.5$. Unnormalized (multiplying both parts by 50 Ω) this is $(140 + j25)$ Ω, reasonably close to the value of $139 + j24$ calculated in (4.14-15), given the graphical accuracy of the Smith chart, an analog instrument. Note the economy of effort with which this result is obtained relative to the solution with the input impedance formula.

As an alternative to the determination of the arc length movement on the Smith chart, we can use the "wavelengths toward generator" scale. Traveling 50° on the line toward the generator results in a clockwise movement on

the chart by $50°/360° = 0.139$ wavelengths. This dimension is shown in Figure 5.4-1 and results in the same travel from point A to point B.

From Figure 5.4-1 the value of the Smith chart is evident. The extended circular contour (shown dashed) at a constant radial distance from the center of the Smith chart reveals all of the normalized impedance values that would be encountered if the transmission line were extended to a half wavelength or longer.

Specifically, we see that, if the transmission line is lengthened until the reflection coefficient angle is $-61°$ (point C) the normalized input impedance will be $1.0 - j1.1$ ($50\ \Omega - j55\ \Omega$ when unnormalized). At point C the total change in reflection coefficient angle measured from point A is $-169°$, corresponding to a total transmission line length of 84.5°. To reach point C we must add an additional 34.5° of 50-Ω line, corresponding to a clockwise movement on the chart of an additional 0.096 wavelengths. At point C the insertion of an inductive reactance of 1.1 (55 Ω unnormalized) in series with the line cancels the capacitive reactance and moves the normalized input impedance to 1 (50 Ω unnormalized) at point D, thereby matching the load to the transmission line.

This example demonstrates that *not only does the Smith chart facilitate the calculation of the input impedance to a line terminated in a mismatched load, it also readily reveals how the mismatch can be tuned to yield a matched input*.

This tuning method is not unique. Actually, there are an infinite number of possible tuning approaches that can be determined with the skillful use of the Smith chart. By continuing clockwise on the dotted circular arc, an additional 0.27 wavelength, between points C and A, respectively, of Figure 5.4-1, we can observe the remainder of the wide range of normalized impedance values that would have been encountered on the transmission line had the electrical length, θ, been varied from 0° to 180°. The total excursion is 0.5 wavelength, after which we are returned to the starting impedance. This duplicate of the load impedance occurs every half wavelength multiple from the load along a lossless line, as was noted in Section 4.14.

In the current example, the real part of the normalized input impedance varies between 0.33 and 3, a 9-to-1 ratio. The imaginary part varies between $-j1.2$ and $+j1.2$. Using only the input impedance formula, these conclusions, obvious with the Smith chart, would have been very laborious to reach. For this reason the Smith chart is a universal presentational format for microwave measurements, usually a direct data output option for network analyzer measurements and network simulation software.

5.5 SLOTTED-LINE IMPEDANCE MEASUREMENT

From the 1930s to the 1950s, before network analyzers were available, the *slotted line* was used to measure microwave impedance. Even today, VSWR, the data value obtained with a slotted line, continues to be a common term used to describe the tolerance of impedance matching.

The slotted-line measurement was effected by cutting a nonradiating slot along the direction of propagation in a coaxial or waveguide transmission line and sliding a rectifying detector diode along the slot to sample the rectified time-average voltage amplitude as a function of position along the line. This fixture was called a *slotted line*. *The ratio of the maximum to minimum voltage values is the voltage standing-wave ratio*, or *VSWR*. The VSWR coupled with the location of the associated voltage minimum (*null*) could be used to determine the load impedance Z_L. The following example demonstrates this procedure.

First, a short circuit is connected at the end of the transmission line prior to connecting the load (Fig. 5.5-1*a*), and the positions of the resulting voltage minima, or nulls, recorded. These null locations are equivalent to the electrical position of the load plane. The rectified voltage amplitude varies sharply with position near a null (Fig. 5.5-1*a*), yielding the most precise measurement of the equivalent load positions. *The distance between two successive minima is one-half wavelength at the operating frequency*, providing a means of determining the actual operating frequency of the generator. *Each set of VSWR and null data is taken at a single microwave frequency*. To obtain data for a band of frequencies, the measurement must be re-performed at each frequency of evaluation.

For example, suppose that, with the short circuit installed, nulls are found at 2, 4.95, and 7.90 in. along the slotted line. The distance between minima is a half wavelength, therefore

$$\lambda = 2(4.95 \text{ in.} - 2.00 \text{ in.}) = 5.90 \text{ in.} \tag{5.5-1}$$

If the slotted line has air dielectric, as is usually the case to permit unimpeded probe movement,

$$f_0 = \frac{11.8 \text{ in.}}{5.9 \text{ in.}} 1000 \text{ MHz} = 2000 \text{ MHz} \tag{5.5-2}$$

This demonstrates that *the slotted line provides a means of frequency determination*.

The locations of the voltage minima can be considered equivalent to the electrical position of the load plane, repeating at half wavelength intervals.

Next, the load is attached and the rectified voltage measurements yield $V_{\text{MAX}} = 1.5$ units and $V_{\text{MIN}} = 0.5$ units (the actual voltage scale is unimportant) with null positions at 2.59 and 5.54 in. From these data

$$\text{VSWR} = \frac{V_{\text{MAX}}}{V_{\text{MIN}}} = \frac{1.5}{0.5} = 3 \tag{5.5-3}$$

$$\rho = \frac{\text{VSWR} - 1}{\text{VSWR} + 1} = \frac{3 - 1}{3 + 1} = 0.5 \tag{5.5-4}$$

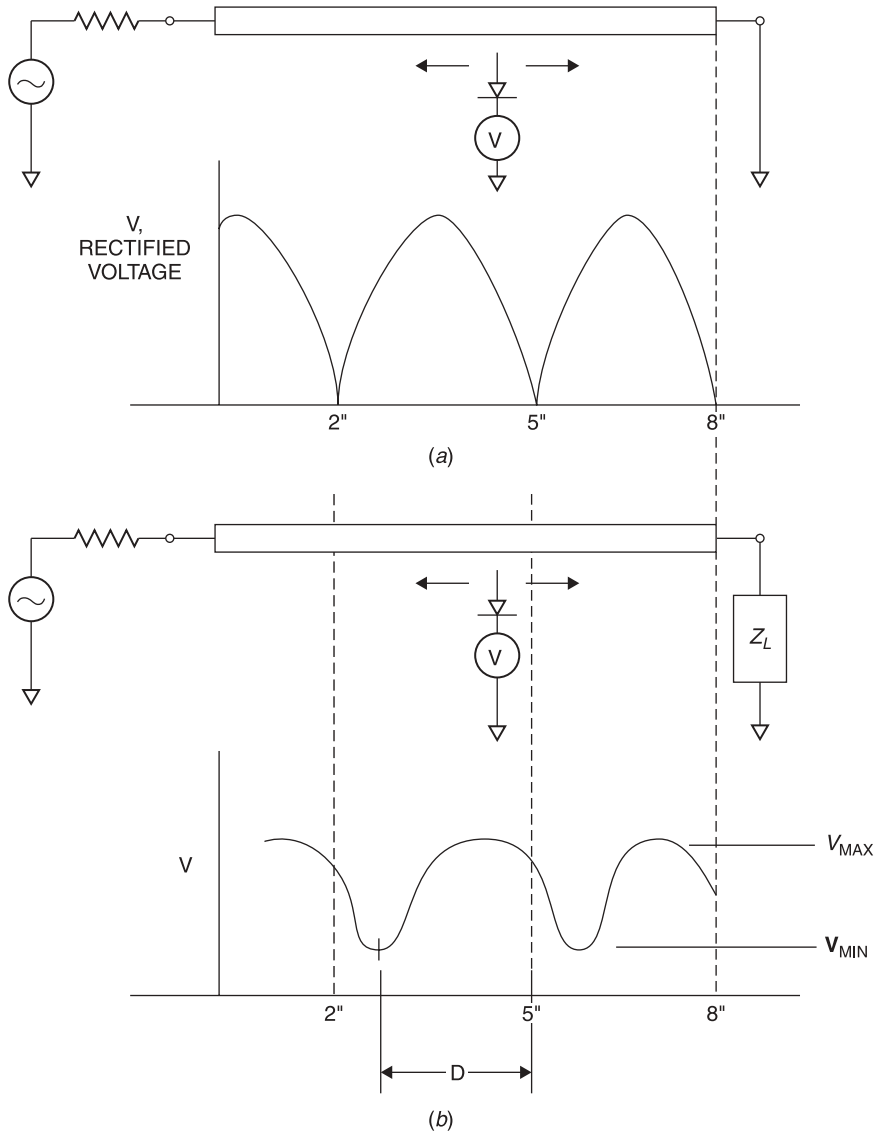


Figure 5.5-1 (*a*) Schematic diagram of slotted-line measurement with short-circuit termination and rectified standing-wave voltage with distance. (*b*) Rectified voltage with an unknown load Z_L.

The null with load in place is located a distance $D = 7.90$ in. $- 5.54$ in. $= 2.36$ in. *toward the generator from the load plane*. To locate the normalized load impedance we convert this distance to wavelengths.

$$D = \frac{2.36 \text{ in.}}{5.90 \text{ in.}/\text{wavelength}} = 0.40 \text{ wavelength} \tag{5.5-5}$$

On the Smith chart a half wavelength of travel on the transmission line produces a 360° change in the argument of the reflection coefficient:

$$0.5\lambda \text{ of travel on line} \Rightarrow 360^\circ \text{ around chart} \tag{5.5-6}$$

Therefore, in degrees

$$D = 0.4\lambda\left(\frac{360^\circ}{0.5\lambda}\right) = 288^\circ \tag{5.5-7}$$

The circumferential scales on the Smith chart read both in degrees of reflection coefficient angle and in wavelengths. Starting from the short-circuit point ($z = 0, \varphi = 180^\circ$) the distance D is shown plotted in the Smith chart of Figure 5.5-2. Note that the 288° (0.4λ) arc is drawn *counterclockwise toward the load*

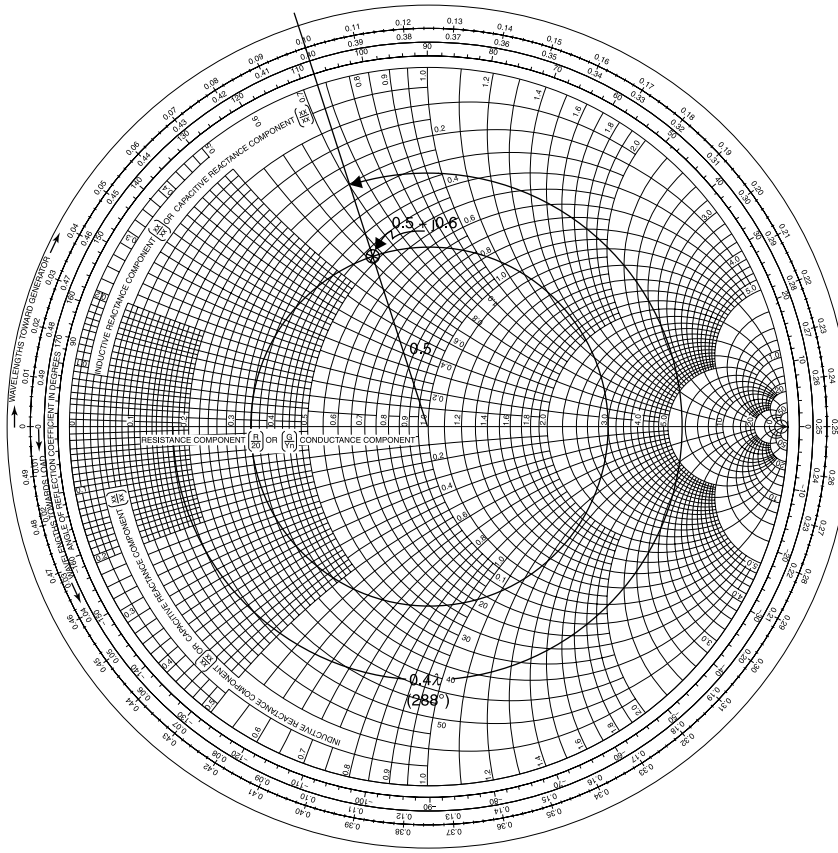


Figure 5.5-2 Smith chart used to compute normalized load impedance from slotted-line measurements. (*Smith chart reproduced through the courtesy of Analog Instrument Co., Box 950, New Providence, NJ 07974.*)

since this is the direction from the load minimum to the plane of the load in question.

The angle of the reflection coefficient at the load is read directly from the peripheral scale on the Smith chart as

$$\varphi = 108^\circ \tag{5.5-8}$$

and hence

$$\Gamma_L = \rho\angle\varphi = 0.5\angle 108^\circ \tag{5.5-9}$$

At this point the normalized load impedance can be read directly from the chart or calculated from

$$z_L = \frac{1+\Gamma_L}{1-\Gamma_L} = \frac{1+0.5\angle 108^\circ}{1-0.5\angle 108^\circ} = 0.5 + j0.6 \tag{5.5-10}$$

Unnormalized this is $(25 + j30)\ \Omega$, the value used in the example of Figure 4.14-5.

Often the difficult part of the slotted-line measurement is deciding in which direction to rotate on the Smith chart from the null position obtained with the load in place. It may help to remember that if the load impedance were known and plotted on the Smith chart, one would make a negative angle rotation (*clockwise from the load toward the generator*) to the position of the minimum impedance (the $\pm 180^\circ$ axis of the chart). Therefore, to travel from the null *toward the load rotate counterclockwise* on the Smith chart.

The reader will notice from this example, in which the VSWR is 3, that the constant reflection coefficient circle intersects the $r = 3$ circle on the 0° axis of the Smith chart. The circle always intersects the r circle at the value of VSWR, as will be demonstrated in the next section.

5.6 VSWR = *r*

When a constant magnitude reflection coefficient circle of radius greater than zero is drawn on the Smith chart, it intersects the horizontal axis of the chart at two points, r and $1/r$. In the case of a matched load, $\rho = 0$, the circle has zero radius, and the two intersections coalesce into one point, the center of the chart at which $r = 1$, $x = 0$.

For the intersection for which $r > 1$, $z = r + j0$, the reflection coefficient has a real value given by

$$\Gamma = \frac{z-1}{z+1} = \rho = \frac{r-1}{r+1} \tag{5.6-1}$$

The corresponding VSWR is

$$\text{VSWR} = \frac{1+\rho}{1-\rho} = \frac{r+1+r-1}{r+1-r+1} = r \tag{5.6-2}$$

For this reason *the VSWR for any normalized load impedance can be obtained by drawing the constant* ρ *circle through it. The VSWR is numerically equal to the resistance value r, corresponding to the circle's intercept with the horizontal Smith chart axis for which* $r > 1$.

5.7 NEGATIVE RESISTANCE SMITH CHART

The area that is outside the unit circle of the Smith chart corresponds to reflection coefficient magnitudes greater than one, the negative resistance and conductance domain. The resistance circles defined by (5.2-9) can be plotted for negative values of r, and the circles for constant reactance can be extended outside the $\rho = 1$ circle. For $\rho \leq 3.16$ (a reflection gain of 10 dB) the result is that shown in Figure 5.7-1. The example of impedance Z_{IN}/Z_0 and admittance Y_{IN}/Y_0 shown on this standard chart is consistent with identifying the admittance of a point on a 180° diametrically opposite location. However, as was noted earlier, the chart can and should be drawn with simultaneously valid Z and Y coordinates in the format of Figure 5.3-2.

Unlike the passive impedance/admittance domain of the Smith chart, the negative resistance domain is unbounded. A new chart must be drawn for each maximum anticipated value of ρ.

The negative resistance domain of the Smith chart (where $\rho > 1$) is defined by the same equations used for the passive Smith chart. This negative resistance domain is useful, for example, to describe reflection amplifiers and to plot the stability circles for transistor amplifiers. However, for most matching designs, the area within the passive impedance Smith chart is sufficient and our matching examples are confined to this area.

5.8 NAVIGATING THE SMITH CHART

For lossless matching, two principal paths, or *contours*, on the Smith chart are employed. These are

1. Constant radius contours for movement along a transmission line and
2. Constant resistance (or conductance) contours for the addition of reactance (or susceptance)

Movement on constant reactance (or susceptance) circles is also possible but implies the addition of resistance (or conductance), consequently lossy match-

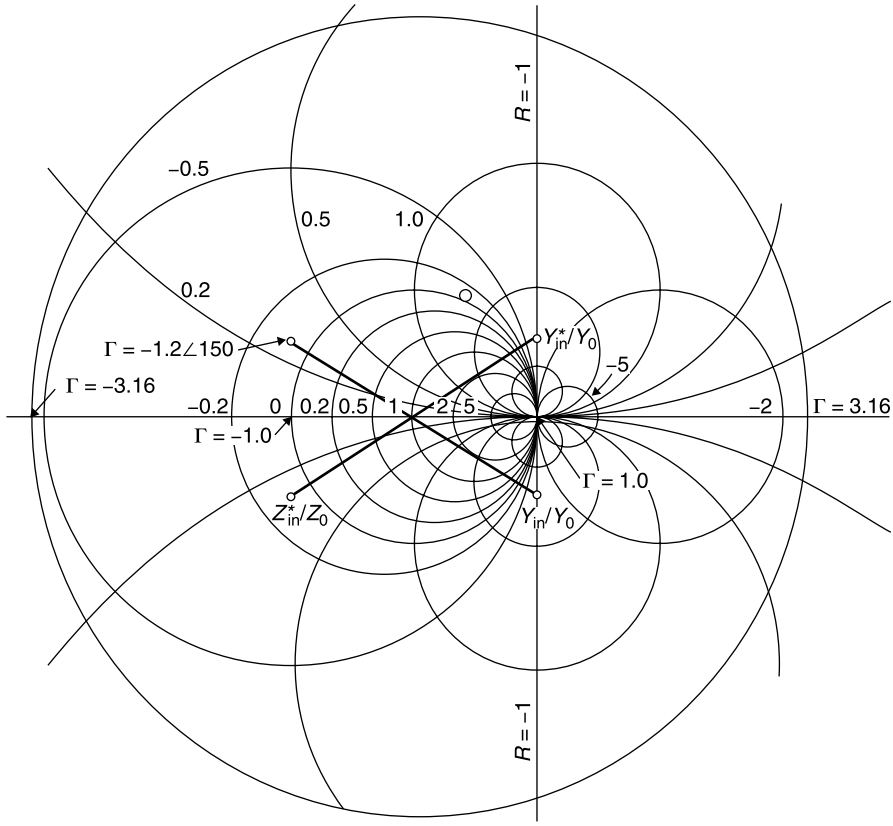


Figure 5.7-1 Smith chart extended to include the negative resistance/conductance domain out to a reflection coefficient magnitude of 3.16. This is sometimes called the "compressed Smith chart." (*Negative* impedance *Smith chart reproduced through the courtesy of Analog Instrument Co., Box 950, New Providence, NJ 07974.*)

ing. Lossy matching has its place, as for very broadband matching, but it is usually avoided in favor of lossless matching.

If the load on a transmission line contains no resistance or conductance, (1) the load can absorb no power, (2) the reflected and incident waves have the same magnitude, (3) the reflection coefficient magnitude is unity ($\rho = 1$), and (4) Γ traces the rim of the passive Smith chart, traversing every reactance (and susceptance) from $-j\infty$ to $+j\infty$ in a transmission line travel of one-half wavelength. This means that open- or short-circuited transmission lines, or stubs, can be used as reactive/susceptive tuning elements.

The next example demonstrates these Smith chart contours and stub matching. A 100-Ω load is connected in series with a 1-pF capacitor. The resulting impedance at 1 GHz of $(100 - j159)$ Ω is to be connected to the end of a 50-Ω transmission line using wire leads having a total of 10 nH of inductance

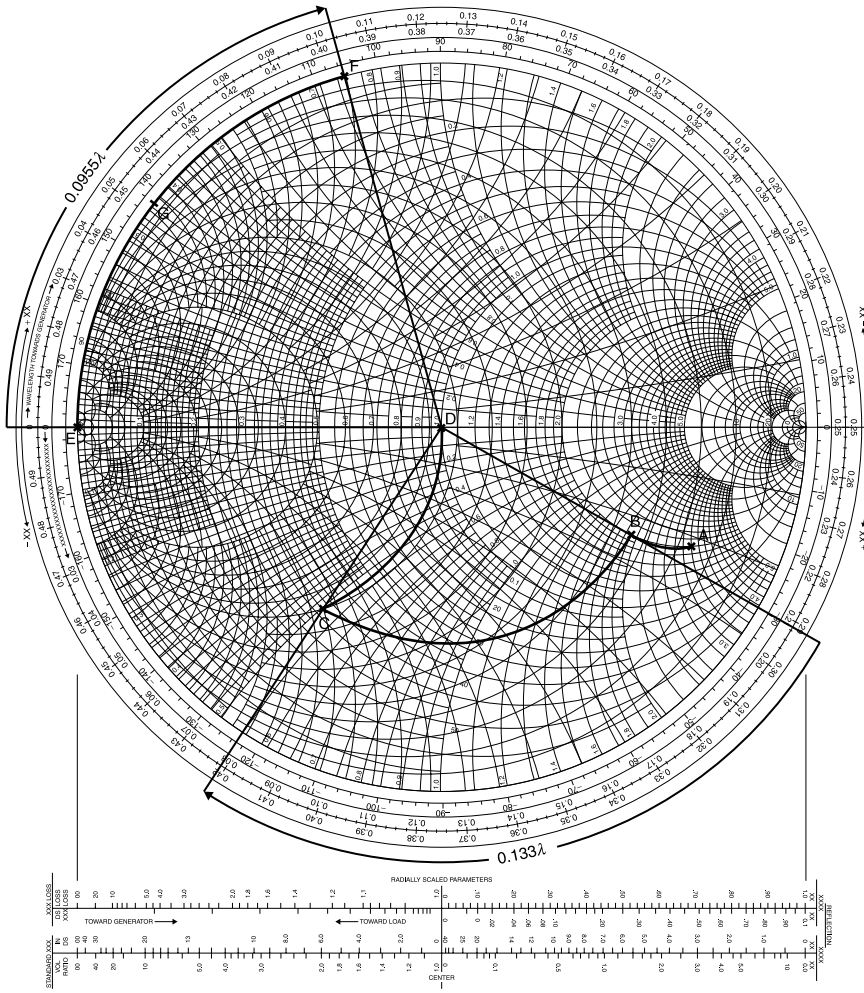


Figure 5.8-1 Procedure to match $(100 - j159)$-Ω load with a $+j63$-Ω series inductance to 50 Ω using a length of 50-Ω line and a short-circuited 50-Ω shunt stub.

($+j63$ Ω at 1 GHz). The resulting load is to be matched to 50 Ω using a shunt tuning stub. The stub may be either open or short circuit terminated. The type of stub termination, the length of the stub, and its location on the main line is to be determined to match this load to 50 Ω (Fig. 5.8-1).

The matching is carried out in the following steps:

1. The reference impedance of the Smith chart is defined to be 50 Ω (the chart is normalized to 50 Ω).

2. The load impedance $(100 - j159)\ \Omega$ is divided by 50 Ω to normalize it, giving $2 - j3.2$ (to the accuracy consistent with the graphical process). This is point A in Figure 4.8-1.
3. The $+j63$-Ω series inductance has a normalized value of $+j1.3$. This is added to the load impedance by *traveling along the constant resistance circle* ($r = 2$) to the reactance intercept $-j1.9$. The initial load (point A) is in the lower half of the Smith chart since its reactance is capacitive. Adding the series inductance causes an upward movement, toward the inductive upper half of the chart, to point B.
4. Drawing a clockwise, constant reflection coefficient magnitude arc intersects the unity *conductance circle* at point C. This point is important. We wish to add a shunt stub. The shunt addition requires the use of admittance coordinates, hence intercept on the unity conductance circle.

   Using the wavelength scale of the chart's periphery, we determine that the starting point B is at 0.292λ and that the $g = 1$ intersection, point C, is at 0.425λ. The distance traveled is $0.425\lambda - 0.292\lambda = 0.133\lambda$, or 48°. This is the distance from the load at which the shunt tuning stub is to be connected. The length of line between load and stub is 48°. This is not to be confused with the angular change in Γ, which is $-2(48°) = -96°$.

5. At point C the normalized admittance is read from the chart to be $1 + j1.5$. To obtain a match, add a normalized susceptance of -1.5 thereby *traveling on the* $g = 1$ *circle* to the center of the chart, point D, completing the match.
6. Short-circuited or open-circuited transmission lines can be used as tuning elements of any desired reactance. We have previously seen that such stubs are described by (4.14-7a) and (4.14-8a), but the Smith chart allows their reactances or susceptances to be read directly. We are free to use any characteristic impedance stub. Suppose for this example that we choose $Z_0 = 50\ \Omega$.

   Then the same Smith chart normalization applies, and we can perform this calculation on the chart used to determine the matching requirement. To obtain the required $-j1.5$ susceptance (an inductance) we begin with a *shorted stub* at the $y = \infty$ location (point E) on the Smith chart and *travel clockwise along the chart's periphery* to the required inductive susceptance at point F. This requires a stub length of 0.0955λ, or about 34.4°. It is emphasized that this is a side calculation and this travel is not part of the original contour from A to D.

   Shorted stubs are inductive when their length is less than a quarter wavelength. An open-circuited, 50-Ω stub could be used, but its length would be a quarter wavelength longer (0.3455λ instead of 0.0955λ). Consequently, its frequency sensitivity would be nearly four times greater than that of the shorted stub.

Finally, the same Smith chart can be used even if the tuning stub impedance is different from 50 Ω. For example, suppose a 100-Ω stub is used, then

$$Y_{\text{STUB}} = -j1.5(0.02\ \text{℧}) = -j0.03\ \text{℧}$$

$$y_{\text{STUB}}\ (\text{normalized to } 100\ \Omega) = \frac{-j0.03}{0.01} = -j3$$

This is point G in Figure 5.8-1 and corresponds to a stub length of about 0.05λ. Notice that for this side calculation the Smith chart is normalized to 100 Ω. This is a shorter stub and would take up much less space in the circuit layout, both because it is half the length of the 50-Ω stub and because 100-Ω stubs have much narrower width, hence can be "serpentined" more easily. Shorter stubs also can have less frequency variation, however, both the 50- and 100-Ω stubs of this example are so short that their susceptances would not vary much more rapidly than inversely with frequency.

The completed circuit with 50-Ω stub is shown in Figure 5.8-2 along with its VSWR performance, referenced to 50 Ω. Errors in round off and graphical inaccuracies make the result slightly off of a perfect 1.00 VSWR and the frequency of best match slightly above the design value of 1000 MHz. This result greatly improves the efficiency of power transmission to the load, as indicated below.

When the arc described by points B and C is extended to the 0° axis of the Smith chart, it intersects it at about $r = 4.2$. Therefore, the VSWR of the load,

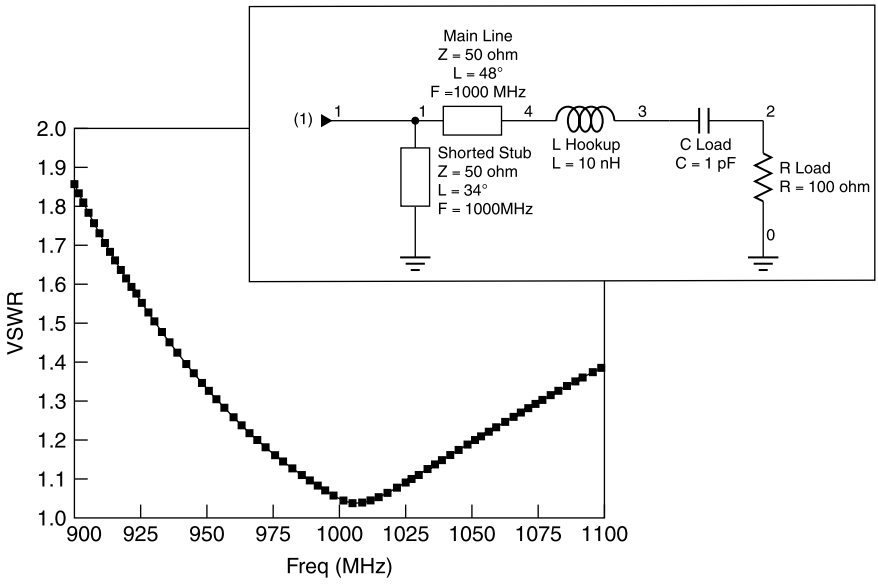


Figure 5.8-2 VSWR performance of stub tuned load.

including its series inductance, is 4.2. The reflection coefficient ρ is 0.62. This is read using the auxiliary scale (not shown in the figure) at the bottom of the standard Smith chart sheet or calculated from VSWR and (4.5-3b).

Without the stub tuning, the return loss (fraction of power returned to the generator) of the load before tuning would have been

$$\mathrm{RL} = \rho^2 \times 100\% = 38\% \tag{5.8-1}$$

Thus the mismatch loss is 62%, which corresponds to a mismatch loss of 2.1 dB. With the stub tuning, the VSWR is below 1.1 and ρ is below 0.05 from (990 – 1020) MHz. The corresponding return loss is less than 0.3%. This improvement in efficiency, easily designed using the Smith chart, is obtained merely by printing a side stub on the transmission line and grounding it.

5.9 SMITH CHART SOFTWARE

Although the procedures for using the Smith chart as a graphical tool are fairly simple, it is easy to err by rotating in the wrong direction, misreading the coordinate values or trying to match using the incorrect circle. To minimize errors, remember to: *Rotate to the $r = 1$ circle when adding a series reactance* and *rotate to the $g = 1$ circle when adding a shunt susceptance.*

Furthermore, the accuracy of the procedure is limited by the graphical process itself. Finally, the Smith chart with both impedance and admittance coordinates is a crowded format that is difficult to read.

These problems are considerably lessened through the use of *Smith chart software* written for the personal computer. The following examples are illustrated using one such program called winSmith [6].

For example, given a load of $(25 + j30)$ Ω at 2000 MHz terminating a 50-Ω transmission line, it is desired to match this load to 50 Ω using a tuning element connected in shunt with the transmission line at an appropriate distance from the load on the generator side. What is the electrical spacing, element type, and value for this tuning element?

This load point is entered as $(25 + j30)$ Ω directly into the program along with the reference impedance for the chart. The program automatically locates impedances on the chart given the reference impedance so there is no need for normalization by the user. The parameters of the cursor-selected point (x in Fig. 5.9-1) are listed in the upper right corner of the display. The units of Z and Y are ohms and mhos, respectively. The S term is S_{11} with magnitude in decibels and angle in degrees (S parameters are discussed in 6.4). The term G stands for gamma (Γ), the reflection coefficient, listed here as $0.49\angle 108°$. This is a more precise value than the $0.5\angle 108°$ value that we read graphically from the Smith chart in the earlier example. The corresponding VSWR, listed as "V" in the display, is 2.89, again a more accurate value than the 3.0 value previously read graphically from the chart.

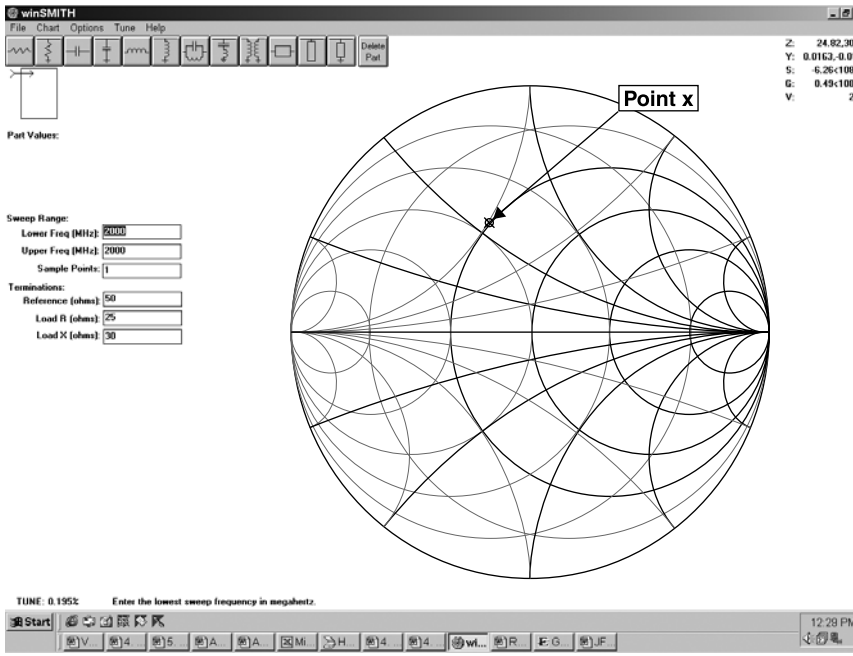


Figure 5.9-1 Personal computer display of Smith chart showing $(25 + j30)$-Ω load on 50-Ω transmission line. (*Produced using winSmith software* [6].)

A selection of pull-down components in the upper left of the display allows the construction of a cascade circuit layout of various topologies. The resulting Smith chart contour is automatically computed and displayed for the circuit chosen and the variable values entered.

Both impedance (dark lines) and admittance (light lines) coordinates are shown. Recall that the location of the load is the same using this Smith chart format, whether the load is expressed in impedance or admittance coordinates. Unlike the graphical Smith chart, the software displayed Smith chart is not normalized, since the value of Z_0 is also entered. The impedance and admittance are read directly in ohms and mhos, respectively.

The tuning procedure is simplified further by selecting for display only the desired coordinates, impedance, or admittance appropriate to the current movement being made on the chart. We wish to move along the 50-Ω line, a constant reflection coefficient magnitude, from the load (now relabeled point A in Fig. 5.9-2) to the intersection with the 0.02-℧ conductance circle (point B). Here the addition of an appropriate shunt susceptance will match the input to the line to 50 Ω (0.02 ℧). The admittance at point B lies in the lower half of the Smith chart, and therefore its imaginary part corresponds to a capacitor (positive susceptance). It is essential to keep in mind that *the lower half of the Smith*

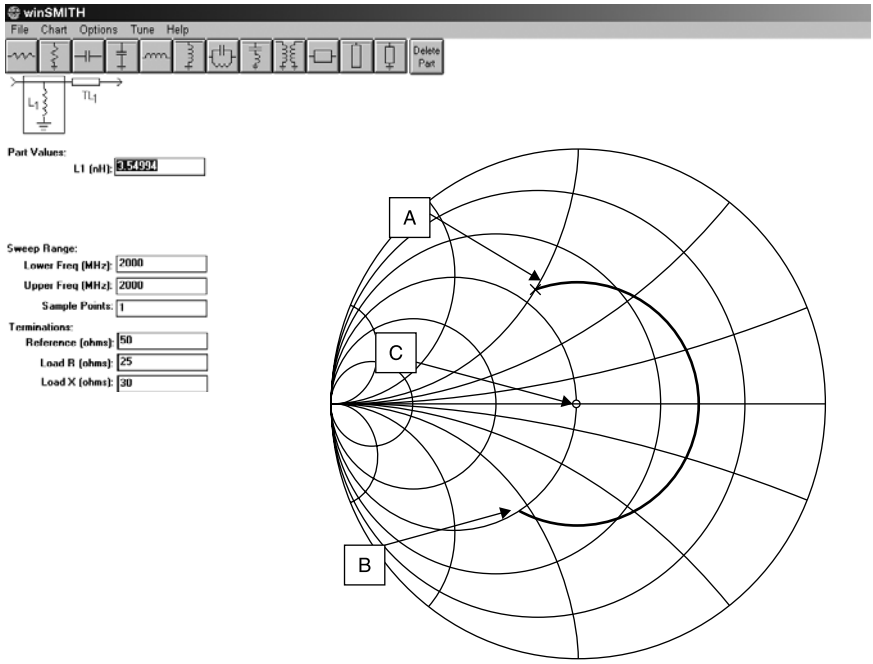


Figure 5.9-2 Smith chart used to determine shunt tuning of the $(25 + j30)$-Ω load on a $Z_0 = 50$ Ω transmission line.

chart is capacitive, and the upper half is inductive, whether expressed as impedance or admittance.

From the Smith chart, the 50-Ω line length required to move from point A to the unity conductance circle (point B) is 114° (0.32λ). Notice that this results in a clockwise rotation of −228° on the Smith chart. From the cursor reading (not visible in this view), the admittance at point B was read to be $(0.02 + j0.022)$ ℧. The addition of a 3.54-nH shunt inductor (having susceptance of $-j0.022$ ℧ at 2 GHz) moves the contour to point C, tuning the load to 50 Ω, as shown in Figure 5.9-2.

5.10 ESTIMATING BANDWIDTH ON THE SMITH CHART

Each contour on the Smith chart applies at a single frequency. However, a measure of the bandwidth of a particular tuning procedure can be obtained by adding a contour at each band edge. Suppose that we are interested in the bandwidth 1800 to 2200 MHz for the previous example, shown plotted in Figure 5.9-2. Adding these band edge contours yields the results in Figure 5.10-1. The load has been modeled as a 25-Ω resistor in series with a 2.39-nH inductor ($+j30$ Ω at 2000 MHz) in order to simulate the frequency variation of the load.

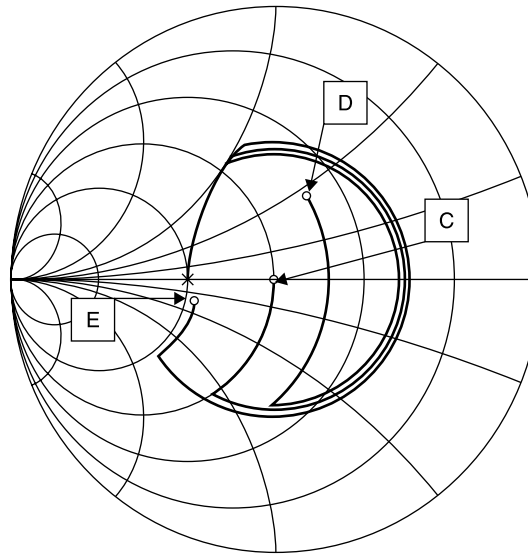


Figure 5.10-1 Smith chart contours for 1800 MHz (point *D*), 2000 (point *C*), and 2200 MHz (point *E*) for the previous tuning example.

Notice that the contours for 1800 MHz (terminating at point *D*) and 2200 MHz (terminating at point *E*) show large variations in the angle for the transmission line section. This is because the line length changes by $\pm 11.4^\circ$ at these band edges. This change is further amplified because on the Smith chart the angular change is doubled to $\pm 22.8^\circ$. Using the cursor, the VSWR values at the end of these excursions were 1.94 at 1800 MHz and 1.96 at 2200 MHz. The resulting mismatch loss is about 0.5 dB.

5.11 APPROXIMATE TUNING MAY BE BETTER

The engineer who first learns matching with the Smith chart is inclined to try for a "perfect match" at the design center frequency, on the assumption that this will also give the best performance over a given bandwidth. But this assumption is not necessarily valid. In some cases an approximate solution at the center frequency may give better overall performance. In addition, settling for a reasonably "close match" at the center frequency might be obtainable using more economical tuning elements, while actually providing a better overall match throughout the required frequency bandwidth. The following example illustrates such a case.

Referring to Figure 5.9-1, we see that the $(25 + j30)\ \Omega$ starting impedance (point *A*) lies close to but is not on the unity conductance circle. At this point

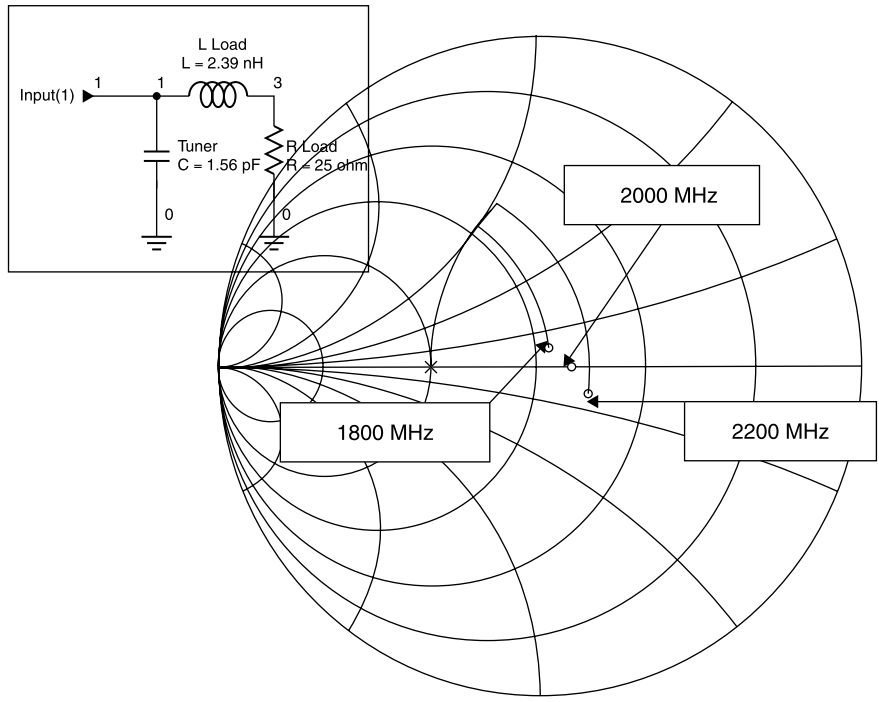


Figure 5.11-1 Result of tuning the $(25 + j30)$-Ω load in an approximate manner at f_0 using a parallel capacitor.

the admittance is $(0.0164 - j0.197)$ ℧. If a capacitor is added in shunt with the load (point A) having a susceptance of $+j0.0197$ ℧ $(-j50.8\ \Omega)$ at 2000 MHz, it must have a capacitance of 1.56 pF. The resulting total admittance at point A would be 0.0164 ℧ (61 Ω real), corresponding to a 1.22 VSWR in a 50-Ω system. With this approximate tuning at the center frequency, the three frequency contours are as shown in Figure 5.11-1.

With this tuning the VSWR values are 1.15, 1.22, and 1.42 at 1800, 2000, and 2200 MHz, respectively. The highest mismatch loss in the band, at 2200 MHz, is only 0.13 dB, much better than was obtained when the load was perfectly matched at 2000 MHz using the transmission line and shunt inductor.

This example further exemplifies the advantage of the Smith chart. The opportunity for approximate tuning was evident when the load was shown with admitttance coordinates on the Smith chart. The example demonstrates two points about matching.

First, *a broader band match usually can be obtained when the tuning is performed close to the load*. Second, *accepting an approximate match at the center frequency may result in a better average match over the operating band.*

5.12 FREQUENCY CONTOURS ON THE SMITH CHART

If the points in Figures 5.10-1 and 5.11-1 that describe low to high frequency impedances for the same location in the circuit are connected with a line, they can be seen to describe a *clockwise arc*. This is always the case, a consequence of Foster's reactance theorem.

Foster's reactance theorem states that the reactive portion of impedances and the susceptive portion of admittances increase positively as frequency increases.

Impedances and admittances on the Smith chart trace clockwise arcs as frequency is increased.

Often the swept frequency response of a network analyzer is presented on a Smith chart display without frequency labels. The clockwise increase with frequency is a ready means of identifying the low and high frequency ends of the display trace.

5.13 USING THE SMITH CHART WITHOUT TRANSMISSION LINES

Notice that in the preceding example there were no transformations along transmission lines (constant reflection coefficient magnitude arcs). The VSWR values were calculated as those which would be incurred *if* the load and its tuner were connected to a 50-Ω line.

The Smith chart can be used without transmission lines, even though it was developed as an aid to transforming impedances along them. It is an orthogonal set of resistance and reactance contours (or conductance and susceptance contours). It is especially useful for performing matching calculations because every passive impedance or admittance, including those with infinite magnitudes, can be plotted within the unit circle enclosed by the $|\Gamma| \leq 1$ Smith chart. This is not true of a rectangular impedance or admittance chart for any given scale factor, whose dimensions increase without bound with the magnitudes of the impedance and admittance components to be plotted.

When using the Smith chart without an associated transmission line, we are free to select arbitrarily its reference impedance, the Z_0 impedance to which all others are normalized. For the best graphical accuracy it is desirable to select the reference impedance near the geometric mean between the largest and smallest impedance values to be plotted.

For example, consider the Q matching example of Figure 3.5-2. We desired to Q match a 5-Ω resistor to 50 Ω at 1 GHz. To demonstrate how this matching could have been performed using the Smith chart instead of the Q matching method, we start by selecting the reference impedance of the chart arbitrarily to be 25 Ω. The 5 Ω resistance is first plotted on the Smith chart (point A) and a value of series inductance added to bring the contour to the intersection with the $G = 0.02$ ℧ conductance circle (point B), as shown in Figure 5.13-1.

At this point the cursor indicates a susceptance of $-j0.06$ ℧. In reactance this is $+j16.7$ Ω, which is resonated at 1 GHz using a capacitor of 9.54 pF.

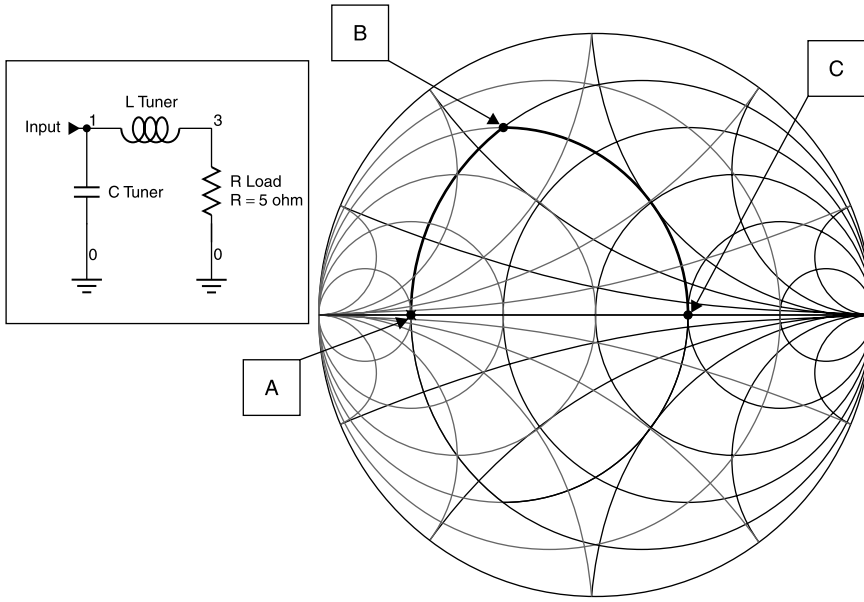


Figure 5.13-1 Smith chart used for LC matching without any transmission line references.

When this value is added, the contour extends to point C, the $R = 50\ \Omega$ value on the 25 Ω reference impedance Smith chart. These tuning elements are the same as those derived earlier using the Q matching procedure (see Figure 3.5-2).

5.14 CONSTANT Q CIRCLES

In a Cartesian-type impedance plane the contours of constant Q are defined by the straight lines $X = QR$, where X is the reactance and R the resistance of a series impedance. We saw that the Smith chart is produced by a bilateral transformation having the property that circles are transformed into circles, straight lines being cases of circles of infinite radius. Thus we expect that constant Q lines drawn in the Smith chart (Γ) plane will be circular or straight lines. This is the case. When the $Q = 3$ circles are drawn in the Smith chart of the previous example, the result is that shown in Figure 5.14-1.

Recall that $Q = 3$ as a condition of the Q matching procedure used in Figure 3.5-2. Given this Q value, the 5- to 50-Ω transformation could be performed on the Smith chart by finding the intersections with the $Q = 3$ circles.

When Q matching is used, consisting of reactances, which alternate in type (L or C) and in topology (series/parallel), adhering to a low Q value ensures broad bandwidth. When viewed on the Smith chart the matching contour

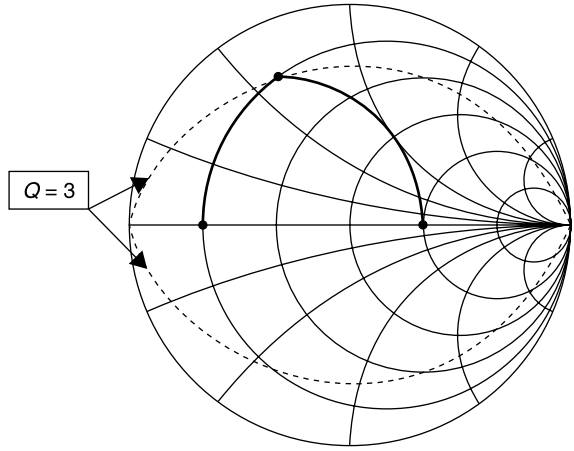


Figure 5.14-1 Smith chart of previous example with $Q = 3$ circles.

remains close to the horizontal $(X = 0)$ axis for low Q values. In the broadband matching example of Figure 3.5-5, the Q was reduced to 1.074. When this example is plotted on the Smith chart, the result is that shown in Figure 5.14-2.

Notice that the Q circles are symmetric with respect to the $x = 0$ axis of the Smith chart. This suggests that the dual of this matching circuit, consisting of a series capacitor connected to the 5-Ω load, followed by a shunt inductor and alternating C and L elements thereafter would produce the same result. This is the case, as shown in Figure 5.14-3.

The element values shown in Figure 5.14-3 were found graphically by extending the contour segment for each tuning element addition to intersect alternately the $Q = 1.074$ circle and the horizontal axis. As such, they are

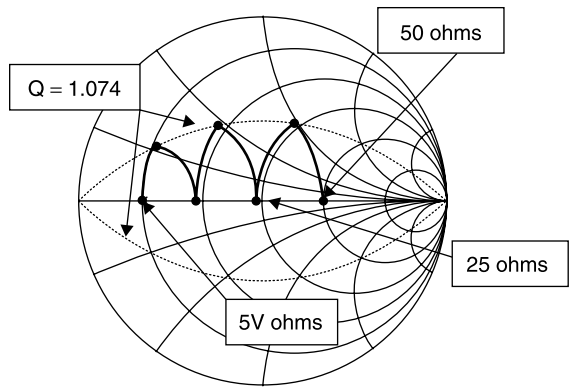


Figure 5.14-2 Broadband LC matching circuit plotted on Smith chart with 25 Ω reference impedance.

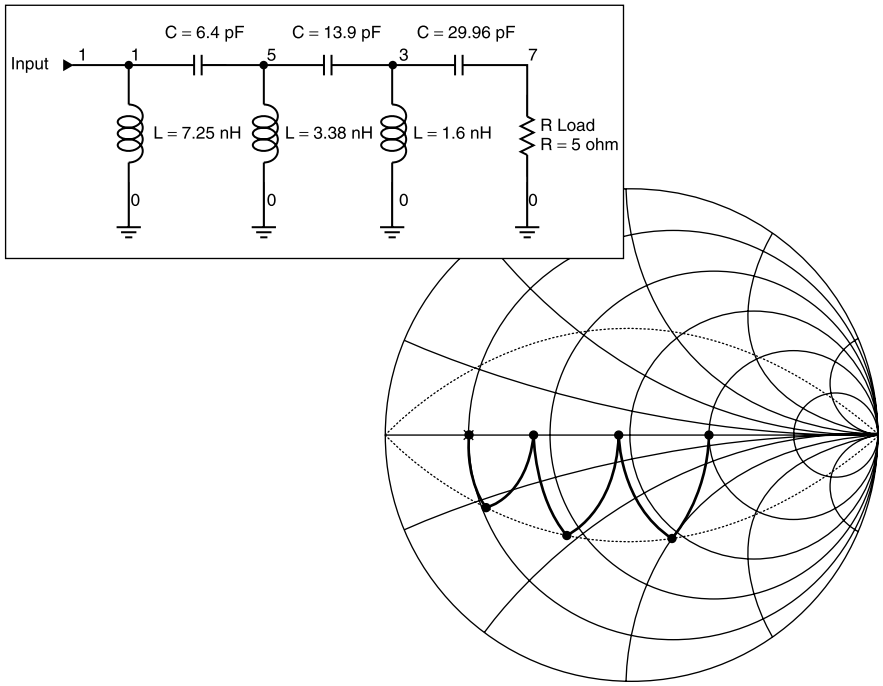


Figure 5.14-3 Dual of broadband tuning circuit derived graphically using lower $Q = 1.074$ circle.

approximate. The final circuit obtained using these approximate values transforms 5 to 49.5 Ω. With more careful graphical techniques, a closer approximation to 50 Ω could be obtained, but the associated VSWR of 1.01 of this result would be sufficient for practically all applications.

5.15 TRANSMISSION LINE LUMPED CIRCUIT EQUIVALENT

Often the properties of a uniform transmission line section are desired but space does not permit the use of an actual section of line. An example is the realization of the branch line hybrid coupler (covered in Section 8.9) using lumped elements. In such cases a lumped equivalent circuit for a transmission line section is useful. The equivalent circuit is only equivalent at one frequency, but the approximation over a modest bandwidth is adequate for many applications.

As an example, suppose that we wish to simulate a 90° length of 50-Ω line using an *LCL* tee circuit at 1 GHz (Fig. 5.15-1). What are the element values for the equivalent circuit?

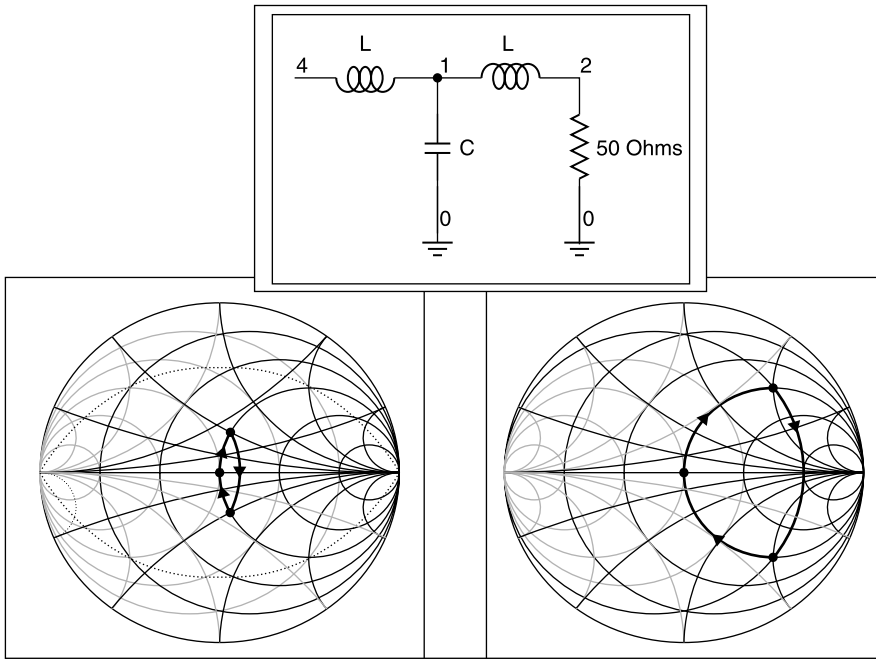


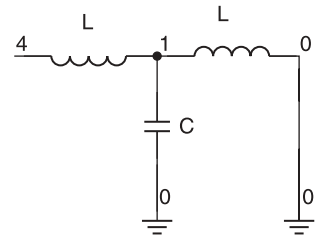
Figure 5.15-1 Combinations 1 and 5 yielding 50-Ω input for *LCL* tee circuit.

We recognize that any length of 50-Ω transmission line presents an input impedance of 50 Ω when the line is terminated in 50 Ω. After some experimentation on the Smith chart, we find that, starting with a load of 50 Ω (and normalizing the chart to this value), there are any number of L and C reactance combinations that will yield 50 Ω input impedance. A set of five combinations is shown in Table 5.15-1.

It might seem that this Smith chart application has failed since it does not provide a unique circuit solution, but this is not so. We have only required that the circuit deliver 50 Ω to the input when the load is 50 Ω, but *any 50-Ω line length does so*; and the Smith chart results properly indicate that numerous equivalent circuits exist for line sections of various lengths. To determine the

TABLE 5.15-1 Sets of L and C Values Yielding 50 Ω Input to Tee Circuit

Combination	1	2	3	4	5
X_L	25	35	50	75	100
X_C	−63	−53	−50	−54.1	−62.6
L	3.98 nH	5.57 nH	7.96 nH	11.94 nH	15.92 nH
C	2.52 pF	3.0 pF	3.18 pF	2.94 pF	2.54 pF



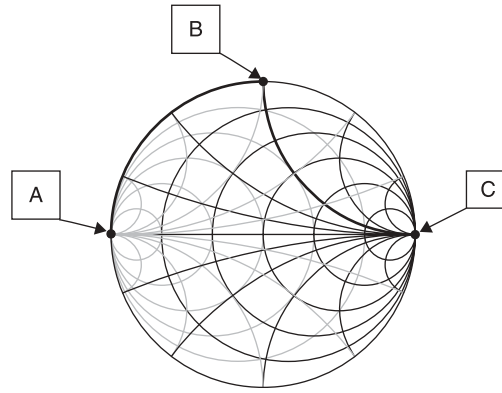


Figure 5.15-2 Smith chart plot of tee circuit with short-circuit load.

equivalent circuit specifically for a 90° long, 50-Ω line section, we must add another condition that relates input and output impedances unique to a 90° long, 50-Ω line section. One such combination is a short-circuit load, which is transformed to an open circuit by a 90° line section as indicated by (4.14-7a). Among the test cases of Table 5.15-1, we find that case 3 yields this result, as shown in Figure 5.15-2.

Initially, we locate the short circuit at point A. Adding a series reactance of $+50\ \Omega$ takes the contour to point B. Note that this is the $z = 0 + j1$ point on the chart normalized to 50 Ω. Next we must use admittance since the capacitor of the tee circuit is a shunt element. However, the contour remains on the periphery of the Smith chart because both the $r = 0$ and the $g = 0$ circles lie on the $|\Gamma| = 1$ circle. To reach the horizontal axis we must add a normalized capacitive susceptance of $+1$. Notice that this brings us to the infinite impedance, point C, on the Smith chart, regardless of the value of the remaining series inductor. However, this second inductor is necessary since only the cases listed in Table 5.15-1 give a matched 50-Ω input when the circuit is loaded with *50* Ω. Given this result, it is now clear why the capacitive and inductive reactances must be equal; they must be *parallel resonant* in order that the input

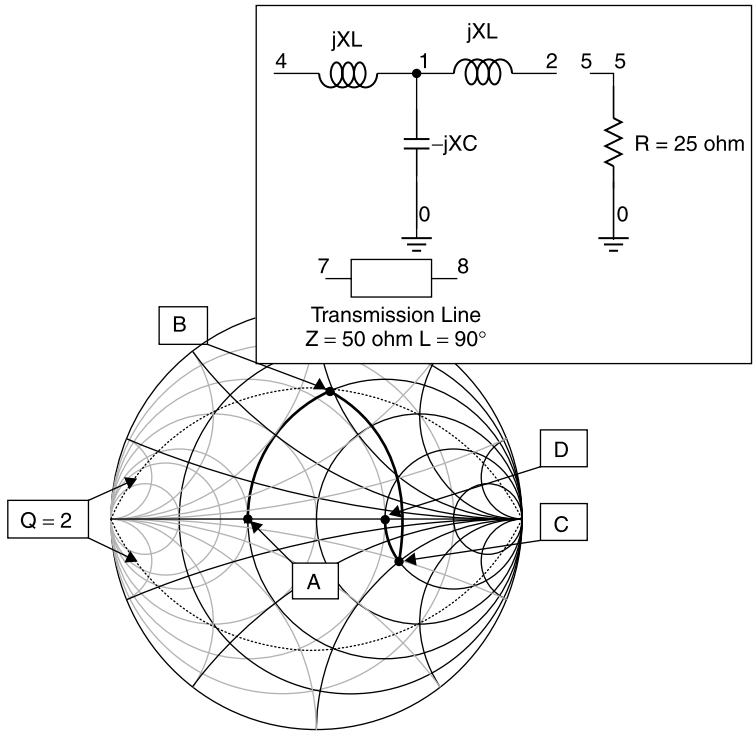


Figure 5.15-3 The circuit and Smith chart contour of the three-element, tee section equivalent circuit of a 90° long, 50-Ω transmission line at 1 GHz.

will be an open circuit when the output of the tee circuit is short circuited. While this is now obvious, the Smith chart revealed this fact graphically.

As a test of this equivalent circuit, consider another pair of load and input impedances. We noted earlier that a quarter-wave transmission line behaves as an impedance inverter. For example, when terminated in *25* Ω, the 90° section of a *50*-Ω line presents *100* Ω at its input terminals. The tee circuit must provide the same transformation if it is to be equivalent.

We begin at point *A* (Fig. 5.15-3), which represents the 25-Ω load on a Smith chart normalized to 50 Ω. Adding a series inductor follows a constant resistance circle to point *B*. Adding a shunt capacitor follows a constant conductance circle to point *C*. Then adding a series inductor of the same value as used between points *A* and *B* follows a constant resistance contour to point *D*, the 100 Ω required input resistance value.

This three-element tee equivalent of a transmission line section can be analyzed exactly using the *ABCD* matrix approach of the next chapter. However, the matrix solution is more difficult to apply for a five-element equivalent circuit, as might be employed to achieve a broader bandwidth of operation.

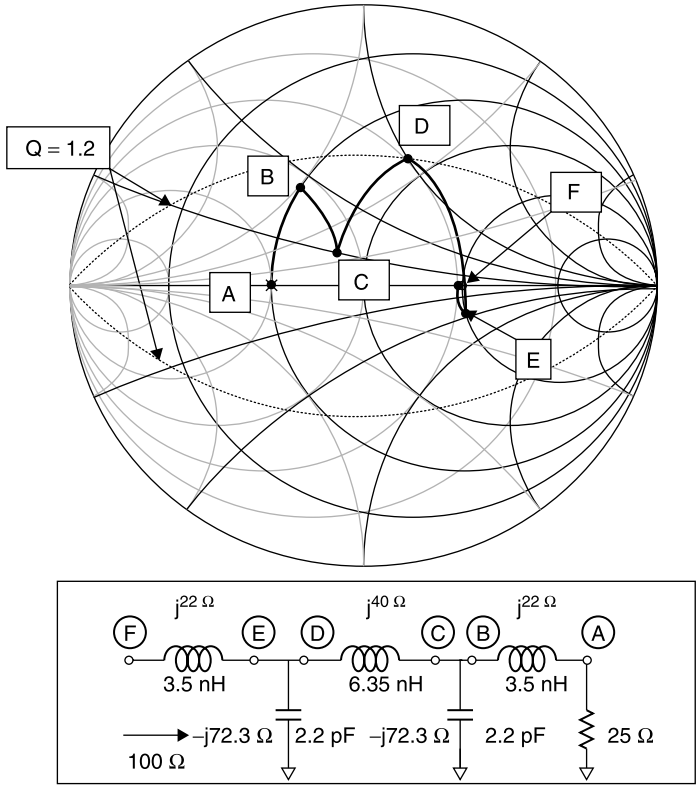


Figure 5.15-4 Smith chart determination of symmetric five-element circuit that is equivalent to a 90°, 50-Ω transmission line at 1 GHz.

The five-element model was designed using the Smith chart, maintaining circuit symmetry (end two elements are equal to each other, as are the second and fourth elements) and requiring that a *25*-Ω load result in a *100*-Ω input. After another cut-and-try process on the Smith chart, the symmetric five-element circuit and its contour shown in Figure 5.15-4 was obtained. The results are approximate, but as will be seen, the match is quite close to 1.00 at the center frequency.

Notice that the contour is just inside a $Q = 1.2$ circle drawn on the chart, while the contour for the three-element circuit is just inside the $Q = 2$ circle. Due to the lower Q of the five-element circuit we would expect broader bandwidth for it, and this is the case as can be seen in Figure 5.15-5. For these simulations the input VSWR is referenced to 100 Ω in order to indicate the deviation of Z_{IN} from the required 100-Ω value.

These examples show that the Smith chart can be employed to design matching networks having considerable complexity, which might otherwise make them more difficult to design by purely analytical means. The examples

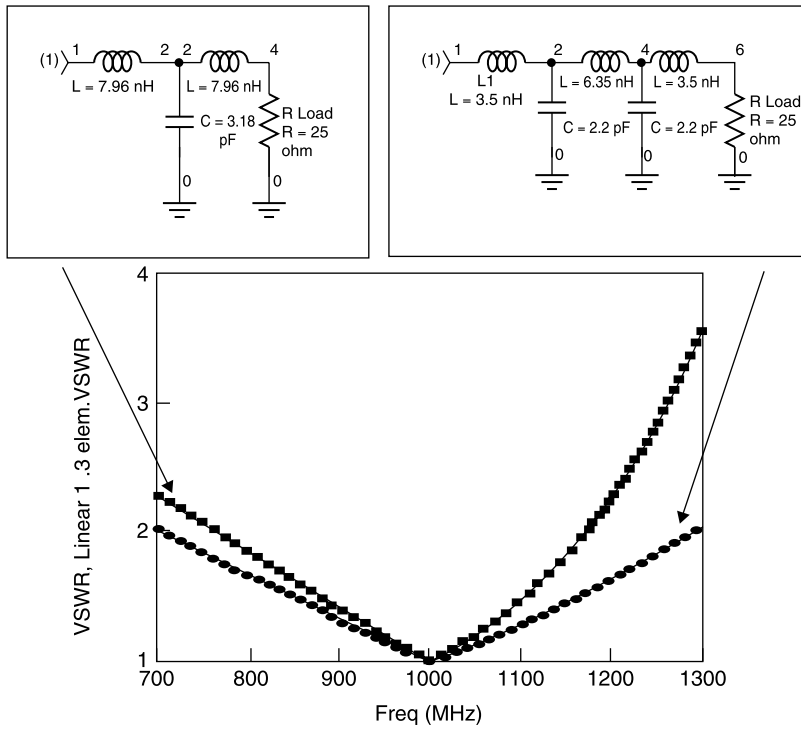


Figure 5.15-5 Performance comparison of three-element and five-element transmission line equivalent circuits designed graphically using the Smith chart.

also are intended to demonstrate that what can be obtained using the Smith chart is only limited by the imagination and skill of its user.

REFERENCES

1. Phillip H. Smith, "A transmission line calculator," *Electronics*, Vol. 12, No. 1, 1939, p. 29. *Smith's original paper describing what became known as the "Smith chart."*
2. Phillip H. Smith, "An improved transmission line calculator," *Electronics*, Vol. 17, No. 1, 1944, p. 130. *Refinements to Smith's original paper*.
3. Phillip H. Smith, *Electronic Applications of the Smith chart*, Noble Publishing, Norcross, GA, 1995. *This is Smith's original book on the Smith chart, first published by McGraw-Hill in 1969 with the assistance of Kay Electric.*
4. Kenneth S. Miller, *Advanced Complex Calculus*, Harper & Brothers, New York, 1960. *An old reference, much too advanced for our purposes, but its development of the properties of the bilinear transformation is fundamental to the realizability of the Smith chart.*

5. Chris Bowick, *RF Circuit Design*, Butterworth-Heinemann, Woburn, MA, 1982. *This is a very readable and practical design guide to filter design and matching circuits.*
6. *winSmith 2.0*, Noble Publishing, Norcross, GA, 1995, 1998. *A personal computer software program for performing Smith chart computations. Mention this textbook for a courtesy discount.*

EXERCISES

E5.1-1 Derive an analytic expression for the input admittance Y_{IN} to a lossless, Z_0 transmission line that is θ degrees long and terminated in an admittance Y_L.

E5.3-1 Derive in terms of Y_0 and electrical length, θ, expressions for the susceptance of (a) open-circuited lossless transmission lines and (b) short-circuited lossless lines.

E5.3-2 A 50-Ω transmission line has a matched generator with available average power of 100 W. What is the voltage 45° from the load if the load is $(50 + j50)$ Ω?

E5.4-1 A *50*-Ω microstrip transmission line has a load of $(50 - j50)$ Ω that is to be matched using a shunt *50*-Ω, open-circuit terminated stub. Since this method of tuning can be "printed" on the board and requires no "via" to connect it to ground, it is expected to be the most economical, reproducible, and vibration-resistant tuning method. Assume the load impedance is constant over the bandwidth.

a. Use the Smith chart to determine the location and electrical length of the stub needed for the matching.

b. Use the Smith chart to determine the VSWR at the edges of a $\pm 10\%$ bandwidth.

E5.4-2 Can you find a 50-Ω perfectly matched input solution to the problem in E5.4-1 using a properly chosen Z_0 and electrical length? If so, what are the values?

E5.4-3 A transistor has an input, which consists of 25 Ω shunted by 6.4 pF. Use the Smith chart to design a matching network to 50 Ω at 1 GHz using only a 50-Ω transmission line and an open-circuited 50-Ω shunt stub.

E5.4-4 Can you find a way to tune the amplifier input of E5.4-2 using only transmission lines that have a lower VSWR over the 900 to 1100 MHz bandwidth?

E5.6-1 What are the maximum and minimum resistances that will be seen along a 50-Ω transmission line having a length of at least one-half wavelength and a load VSWR of 3?

E5.11-1 A patch antenna at 850 MHz is measured to have an impedance of $(5 - j25)\ \Omega$. Use the Smith chart to design a matching network to 50 Ω as well as you can (using an approximate method is acceptable) *using two in-line (cascaded) transmission line sections* of appropriate characteristic impedances.

E5.11-2 Repeat Exercise 5.11-1, this time using lumped elements and the Q matching method.

E5.11-3 Repeat Exercise 5.11-2, this time using only a cascade section of a 50-Ω line and an open-circuited, shunt 50-Ω line stub. Use the Smith chart to determine the length of the shunt stub and its location on the main line.

E5.11-4 Use a network simulator to compare the bandwidths for −20-dB return loss for the solutions that you obtained for Exercises 5.11-1, 5.11-2, and 5.11-3.

E5.11-5 **a.** Can you find a way to tune the load in E5.4-1 using an approximate method that would satisfy the requirement for an all printed matching network and give a VSWR over the band no higher than that achieved in the solution of E5.4-1.

b. Explain why this matching method gives lower VSWR over the operating bandwidth.

E5.14-1 Only two reactive elements are needed to conjugately match any complex impedance load to any complex impedance generator. Demonstrate that there is an unlimited number of matching networks possible by using 6 to 10 reactive elements (L's and C's) to transform 10 to 50 Ω. Use a "random walk" on the Smith chart that starts at 10 Ω and eventually, within 6 to 10 elements, "arrives" at 50 Ω.

E5.14-2 **a.** Estimate the Q of a single LC matching network needed to transform 7 to 50 Ω at 1000 MHz.

b. Draw Q circles on the Smith chart and use them as a guide to design a matching network consisting of a series L and a shunt C.

c. Change the reactances of the two tuning elements to what they become at 900 and 1100 MHz. Plot their respective contours on the Smith chart. They will not quite transform to 50 Ω but to some other impedances. What are the VSWR values at these impedances (at therefore at 900 and 1100 MHz)?

E5.14-3 Repeat E5.14-2, but this time use a three-section (six-element) tuner. Does this give a better match (lower VSWR) over the 900 to 1100 MHz bandwidth?

CHAPTER 6

Matrix Analysis

6.1 MATRIX ALGEBRA

For our purposes, *a matrix is an ordered array of the coefficients of a set of linear simultaneous equations*. For example, a linear equation in two variables, x and y, describes a line in the xy plane. Two such equations describe two lines. If the lines are not parallel and one does lie on top of the other, they will intersect. The intersection corresponds to a common solution for x and y that satisfies both equations:

$$a_{11}x + a_{12}y = a_{13} \tag{6.1-1a}$$

$$a_{21}x + a_{22}y = a_{23} \tag{6.1-1b}$$

In matrix notation these equations are written

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} \tag{6.1-2}$$

The rules for matrix manipulation allow us to recover from (6.1-2) the same information contained in (6.1-1a,b). *To multiply an $n \times n$ square matrix A by an $n \times 1$ column matrix B, form the resultant column matrix C whose respective elements j are equal to the sum of the products of the jth row of A by the corresponding column elements of B.*

Explicitly this rule is demonstrated for a 2×2 matrix multiplying a 2×1 matrix as

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_{11}b_1 + a_{12}b_2 \\ a_{21}b_1 + a_{22}b_2 \end{pmatrix} = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} \tag{6.1-3}$$

To multiply a square matrix by another square matrix the rule is: *Multiplying an $n \times n$ matrix by another $n \times n$ matrix produces an $n \times n$ product matrix whose*

High Frequency Techniques: An Introduction to RF and Microwave Engineering, By Joseph F. White.
ISBN 0-471-45591-1

ij terms are the sum of the products obtained by multiplying each element in the ith row of the first matrix by the corresponding elements in the jth column of the second matrix.

For example, multiplying a 2×2 by another 2×2 matrix yields

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} (a_{11}b_{11} + a_{12}b_{21}) & (a_{11}b_{12} + a_{12}b_{22}) \\ (a_{21}b_{11} + a_{22}b_{21}) & (a_{21}b_{12} + a_{22}b_{22}) \end{pmatrix} \quad (6.1\text{-}4)$$

We also define that *two matrices are equal to each other if every element of the first is equal to the corresponding element of the second.*

Given these rules, (6.1-3) is equivalent to (6.1-1a,b) where in this case $b_1 = x$ and $b_2 = y$. Some additional rules are: *To add two matrices together, form the sum matrix whose elements are the sum of the respective elements of the two matrices*. Thus

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} (a_{11} + b_{11}) & (a_{12} + b_{12}) \\ (a_{21} + b_{21}) & (a_{22} + b_{22}) \end{pmatrix} \quad (6.1\text{-}5)$$

Similarly, *to subtract a matrix B from matrix A, form the resulting matrix C whose elements are* $c_{ij} = a_{ij} - b_{ij}$ *for all i and j*. Thus

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} - \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} (a_{11} - b_{11}) & (a_{12} - b_{12}) \\ (a_{21} - b_{21}) & (a_{22} - b_{22}) \end{pmatrix} \quad (6.1\text{-}6)$$

If two simultaneous equations have a common solution, we could find it algebraically by solving one equation in terms of one unknown and then substituting this value into the other equation to solve for the remaining unknown. Similar manipulations would allow the simultaneous solution of *n*-independent equations in *n* unknowns.

However, the use of matrix notation and its rules allows us to formalize this solution to a simple procedure called *Cramer's rule* [1]. Before describing the rule we must first define the value of the determinant, (pronounced "del").

For a 2×2 matrix, the value of del is obtained by cross multiplying terms beginning with the first product at the upper left and subtracting the second cross product as shown below:

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \quad (6.1\text{-}7)$$

If the simultaneous equations have a common solution,

$$\Delta \neq 0$$

Given the nonzero value for Δ, Cramer's rule can be applied to the solution of

(6.1-1a,b) as

$$x = \frac{\begin{vmatrix} a_{13} & a_{12} \\ a_{23} & a_{22} \end{vmatrix}}{\Delta} \tag{6.1-8a}$$

$$y = \frac{\begin{vmatrix} a_{12} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}}{\Delta} \tag{6.1-8b}$$

in which the determinants in the numerators of x and y are evaluated in the same manner described for the del in (6.1-6).

For matrices larger than 2×2, we must use a reduction process, successively reducing the size of the starting matrix by evaluating *minors* and *cofactors. The minor of an element with subscript ij in a determinant is the determinant that results when the ith row and the jth column are deleted.*

For example, the minor of the a_{11} element of the 3×3 matrix is

$$\text{Minor of } a_{11} \text{ in } \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \tag{6.1-9}$$

The *cofactor* is the minor with the appropriate algebraic sign. *The determinant is equal to the sum of the products of the elements of any row or column by their respective cofactors, where the cofactor of the element in the ith row and the jth column is* $(-1)^{i+j}$ *times its minor.*

Thus, the evaluation of the determinant of a 3×3 matrix first proceeds by reducing the value of del to three terms, each containing a 2×2 matrix, whose numeric value is obtained by forming the diagonal products with appropriate signs, as shown below:

$$\begin{aligned} \Delta &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{33}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}) \end{aligned} \tag{6.1-10}$$

Notice that the negative sign of the second term results because the sum of the subscripts of a_{12} is odd, hence the negative sign.

With this basis, Cramer's rule can be stated formally as: *Given a nonzero* $n \times n$ *determinant* Δ *of the coefficients of a set of n simultaneous linear equations, the common solution for the ith unknown is equal to the fraction of two determinants having a denominator equal to* Δ *and a numerator determinant obtained by replacing the ith column of* Δ *with the* $a_{i,n+1}$ *values of the equation set.*

For example, consider a two-equation set of the form

$$a_{11}x + a_{12}y = a_{13} \tag{6.1-11a}$$

$$a_{21}x + a_{22}y = a_{23} \tag{6.1-11b}$$

with specific coefficients

$$x - y = 0 \tag{6.1-12a}$$

$$2x + y = 6 \tag{6.1-12b}$$

First, note that the determinant for the equation set is not equal to zero:

$$\Delta = 1 - (-2) = 3 \tag{6.1-13}$$

Therefore the equations have a common, nontrivial solution.

Next, using Cramer's rule, the simultaneous solution of the two equations is obtained as

$$x = \frac{\begin{vmatrix} 0 & -1 \\ 6 & 1 \end{vmatrix}}{\Delta} = \frac{1 \cdot 0 - (-1) \cdot 6}{3} = 2 \tag{6.1-14}$$

$$y = \frac{\begin{vmatrix} 1 & 0 \\ 2 & 6 \end{vmatrix}}{\Delta} = \frac{1 \cdot 6 - 2 \cdot 0}{3} = 2 \tag{6.1-15}$$

The reader can verify that the solution $x = 2,\ y = 2$ satisfies both equations, and that graphing the straight lines that they represent reveals that their common solution (intersection) is at $(2, 2)$ in the xy plane.

While the rules given for the evaluation of determinants and the application of Cramer's rule apply for any size determinant, most of our applications can be accommodated with no more than a 3×3 determinant.

6.2 *Z* AND *Y* MATRICES

A two-port network has two ports, each having a pair of terminals. *The behavior of a linear two-port network can be described entirely at one frequency by a pair of equations, any of whose terms may be complex*. One of several such equation descriptions employs the *Z matrix* or *impedance parameters* (Fig. 6.2-1).

The equations that define the circuit's behavior are

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \tag{6.2-1}$$

$$V_2 = Z_{21}I_1 + Z_{22}I_2 \tag{6.2-2}$$

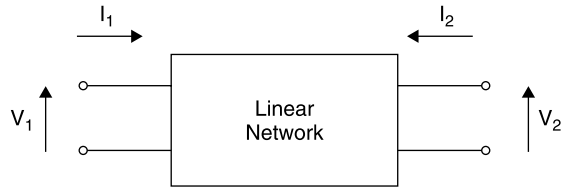


Figure 6.2-1 Voltage and current conventions for Z and Y matrices.

where V and I are phasor quantities. The Z parameter values are found by solving circuit equations for the various ratios of network voltages and currents. Thus,

$$Z_{11} = \frac{V_1}{I_1}\bigg|_{I_2=0} \qquad Z_{12} = \frac{V_1}{I_2}\bigg|_{I_1=0} \qquad Z_{21} = \frac{V_2}{I_1}\bigg|_{I_2=0} \qquad Z_{22} = \frac{V_2}{I_2}\bigg|_{I_1=0} \tag{6.2-3}$$

As can be seen from (6.2-3), *to measure the Z parameters requires that open circuits be placed at the input and output terminals* at various times. This may be undesirable to do with microwave networks, but if the Z parameters can be derived analytically, this poses no problem.

Expressed in matrix notation,

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \tag{6.2-4}$$

The definition of the Z matrix can be applied to a network having an arbitrary number of ports, the requirement being a separate independent equation relating the voltage at every port to the currents at all of the ports. Thus, an N-port network requires an $N \times N$ matrix representation.

One can also write an admittance matrix based on Kirchhoff's node equations. The nodal format of the Y matrix is the basis of the *nodal analysis* format used extensively in network analysis and simulation software, such as the *Genesys* program employed for the examples in this text.

The *admittance parameters*, or *Y matrix*, use the same voltage and current conventions (Fig. 6.1-1) that are employed to define the Z matrix:

$$\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \tag{6.2-5}$$

and the elements are defined in a manner similar to that used to define the Z matrix elements, that is

$$Y_{11} = \frac{I_1}{V_1}\bigg|_{V_2=0} \qquad Y_{12} = \frac{I_1}{V_2}\bigg|_{V_1=0} \qquad Y_{21} = \frac{I_2}{V_1}\bigg|_{V_2=0} \qquad Y_{22} = \frac{I_2}{V_2}\bigg|_{V_1=0} \tag{6.2-6}$$

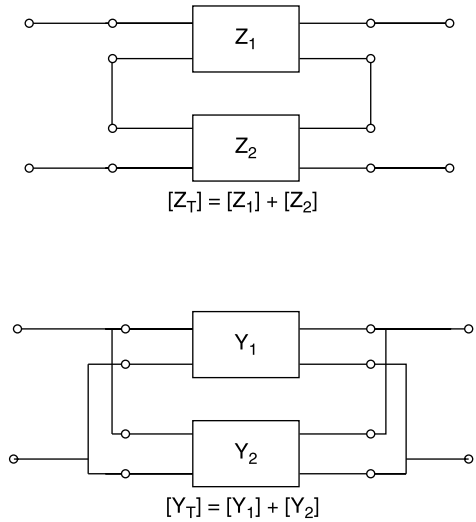


Figure 6.2-2 Series and parallel network combinations represented by Z and Y matrix additions [5, p. 190].

Notice that the Y parameters are the *dual* of the Z parameters in every manner. Impedance is replaced by admittance, currents and voltages are interchanged, and the elements are evaluated using a zero voltage at a terminal (a short circuit) rather than the zero current (open circuit) used to evaluate the Z matrix elements.

The impedance matrix, and its dual, the admittance matrix, have the interesting and very useful properties that the overall matrix of a series or parallel combination of circuits can be obtained by adding together the corresponding elements, respectively, of the Z or Y matrices. This is shown schematically below for pairs of two ports, but the result can be applied to arbitrarily extensive series hookups for Z parameter circuits or parallel hookups for Y parameter circuits.

The relationships in Figure 6.2-2 demonstrate the value of using matrix notation. It facilitates the recognition of identities that might not be so evident were only the underlying simultaneous equations employed.

6.3 RECIPROCITY

A network is reciprocal if a zero impedance source and a zero impedance ammeter can be placed at any locations in a network and their positions interchanged without changing the ammeter reading [3, p. 544]. This is a general relationship attributed to Lorentz and is derived in the general case by applying Maxwell's equations to a general isotropic medium [4, Sec. 1.9]. The conditions for the

evaluation of the Z_{ij} and Z_{ji} matrix elements is equivalent to the Lorentz experiment of interchanging generator and ammeter. All of the circuits in Figure 6.4-3 are reciprocal [3, p. 544].

Accordingly, as a consequence of reciprocity, for the Z matrix description of a reciprocal network

$$Z_{12} = Z_{21} \tag{6.3-1}$$

and for the Y matrix,

$$Y_{12} = Y_{21} \tag{6.3-2}$$

Basically, reciprocity results with network elements that are linear and bilateral, that is, they have the same behavior for currents flowing in either direction. Invoking reciprocity is a powerful tool that provides an additional relationship useful in the analysis of circuits and electromagnetic field cases. It is customarily applied to show that characteristics of transmitting antennas apply equally to receiving antennas and vice versa.

6.4 THE *ABCD* MATRIX

The *ABCD parameters, or ABCD matrix*, are similar to the Z and Y parameters but with three important differences [5, pp. 187–188; 2, Sec. 11.08]. First, *the ABCD matrix is defined only for a two-port network*. Second, the output current direction exits the network. Third, the dependent and independent variables are mixed, each containing both a voltage and a current (Fig. 6.4-1). The result is that the outputs of the network, V_2 and I_2, have the same directions and definitions as the inputs of a following network. For this reason *ABCD matrices can be multiplied together to form the overall matrix of a cascade of circuits*. Because of this property the *ABCD* matrix is also called the *chain matrix*.

The defining equations for the *ABCD* parameters are

$$V_1 = AV_2 + BI_2 \tag{6.4-1}$$

$$I_1 = CV_2 + DI_2 \tag{6.4-2}$$

in which the A, B, C, and D parameters are evaluated using the relationships in

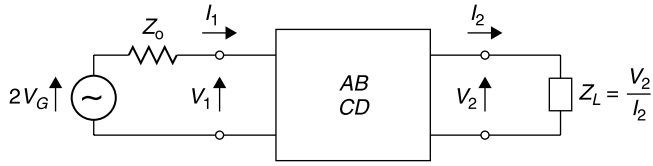


Figure 6.4-1 Defining voltage and current conventions used for the *ABCD* matrix.

(6.4-3) and noting that $I_2 = 0$ corresponds to an open-circuit termination and $V_2 = 0$ corresponds to a short-circuit termination:

$$A = \frac{V_1}{V_2}\bigg|_{I_2=0} \qquad B = \frac{V_1}{I_2}\bigg|_{V_2=0} \qquad C = \frac{I_1}{V_2}\bigg|_{I_2=0} \qquad D = \frac{I_1}{I_2}\bigg|_{V_2=0} \tag{6.4-3}$$

As is true of the elements of the Z and Y matrices, the A, B, C, and D coefficients may be complex quantities. The $ABCD$ matrices for commonly encountered two-port networks are listed in Figure 6.4-2 and can be verified using (6.4-3).

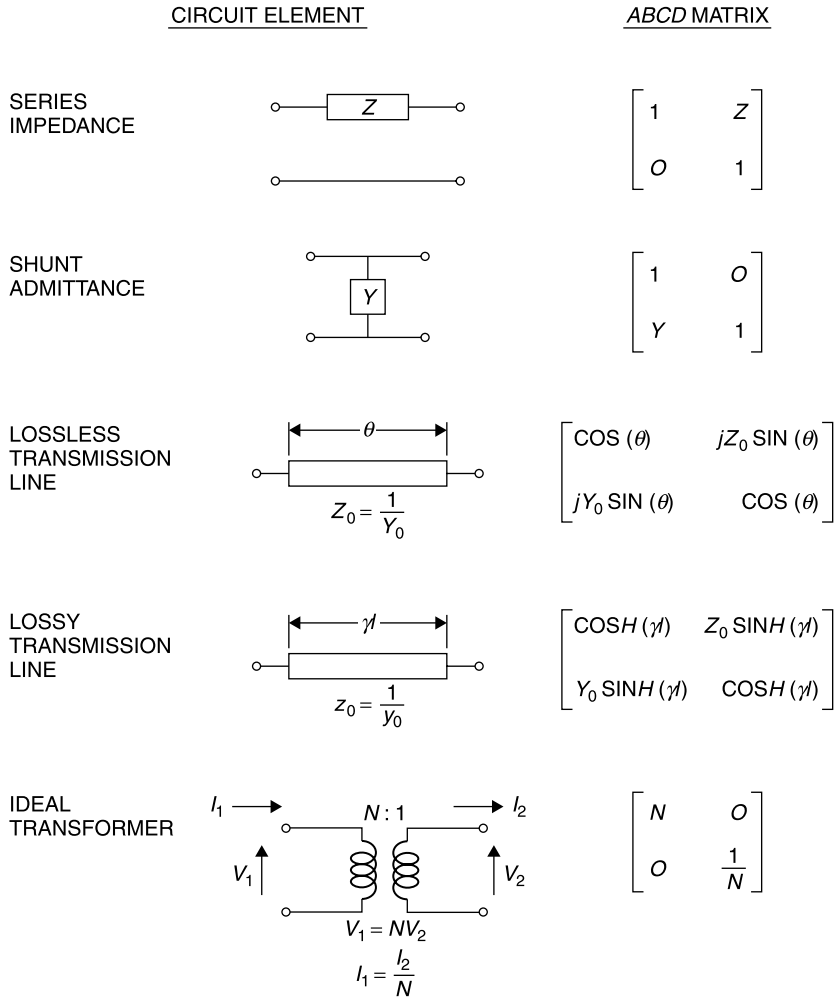


Figure 6.4-2 $ABCD$ matrices for common circuit elements.

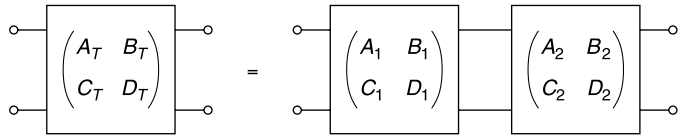


Figure 6.4-3 Cascaded circuits can be represented by a new circuit whose *ABCD* parameters are the result of matrix multiplying the *ABCD*s of the individual circuits.

Remarkably, the modest library of elements and their *ABCD* matrices shown in Figure 6.4-2 suffices to permit the calculation of numerous practical RF and microwave cascaded networks. The principal advantage of the *ABCD* parameters is the fact that the overall matrix representation of a cascade of circuits, $ABCD_T$, can be formed by matrix multiplying their individual *ABCD* matrices. This is shown schematically in Figure 6.4-3.

The required matrix multiplication is

$$\begin{pmatrix} A_T & B_T \\ C_T & D_T \end{pmatrix} = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix} \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} = \begin{pmatrix} (A_1A_2 + B_1C_2) & (A_1B_2 + B_1D_2) \\ (C_1A_2 + D_1C_2) & (C_1B_2 + D_1D_2) \end{pmatrix} \tag{6.4-4}$$

The matrix multiplication defined by (6.4-4) is not *commutative*. That is, $[ABCD]_1 \times [ABCD]_2$ is not necessarily equal to $[ABCD]_2 \times [ABCD]_1$. Therefore, care must be exercised to preserve the order of the matrix multiplications of the cascaded circuits shown in Figure 6.4-3 and defined by (6.4-4). When a cascade of more than two networks is encountered, *a good procedure to find the overall ABCD matrix is to begin by combining the two networks nearest to the load*. Then that result is combined with the next nearest network to the load. The process can be repeated until the *ABCD* matrix for an arbitrarily long cascade is obtained.

By dividing (6.4-1) by (6.4-2) we obtain an expression for the input impedance, Z_{IN}, to the network in terms of the load impedance, Z_L, as follows:

$$Z_{\text{IN}} = \frac{AV_2 + BI_2}{CV_2 + DI_2} = \frac{A(V_2/I_2) + B}{C(V_2/I_2) + D} = \frac{AZ_L + B}{CZ_L + D} \tag{6.4-5}$$

The insertion loss and insertion phase of a two-port network are readily determined from its *ABCD* matrix. Since these values are referenced to the voltage of the generator, V_G, rather than V_1, it is useful to find the total matrix, $ABCD_T$, of the network that comprises both the two port as well as the impedance of the generator, Z_G. This is diagrammed in Figure 6.4-3. This overall

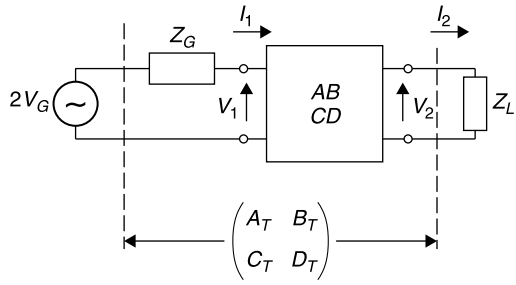


Figure 6.4-4 The *ABCD* matrix of a two-port and the generator impedance is used to compute insertion loss and insertion phase.

matrix is shown in Figure 6.4-4 and is calculated as follows:

$$\begin{pmatrix} A_T & B_T \\ C_T & D_T \end{pmatrix} = \begin{pmatrix} 1 & Z_G \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} (A + CZ_G) & (B + DZ_G) \\ C & D \end{pmatrix} \quad (6.4\text{-}6)$$

Applying (6.4-1),

$$2V_G = A_T V_2 + B_T I_2 = A_T V_2 + \frac{B_T V_2}{Z_L}$$

$$\frac{2V_G}{V_2} = A_T + \frac{B_T}{Z_L}$$

$$\frac{2V_G}{V_2} = A + CZ_G + \frac{B}{Z_L} + \frac{DZ_G}{Z_L}$$

This expression can be used to find the insertion loss and insertion phase resulting when a network described by the matrix, *ABCD*, is inserted between a generator, Z_G, and load Z_L. When $Z_G = Z_L = Z_0 =$ a real quantity, and since the maximum power deliverable to the load is

$$P_A = V_G^2 / Z_0$$

then the transducer loss and insertion loss (IL) of a two-port network defined by its *ABCD* matrix are the same and equal to [2, p. 187]

$$\text{IL} = \left|\frac{V_G}{V_2}\right|^2 = \frac{1}{4}\left|A + \frac{B}{Z_0} + CZ_0 + D\right|^2 \qquad (Z_L = Z_0 = \text{real}) \quad (6.4\text{-}7)$$

Under these conditions the initial phases of V_G and V_L (more precisely the respective arguments of their phasor representations) are the same and can be defined as the reference phase, 0°. The change in the argument of the voltage at

the load incurred by inserting the *ABCD* network is that of the ratio of V_G/V_2. It is called the *insertion phase*, φ_I, of the *ABCD* network and given by

$$\varphi_I = \tan^{-1}\left\{\frac{\mathrm{Im}(A + B/Z_0 + CZ_0 + D)}{\mathrm{Re}(A + B/Z_0 + CZ_0 + D)}\right\} \tag{6.4-8}$$

Note that a positive value of φ_I indicates a phase delay through the two-port network while a negative value indicates a phase advance.

All of the circuits shown in Figure 6.4-2 are reciprocal. If the network is passive and reciprocal [3, p. 534],

$$AD - BC = 1 \tag{6.4-9}$$

and if the network is symmetrical with respect to its input and output ports [3, p. 534]

$$A = D \tag{6.4-10}$$

As a further example in the use of the *ABCD* matrix, suppose that an integrated circuit requires a quarter wavelength transmission line. If there is insufficient space for the line itself, its behavior can be *mimicked at one frequency* by an equivalent lumped element Pi circuit containing two capacitors and an inductor (Fig. 6.4-5).

The series impedance jX_L of the inductance L is $j\omega L$ and the shunt admittance jB_C of each capacitance C is $j\omega C$. To find the overall matrix for this circuit, $ABCD_{\mathrm{Pi}}$ find the respective *ABCD* matrices for the elements (Fig. 6.4-2) and perform successive multiplication of them *beginning at the load end* [5, p. 189]:

$$\begin{aligned}
\begin{pmatrix} A_{\mathrm{Pi}} & B_{\mathrm{Pi}} \\ C_{\mathrm{Pi}} & D_{\mathrm{Pi}} \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ jB_C & 1 \end{pmatrix}\left(\begin{pmatrix} 1 & jX_L \\ 0 & 1 \end{pmatrix}\left(\begin{pmatrix} 1 & 0 \\ jB_C & 1 \end{pmatrix}\right.\right. \\
\begin{pmatrix} A_{\mathrm{Pi}} & B_{\mathrm{Pi}} \\ C_{\mathrm{Pi}} & D_{\mathrm{Pi}} \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ jB_C & 1 \end{pmatrix}\left(\begin{pmatrix} (1 - B_C X_L) & jX_L \\ jB_C & 1 \end{pmatrix}\right. \\
\begin{pmatrix} A_{\mathrm{Pi}} & B_{\mathrm{Pi}} \\ C_{\mathrm{Pi}} & D_{\mathrm{Pi}} \end{pmatrix} &= \begin{pmatrix} (1 - B_C X_L) & jX_L \\ jB_C(2 - B_C X_L) & (1 - B_C X_L) \end{pmatrix}\left(\right.
\end{aligned} \tag{6.4-11}$$

Comparing $ABCD_{\mathrm{Pi}}$ matrix with that of a lossless transmission line of Z_0 im-

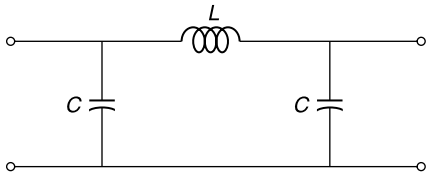


Figure 6.4-5 Pi circuit used to simulate a 90° transmission line.

pedance and θ electrical length (Fig. 6.4-2) equivalence requires that

$$A_{\text{Pi}} = \cos\theta = 1 - B_C X_L \tag{6.4-12}$$

However, for a 90° line section, $\cos\theta = 0$, and therefore

$$X_L = \frac{1}{B_C} \tag{6.4-13}$$

Next solve for B/C for the lossless line length and equate it to the same quotient for the Pi circuit, yielding

$$\frac{B_{\text{Pi}}}{C_{\text{Pi}}} = \frac{Z_0}{Y_0} = Z_0^2 = \frac{X_L}{2B_C - B_C^2 X_L} \tag{6.4-14}$$

and for the case of the 90° line length, for which $X_L = 1/B_C$, (6.4-14) reduces to

$$X_L = \frac{1}{B_C} = Z_0 \tag{6.4-15}$$

For example, a 50-Ω 90° long line section can be simulated as a pair of shunt 50-Ω capacitive reactances and a series 50-Ω inductor. The broad utility of this derivation is that it allows the synthesis of any line length and any impedance. The usefulness of the *ABCD* matrix representation can be appreciated by considering the additional computational labor that would have been required to derive this very general result by other means.

6.5 THE SCATTERING MATRIX

The *Z*, *Y*, and *ABCD* matrices are useful at low frequencies and for proving general analytical results. However, they can be inconvenient to apply to microwave measurements of devices because they require application of short- and open-circuit terminations for evaluation. In addition, it is not practical to measure discrete voltages and currents at high frequencies. Instead, it is relatively easy to measure traveling waves through the use of directional couplers.

For these reasons, the common practice to evaluate the behavior of a multiport network is to use *incident waves* as excitations at each port and to measure the resulting exiting waves [5, p. 541]. The exiting waves are called *reflected waves*. In actuality, the wave exiting a given port may be the result of an incident wave at another port, and therefore more properly might be called a *transmitted wave*, but this is not the practice. The waves that impinge on the network are called *incident waves* and those that exit are called *reflected waves*, regardless of how they come about.

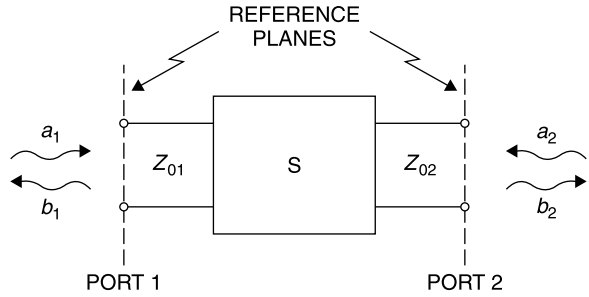


Figure 6.5-1 Defining conventions for the incident, a, and reflected, b, waves.

The incident waves are designated as a_i and the reflected waves are designated as b_i where i is the number of the port. Both the *a and b waves are phasors, having both magnitude and phase at the specified terminals of the network port*. The wave directions and terminal planes are shown schematically in Figure 6.5-1 for a two-port network; however, the convention can be applied to a network with any number of ports, and *the ports need not all be of the same type!* For example, a network might have a coaxial line port, a waveguide port, and a balanced transmission line port.

For each port the incident and reflected waves are defined in terms of the characteristic impedance of the transmission line connected to that port. The wave dimensions are such that when an a or b wave is multiplied by its respective complex conjugate the resulting product has the dimensions of power:

$$a_i = \frac{V_{iI}}{\sqrt{Z_{0i}}} = I_{iI}\sqrt{Z_{0i}} \tag{6.5-1}$$

$$b_i = \frac{V_{iR}}{\sqrt{Z_{0i}}} = I_{iR}\sqrt{Z_{0i}} \tag{6.5-2}$$

The I subscript stands for incident wave, and the R for reflected wave, and Z_{0i} is the characteristic impedance of the transmission line connected to the ith port of the network. Note that a, b, V, and I are phasors and their phases are specified relative to their respective *network ports*.

Often the a and b waves are loosely called *power waves*, but this practice is misleading and should be avoided because *the power associated with each wave is equal to the magnitude of the square of its amplitude*. It is more precise to recognize that they are traveling *waves having amplitudes and phases* just as do incident voltages and currents.

By using a and b waves, a *linear* network can be characterized by a set of simultaneous equations describing the reflected waves from each port in terms of the incident waves at all of the ports. The constants that characterize the network under these conditions are called *S parameters*.

For example, for a two-port network, the b wave leaving port 1 is the phasor sum of a wave reflected from the input port ($S_{11}a_1$) plus a wave that passed through the two-port from port 2 ($S_{12}a_2$):

$$b_1 = S_{11}a_1 + S_{12}a_2 \tag{6.5-3}$$

Similarly

$$b_2 = S_{21}a_1 + S_{22}a_2 \tag{6.5-4}$$

This descriptive approach can be applied to a network with an arbitrary number of ports. An independent equation is needed for each port. The S parameters can be expressed in matrix format, yielding a square, $n \times n$ matrix, where n is the number of network ports. For the two ports the matrix format is

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{6.5-5}$$

Each of the S parameters has its own physical significance. For example, suppose that the S parameters apply to a transistor that is embedded in a matched 50-Ω system. Then S_{11} corresponds to the reflection coefficient at the input port when the output is match terminated ($a_2 = 0$). Similarly, S_{22} is the reflection coefficient at the output port when the input is match terminated ($a_1 = 0$). Parameter S_{12} is the feedback to the input from any wave entering port 2, such as a wave reflection from the load; and S_{21} is the amplification (or loss if $|S_{21}| < 1$) of the input wave amplitude as it travels through the transistor to port 2. Accordingly, $|S_{21}|^2 = S_{21}S_{21}^*$ is the power gain of the transistor (or loss if $|S_{21}| < 1$). These interpretations apply not only to this transistor example but to any linear two-port network to which the S parameters are applied. This separation of the functions of the parameters is especially useful. For example, consider the S parameter file for the 2N6679 transistor shown in Table 6.5-1.

Merely by examining this file, most of the device performance characteristics can be read directly. For example, at 100 MHz the voltage gain $|S_{21}|$ is 38.6. At this frequency the power gain with 50-Ω source and load is $38.6^2 = 1490$, or 32 dB. With 50-Ω source and load, the gain of the transistor drops to unity, 0 dB, at 6.5 GHz. However, with proper matching networks the device can provide more than unity gain at 6.5 GHz, since some of the power is being reflected at the input and output ports.

For example, at 100 MHz, the input reflection coefficient magnitude is 0.6. This means that the input return loss is 36%, which corresponds to a mismatch loss of 1.9 dB. Similarly, the output return loss is 69%, which corresponds to a mismatch loss of 5.1 dB. This means that with proper tuning, the transistor could provide $1.9 + 38.6 + 5.1 = 45.6$ dB total gain at 100 MHz. We also note that the feedback term, S_{12}, is low at 100 MHz, which, as we will see later, has encouraging implications for stability at this frequency.

TABLE 6.5-1 *S* Parameter File for the Motorola 2N667A Bipolar Transistor[a]

! 2N6679A.S2P
! 2N6679
! VCE = 15 V; IC = 25 mA

GHZ S MA R 50

! S-PARAMETER DATA

0.1	0.60	−76	38.6	141	0.01	55	0.83	−20
0.5	0.67	−158	12.7	95	0.02	40	0.50	−27
1	0.68	−178	6.6	77	0.03	53	0.46	−32
1.5	0.68	170	4.4	64	0.04	54	0.47	−41
2	0.69	162	3.4	54	0.05	54	0.47	−50
2.5	0.69	154	2.7	42	0.06	55	0.49	−59
3	0.69	146	1.3	31	0.07	55	0.53	−70
3.5	0.69	138	1.9	21	0.08	54	0.55	−79
4	0.69	131	1.7	11	0.09	51	0.57	−89
4.5	0.69	123	1.5	1	0.10	49	0.59	−97
5	0.69	114	1.4	−9	0.12	44	0.62	−10 6
5.5	0.69	106	1.2	−19	0.14	39	0.64	−11 3
6	0.69	98	1.1	−28	0.15	33	0.68	−12 2
6.5	0.69	90	1.0	−37	0.17	31	0.69	−13 0

[a] By convention the columns contain left to right: f(GHz), $|S_{11}|$, arg(S_{11}), $|S_{21}|$, arg(S_{21}), $|S_{12}|$, arg(S_{12}), $|S_{22}|$, arg(S_{22}) [6].

The appeal of S parameters is apparent in that all of this information is directly manifest from the S parameter data table. For microwave applications, especially transistor characterization, the S parameters are particularly desirable because their evaluation *does not require short- or open-circuit terminations*. Instead *matched loads are placed at the input or output* ports and either the transmission or reflection coefficients is measured. The individual S parameters are evaluated using these relationships:

$$S_{11} = \frac{b_1}{a_1}\bigg|_{a_2=0} \qquad S_{12} = \frac{b_1}{a_2}\bigg|_{a_1=0} \qquad S_{21} = \frac{b_2}{a_1}\bigg|_{a_2=0} \qquad S_{22} = \frac{b_2}{a_2}\bigg|_{a_1=0} \tag{6.5-6}$$

where $a_i = 0$ denotes a matched load at port i.

All excitations, *a waves*, are incident traveling waves, easily measured using directional couplers. Similarly, all exiting waves, *b waves*, are traveling waves in the opposite direction, also readily measured using directional couplers. More about directional couplers and measurements in Chapter 8.

If the network is reciprocal [3, p. 544] (which a transistor is not), then

$$S_{ij} = S_{ji} \tag{6.5-7}$$

The S parameters are complex quantities, having phase angles that correspond to the respective network port reference planes used in their measurement. Yet another convenience of S parameters is that a change of reference planes *farther away from the network by an electrical distance* θ *degrees results in only a negative argument change in the S parameter. For movement closer to the network the argument change is positive*. For diagonal elements, a movement away from the network gives

$$s_{nn}^{\text{new}} = s_{nn}e^{-j2\theta_n} \tag{6.5-8}$$

and for off-diagonal elements a movement away from the network gives

$$s_{mn}^{\text{new}} = s_{mn}e^{-j(\theta_m+\theta_n)} \tag{6.5-9}$$

where the m and n subscripts correspond to the respective m and n reference plane movements. This feature of S parameters allows convenient network *deembedding* in which the S parameters of a network can be easily inferred, even when the measurements were made using reference planes that were remote from the actual network ports as, for example, when a device is measured in a test fixture. That is, the *S parameters are most convenient for measurements, since a change of reference planes is easily calculated by adjusting only the S parameter phase angles*. This reference plane shifting is an option built into most network analyzers.

For example, suppose we wish to extend the reference planes an additional 10° from the input port and an extra 15° from the output. The new matrix, in terms of the original S parameters, becomes

$$\begin{pmatrix} s_{11}e^{-j20^\circ} & s_{12}e^{-j25^\circ} \\ s_{21}e^{-j25^\circ} & s_{22}e^{-j30^\circ} \end{pmatrix}_{\text{New}} \tag{6.5-10}$$

Similarly, by reversing the signs in the exponential factors, we can remove line lengths that may have separated an actual device, such as a transistor from the test ports of a test fixture, a process called *deembedding*.

Since the net input power to a lossless network must equal the net output power, it is possible to show that certain relationships hold for the S parameters of a lossless network [3, p. 544]. For example, *for a lossless network, the sum of the squares of the magnitudes of the S parameters of any column is unity:*

$$\sum_{n=1}^{N} |S_{ni}|^2 = 1 \tag{6.5-11}$$

where i is any column from 1 to N for a $N \times N$ matrix.

This is easily proved. Consider, for example, a lossless two-port network that we excite at port 1 with one unit of power and place a matched load at port 2. The total power into the lossless network must equal the total power that emerges, or

$$|a_1^2| = 1 = |b_1^2| + |b_2^2| = |S_{11}^2| + |S_{21}^2| \tag{6.5-12}$$

This argument can be used by successive application of power to any port with matched loads at the other ports. It applies to a lossless network having any number of ports. For a lossless two-port network this means that

$$|S_{11}|^2 + |S_{21}|^2 = 1 \tag{6.5-13}$$

$$|S_{12}|^2 + |S_{22}|^2 = 1 \tag{6.5-14}$$

It is also true that *for a lossless network, the sum of the products of the elements of any column by the complex conjugates of the corresponding elements of another column is zero:*

$$\sum_{n=1}^{N} S_{na} S_{nb}^* = 0 \qquad a \neq b \tag{6.5-15}$$

A matrix whose elements satisfy (6.5-11) and (6.5-15) is called a *unitary matrix* [7, p. 175]. If the network is reciprocal, then $S_{ij} = S_{ji}$ and (6.5-11) and (6.5-15) also apply to rows. Applying this to a lossless two-port network,

$$S_{11} S_{12}^* + S_{21} S_{22}^* = 0 \tag{6.5-16}$$

Because of their numerous advantages, S parameters are standard output format options for nearly all microwave measurements.

However, the advantages of S parameters come at a price. Unlike Z, Y, and $ABCD$ parameters, whose values are independent of the measurement system, the evaluation of S parameters is dependent upon the characteristic impedances (port reference impedances) of the transmission lines connected to their ports. *The S parameters are only valid when employed with the reference impedances used for their measurement*. Usually this is 50 Ω. For example, given 50-Ω S parameters for a transistor, they cannot be used directly for an amplifier design intended for 75 Ω characteristic impedance. While a mathematical conversion could be performed, common practice is to remeasure the transistor using a 75-Ω characteristic impedance network analyzer.

6.6 THE TRANSMISSION MATRIX

Since the a waves used with the S parameters are directed inward and the b waves outward, the overall S matrix of a cascade of two ports cannot be

formed by matrix multiplication of the S matrices of the elements of the cascade. However, a and b waves can be rearranged in the defining equations to overcome this limitation. The resultant matrix is known as the *wave transmission matrix*, or *T matrix*. The T matrix uses the same definitions for the a and b waves shown in Figure 6.5-1.

As was true of the $ABCD$ matrix, the T matrix achieves its cascading attribute merely by redefining the input and output equation format so that it is compatible with cascade multiplication. One such T matrix definition is

$$\begin{pmatrix} b_1 \\ a_1 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} \tag{6.6-1}$$

in which the a and b waves have the same definitions as previously. The necessary changes of direction of the waves is accommodated in the signs of the respective T parameters. The T matrices can be cascade multiplied in the same manner used for $ABCD$ matrices.

Since measurements are commonly made using S parameters only, conversions between T and S matrices are useful. The equivalences, assuming $Z_{01} = Z_{02}$, are [3, p. 548]:

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{s_{12}s_{21} - s_{11}s_{22}}{s_{21}} & \dfrac{s_{11}}{s_{21}} \\ \dfrac{-s_{22}}{s_{21}} & \dfrac{1}{s_{21}} \end{pmatrix} \quad (Z_{01} = Z_{02}) \tag{6.6-2}$$

and

$$\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{T_{12}}{T_{22}} & \dfrac{T_{11}T_{22} - T_{12}T_{21}}{T_{22}} \\ \dfrac{1}{T_{22}} & \dfrac{-T_{21}}{T_{22}} \end{pmatrix} \quad (Z_{01} = Z_{02}) \tag{6.6-3}$$

Reciprocity requires that [5]

$$S_{12} = S_{21} \tag{6.6-4}$$

and equivalently,

$$T_{11}T_{22} - T_{12}T_{21} = 1 \tag{6.6-5}$$

Care must be taken in the use of the conversions in (6.6-2) and (6.6-3) since they are based on the defining relations in (6.6-1). There are other defining relationships used for the transmission matrix for which different conversion formulas must be used.

REFERENCES

1. George B. Thomas, Jr., *Calculus and Analytic Geometry*, Addison-Wesley, Cambridge, MA, 1953. *This is a classic calculus textbook.*
2. Simon Ramo and John R. Whinnery, *Fields and Waves in Modern Radio*, Wiley, New York, 1944, 1953 (later revised with a third author, Van Duzer). *This is a classic introductory text describing fields and Maxwell's equations.*
3. Peter A. Rizzi, *Microwave Engineering, Passive Circuits*, Prentice-Hall, Englewood Cliffs, NJ, 1988. *Excellent microwave engineering textbook covering theory and design of transmission lines, couplers, filters, and numerous other passive devices.*
4. Robert E. Collin, *Field Theory of Guided Waves*, McGraw-Hill, New York, 1960.
5. Joseph F. White, *Microwave Semiconductor Engineering*, Noble Publishing, Norcross GA, 1995. *Extensive treatment of PIN diodes and their switching and phase shifting applications. Also includes many fundamental treatments of microwave circuits.*
6. S Parameter Directory, within the *Genesys* network simulator software, Eagleware, 630 Pinnacle Court, Norcross, GA 30071 circa 2000.
7. Robert E. Collin, *Foundations of Microwave Engineering*, McGraw-Hill, New York, 1966. *Collin is a superb theoretician. This textbook gives a fundamental and precise introduction to microwave engineering.*

EXERCISES

E6.2-1 Show that the Z parameters for a two-port network consisting of a length of lossless transmission line of characteristic impedance Z_0 and electrical length θ are

$$\begin{pmatrix} Z_{11} & Z_{12} \\ Z_{211} & Z_{22} \end{pmatrix} = \begin{pmatrix} -jZ_0 \cot\theta & jZ_0 \csc\theta \\ jZ_0 \csc\theta & -jZ_0 \cot\theta \end{pmatrix}$$

E6.2-2 Derive an expression for the input impedance Z_{IN} to a two-port network in terms of its Z parameters. How does this derivation compare in difficulty to that of (6.4-5) for use with the *ABCD* matrix?

E6.2-3 Use the Z parameters you verified in E6.2-1 along with the input impedance formula you derived in E6.2-2 to show that the input impedance Z_{IN} to the network is always Z_0 when the network consists of a Z_0 transmission line of any length, θ, terminated in Z_0.

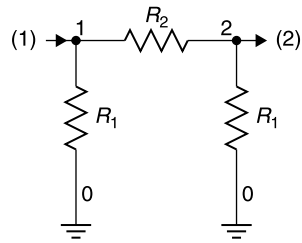
E6.4-1 **a.** Design a Pi circuit of the form shown in E6.4-2 so that it is equivalent at 1 GHz to a 90° length of 50-Ω transmission line. Then use a network simulator to answer.

b. What is the bandwidth of a maximum insertion loss of 0.1 dB?

c. Over what bandwidth is the insertion phase shift, φ, within 10° for the two circuits?

d. The group delay for the line section is 0.25 ns (a quarter period at 1 GHz). Over what bandwidth does the Pi circuit have a group delay ($-\partial\varphi/\partial\omega$) that is within 10% of this value (within 0.025 ns of 0.25 ns)?

E6.4-2 Design a Pi circuit variable attenuator in the format shown in the figure below.



a. Solve for the relationship of R_2 in terms of R_1 such that the attenuator is matched to a Z_0 load.

b. From the results of your analysis, are there any restrictions on the values of R_1 or R_2?

E6.4-3 **a.** Derive an expression for the attenuation (insertion loss) of the Pi attenuator circuit of E6.4-2. *Hint:* Use the formula for insertion loss in terms of the $ABCD$ parameters.

b. Find the numeric values for R_1/Z_0 and R_2/Z_0 to obtain 10 dB of insertion loss.

E6.4-4 A lumped element phase shifter consists of switching between the *low-pass* and *high-pass* networks shown in the figure below. Note that the values are normalized to $Z_0 = Y_0 = 1$.

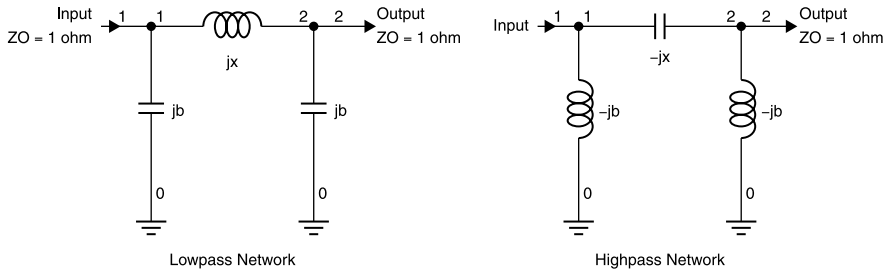
a. Calculate the overall matrix, $ABCD_T$, for the three-element network treating the shunt elements as jb and the series element as jx, where the signs of x and b can be plus or minus (same matrix will apply to both networks).

b. Find the relation between x and b to obtain a match ($Z_{IN} = Z_0 = 1$) in general jb and jx terms applicable to either network.

c. Find the values of x and b for each network that provide a match.

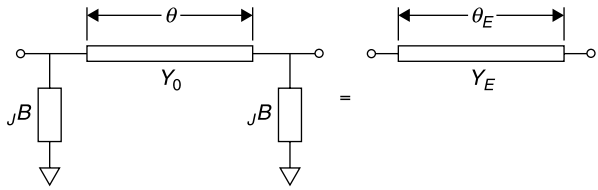
d. Find the values of x and b that give -90° of transmission phase through the low-pass network and $+90^\circ$ through the high-pass network. (Then switching between these two circuits will provide a phase shift of 180°.)

e. Use a network simulator to determine the bandwidth over which each circuit has an input VSWR < 1.2 (or a return loss < 20 dB).



E6.4-5 **a.** Find the overall matrix for the three-element, loaded-line network, $ABCD_T$, shown in the figure below.

b. Use the result to find an equivalent length of transmission line of Y_E characteristic admittance and θ_E electrical length which represents the three-element network. *Hint:* To do this, compare the $ABCD_T$ with the $ABCD$ matrix of a uniform line of Y_E admittance and θ_E electrical length.



c. Use the result in part (b) to find the electrical spacing, θ, that matches two similar lossless discontinuities (each represented as a shunt susceptance, $\pm jB$) on a transmission line ($Z_{\text{IN}} = Z_0$ when $Z_L = Z_0$). Assume the line, source, and load impedances are Z_0 (characteristic admittance $Y_0 = 1/Z_0$). Express the result in terms of normalized susceptance, $\pm b$ (normalized to Y_0).

E6.4-6 It is proposed to make a phase trimmer for a 50-Ω cable that consists of a pair of capacitive screws spaced a quarter wavelength apart. This is called a *transmission phase shifter*.

a. Use the equivalent line length formulation of E6.4-5 to determine the maximum amount of capacitance each screw must have to provide an insertion phase tuning range of 0° to 11° at 1 GHz.

b. Use a network simulator to determine the bandwidth over which the circuit is well matched (VSWR < 1.10) if the design center frequency is 1 GHz and the tuning capacitors are adjusted to provide the maximum "line lengthening" of $+11°$.

c. Suppose that a long cascade of these circuits is formed, each loaded-line network beginning at the output of the preceding one.

What would be the maximum VSWR of the cascade? *Hint:* Find the Y_E of the cascade.

E6.5-1 **a.** Determine the S parameters for a lossless 50-Ω transmission line having an electrical length, θ, and located between Z_0 load and Z_0 source.

b. How does this solution compare with that obtained for the Z parameters of a lossless line in E6.2-2? What differences would there be in using these two different network representations for a length of transmission line?

E6.5-2 Find the S matrix for a series inductor of 7.96 nH at 1 GHz and referenced to 50-Ω.

E6.5-3 Verify the S matrix that you obtained in E6.5-2 is unitary, as it should be for a lossless circuit.

E6.5-4 How does the insertion loss of the series inductor network calculated in E6.5-2 using the S matrix compare with that for a series impedance given by (2.10-4)? Assume matched generator and load ($Z_G = Z_L = Z_0$).

CHAPTER 7

Electromagnetic Fields and Waves

7.1 VECTOR FORCE FIELDS

Phenomena by which noncontact forces can be exerted from a distance are called *vector force fields*. From our earliest experiences, we are familiar with the noncontacting force of gravity. On the Earth's surface it is essentially a constant downward force. More precisely, the gravitational vector force field can be described as *the force of gravity F between any two masses m_1 and m_2 is attractive, directly proportional to their product and inversely proportional to the square of the distance between them and directed along the radius line r connecting their centers of mass:*

$$\vec{\mathbf{F}} = \frac{m_1 m_2 G}{r^2}\vec{r} \tag{7.1-1}$$

where $\vec{r}$ is a unit vector in the r direction. For this expression a positive force $\vec{\mathbf{F}}$ is attractive.

Based upon Earth-borne and celestial observations, it is believed that (7.1-1) is applicable to all masses throughout the universe. It can be applied on the moon or any other location. Of course, gravitational forces from every star, planet, moon, asteroid, and every other mass in the universe exert a force component on us; but the Earth exerts so much larger a force that we can neglect the others with only small errors. The gravity field and its mathematical expression in (7.1-1) is an empirically developed *primary concept*. It cannot be derived from more fundamental principles (Fig. 7.1-1).

We cannot see the force field, $\vec{F}$, due to gravity: but since its effect is a daily experience, the gravity field seems intuitively obvious to most individuals. Few persons claim not to understand gravity because it is an invisible force.

In the same manner the presence and behavior of equally invisible *electric and magnetic fields*, or simply *EH fields*, must be intuitively familiar to the insightful RF and microwave engineer. Their presence is experienced in varying

High Frequency Techniques: An Introduction to RF and Microwave Engineering, By Joseph F. White.
ISBN 0-471-45591-1

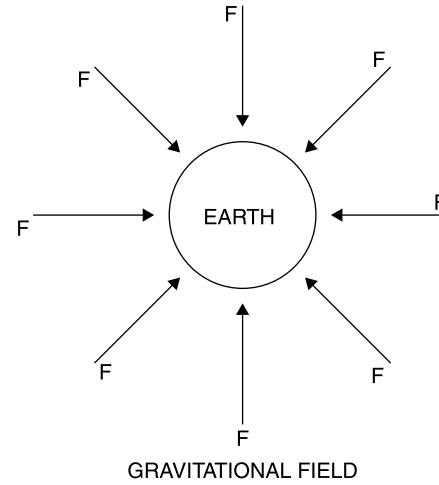


Figure 7.1-1 Earth's noncontacting force field, gravity.

degrees within all circuits and throughout all spaces, even extending throughout the entire known universe. Gravity and electromagnetic waves have in common that they are the only known wave phenomena that can propagate through a vacuum. However, gravity "waves," that is, a time-varying change in the gravitational force have not yet been observed, the difficulty being that a change of mass is required to produce them. Since this requires the destruction or creation of matter, gravity waves have not been and probably never will become practical for routine communicative purposes.

The $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields, however, are eminently practical for communication, since their creation requires merely the separation of electrical charges or the creation of electric currents, both being stimuli that are easy to arrange. Like gravity, behaviors of $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields conform to empirically developed laws.

These are summarized by *Maxwell's equations, which* he formulated from various empirical observations made by others [1, Chapters 1 and 2]. These equations were integrated into a unified theory of electromagnetism to give precise mathematical expressions to Gauss's laws for electric and magnetic fields, Faraday's law (with Lenz's law) of induction, and Ampere's law, with displacement current added by Maxwell.

Using his equations James Clerk Maxwell proved mathematically that electromagnetic waves could propagate through a nonconducting medium at the speed of light. He formulated his unified theory in about 1862 [2, p. 73], but his equations were more numerous at that time. It remained for Oliver Heaviside [3] to reduce them to the four equations, as we now know them. Maxwell had stated that *"we have strong reason to conclude that light itself is an electromagnetic disturbance in the form of waves."* Later, Heinrich Hertz verified Maxwell's theory of electromagnetic waves, but the ideas were not widely accepted.

It took time to displace the erroneous but commonly held view that the

transmission of visible light required a "luminiferous aether" that pervaded all space and material. It had no weight but elastic properties to permit it to propagate the disturbance that light waves represented (as the elasticity of a drum's head allows the disturbance caused by striking it with a drumstick to propagate across the membrane).

> Faraday had speculated upon this possibility [that light was a form of electric wave] several years earlier [than Maxwell's similar conclusion], and had pointed out that whereas one infinite and all-pervasive imponderable aether is a severe strain upon one's imagination, belief in two co-existent, infinite, all-pervasive, and imponderable aethers, one for light and one for electricity, is simply beyond the limits of credulity. Therefore he suspected, on this basis alone, that light is an electric phenomenon [4, p. 112].

Usually Maxwell's equations are introduced on a mathematical basis. Since they describe three-dimensional field quantities, this requires a knowledge of vector mathematics which, when studied separately, can be complex and may seem to be arbitrary in definition. In our treatment EH fields will be introduced through the definitions, applications, and experiments by which they were first discovered. Within this context we shall introduce, as needed, the vector mathematics required to describe them. The author believes that this approach provides a more intuitive understanding, not only of the behavior of EH fields and waves but also of the fundamentals of vector mathematics needed to describe and analyze them.

7.2 E AND H FIELDS

Wireless transmission occurs by means of propagating $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields, or electromagnetic waves. *Electromagnetic waves—comprising radio, microwave, millimeter waves, infrared, visible light, ultraviolet and X rays—are the only practical means of propagating energy and thereby communicating through a vacuum.*

Electromagnetic waves are as fundamental to radio as the Big Bang is to the origin of the known universe. The $\vec{E}$ *(electric) and* $\vec{H}$ *(magnetic) fields are vector force fields, having both magnitude and direction at every point. They can be static (non-time-varying), but for propagating applications they must vary in time, and when employed for communication this variation usually is sinusoidal or can be approximated by a sum of sinusoidal components.*

7.3 ELECTRIC FIELD E

The electric field $\vec{E}$ *at a given position is the magnitude and direction of the force* $\vec{F}$ *which is experienced by a unit positive charge q when it is exposed to the* $\vec{E}$ *field at that position:*

$$\vec{E} = \frac{\vec{F}}{q} \tag{7.3-1}$$

In the MKS system, $\vec{F}$ is in newtons, q is in coulombs, and $\vec{E}$ is in newtons/coulomb (or equivalently, volts/meter).

Electrostatically, the force between two charges is expressed by *Coulomb's law*. It is analogous to the gravitational force between two masses. Coulomb's law states that the force exerted on a point charge q due to another point charge q_1 is directly proportional to the product qq_1 and inversely proportional to the square of the distance r between them. That is,

$$\vec{F} = \frac{qq_1}{4\pi\varepsilon r^2}\vec{r} \tag{7.3-2}$$

where $\varepsilon = \varepsilon_0\varepsilon_R$, ε_0 is a constant (see inside cover for value), ε_R is a function of the dielectric material, and $\vec{r}$ is a unit vector along the line between the two charges and directed away from q. Thus, a positive value of $\vec{F}$ is a force of repulsion while a negative value indicates an attractive force. It follows from (7.3-1) that the electric field due to a point charge q_1 is given by

$$\vec{E} = \frac{q_1}{4\pi\varepsilon r^2}\vec{r} \tag{7.3-3}$$

For example, suppose a 9-V battery is connected to a vertically oriented capacitor with parallel plates 1 cm apart in a vacuum. Since the $\vec{E}$ field can be presumed uniform between the plates, its integral with respect to distance, $E \times d = V$, is potential energy or voltage. The $\vec{E}$ field will be 9 V/cm or 900 V/m. Then suppose that a single electron is placed on the lower plate and released [5, Vol. 2, pp. 420–421].

What is the ratio of the gravitational force on the electron compared to the force of the electric field?

$$\text{Electronic charge} = e = 1.60 \times 10^{-19}\ \text{C}$$

$$\text{Electronic mass} = 9.1 \times 10^{-31}\ \text{kg}$$

$$F_{\text{electric}} = eE$$

$$= (1.6 \times 10^{-19}\ \text{C}) \times (0.9 \times 10^{3}\ \text{N/C}) = 1.4 \times 10^{-16}\ \text{N} \tag{7.3-4}$$

$$F_{\text{gravity}} = mg$$

$$= (9.1 \times 10^{-31}\ \text{kg}) \times (9.8\ \text{N/kg}) = 8.9 \times 10^{-30}\ \text{N} \tag{7.3-5}$$

The ratio of the electric to gravitational forces is

$$\frac{F_e}{F_g} = \frac{1.4 \times 10^{-16}}{8.9 \times 10^{-30}} = 1.6 \times 10^{13} \tag{7.3-6}$$

Thus, the gravitational force, even in this fairly mild electric field, is negligible compared to the electric force.

Starting from rest, to what vertical speed will the electron accelerate before it strikes the opposite plate?

$$\text{Acceleration} = a = \frac{F}{m} = \frac{eE}{m}$$
$$= \frac{(1.60 \times 10^{-19}\ \text{C}) \times (0.9 \times 10^{3}\ \text{N/C})}{9.1 \times 10^{-31}\ \text{kg}} = 1.6 \times 10^{14}\ \text{m/s}^2 \quad (7.3\text{-}7)$$

The velocity after being accelerated through a distance x is

$$\text{Velocity} = \sqrt{2ax}$$
$$= \sqrt{(2 \times 1.6 \times 10^{14}\ \text{m/s}^2) \times 10^{-2}\ \text{m}} = 1.8 \times 10^{6}\ \text{m/s}$$
$$= 6 \times 10^{6}\ \text{ft/s} = 8{,}530{,}000\ \text{mph!} \quad (7.3\text{-}8)$$

How long will it take for the electron to traverse the 1 cm distance?

$$t = \frac{v}{a} = \frac{1.8 \times 10^{6}\ \text{m/s}}{1.6 \times 10^{14}\ \text{m/s}^2} = 1.1 \times 10^{-8}\ \text{s} = 11\ \text{ns} \quad (7.3\text{-}9)$$

7.4 MAGNETIC FLUX DENSITY

For magnetic force problems we use the flux density vector $\vec{\mathbf{B}}$ to calculate the magnitude and direction of the force. For homogenous media $\vec{\mathbf{B}}$ and $\vec{\mathbf{H}}$ are related by a constant. In the case of magnetic forces acting upon a charge, the situation is more complex than that of electric field forces. This is because *the magnetic force on a charge only acts when the charge is moving relative to the field and then the direction of the force is at right angles to the plane that contains the field and the charge's velocity vectors* (Fig. 7.4-1).

The magnetic induction ($\vec{\mathbf{B}}$ field) is one weber/m^2 (one Tesla) when one coulomb of charge, moving with a perpendicular component of velocity of one m/s ($\vec{\mathbf{v}} \sin \varphi$) experiences a force of one newton.

The condition of maximum force on the moving charge occurs when the charge's velocity $\vec{\mathbf{v}}$ is oriented exactly 90° relative to the $\vec{\mathbf{B}}$ field. However, in general, the charge's velocity is at an angle φ relative to the $\vec{\mathbf{B}}$ field and the magnitude of the force $\vec{\mathbf{F}}$ is given by

$$\mathrm{F} = \mathrm{qvB} \sin \varphi \quad (7.4\text{-}1)$$

where $\vec{\mathbf{F}}$ is a force in the plane orthogonal to the plane of $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$ and in the

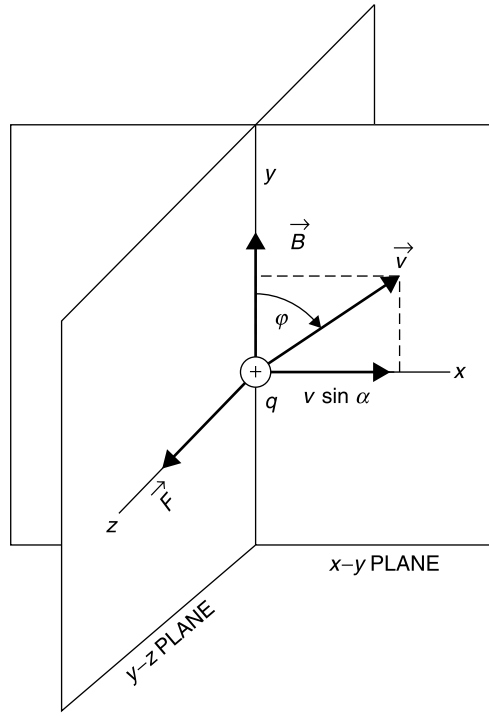


Figure 7.4-1 Vectors $\vec{\mathbf{B}}$ and $\vec{\mathbf{v}}$ lie in a plane that is orthogonal to the resulting force $\vec{\mathbf{F}}$ on the moving charge q.

direction of a right-hand screw's advance when $\vec{\mathbf{v}}$ is rotated into $\vec{\mathbf{B}}$ through the smaller angle φ between $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$.

The expression in (7.4-1) relates the magnitudes of the vectors $\vec{\mathbf{F}}, \vec{\mathbf{v}}$, and $\vec{\mathbf{B}}$ but further explanation is necessary to convey their relative directions. For this situation the definition of the *vector cross product* proves useful.

7.5 VECTOR CROSS PRODUCT

The vector cross product, (pronounced "$\vec{\mathbf{A}}$ *cross* $\vec{\mathbf{B}}$*") of two vectors* $\vec{\mathbf{A}}$ *and* $\vec{\mathbf{B}}$ *is equal to a third vector* $\vec{\mathbf{C}}$ *whose magnitude is equal to the product of the magnitudes of* $\vec{\mathbf{A}}, \vec{\mathbf{B}}$ *and the sine of the angle* φ *between them and whose direction is perpendicular to the plane of* $\vec{\mathbf{A}}$ *and* $\vec{\mathbf{B}}$ *with a positive sense defined as that of the advance of a right-hand screw were* $\vec{\mathbf{A}}$ *to be rotated into* $\vec{\mathbf{B}}$ *through the smaller angle between them:*

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} \equiv \vec{\mathbf{C}} \tag{7.5-1}$$

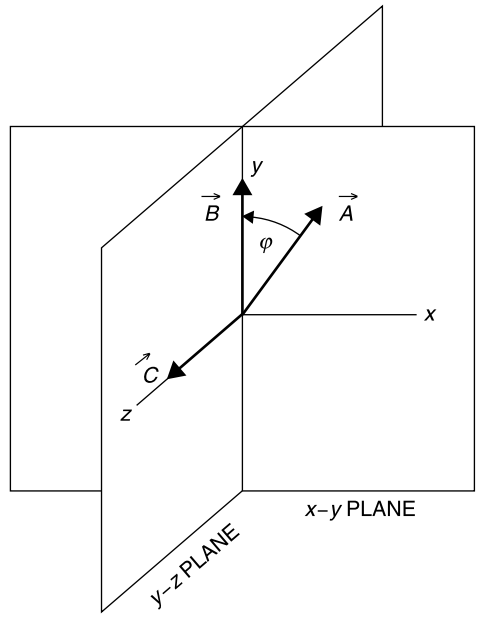


Figure 7.5-1 Vector cross product.

This definition, although lengthy in words, completely describes the three-dimensional geometry of Figure 7.5-1. Once learned, only two rules are required to remember and apply the vector cross product. *First*, $\vec{\mathbf{A}}$ is rotated into $\vec{\mathbf{B}}$ to produce a vector $\vec{\mathbf{C}}$ that advances in the direction of a right-hand screw. *Second*, the magnitude of $\vec{\mathbf{C}}$ is the product of the magnitudes of $\vec{\mathbf{A}}$, $\vec{\mathbf{B}}$, and sin φ, where φ is the angle (less than 90°) between them.

Keep in mind that a positive or negative sign applies to the curl according to the sense of rotating $\vec{\mathbf{A}}$ into $\vec{\mathbf{B}}$. For example, in Figure 7.5-2, the rotation of $\vec{\mathbf{A}}_1$ into $\vec{\mathbf{B}}$ through φ_1 produces a resultant, $\vec{\mathbf{C}}_1$ in the $+\vec{\mathbf{z}}$ direction. However, the rotation of $\vec{\mathbf{A}}_2$ through φ_2 produces the opposite rotational sense and, accordingly, a $\vec{\mathbf{C}}_2$ in the $-\vec{\mathbf{z}}$ direction. In either case it is necessary to *use the smaller angle*, φ, *between* $\vec{\mathbf{A}}$ *and* $\vec{\mathbf{B}}$ to determine the rotational sense of $\vec{\mathbf{A}}$ into $\vec{\mathbf{B}}$ and from it the corresponding sign of $\vec{\mathbf{C}}$.

Because the sense of the advance of a right-hand screw changes when the order of the vectors of the cross product is reversed, it follows that

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = -\vec{\mathbf{B}} \times \vec{\mathbf{A}} \tag{7.5-2}$$

Notice that Figure 7.4-1 displays a *right-handed coordinate system* wherein crossing $\vec{\mathbf{x}}$ into $\vec{\mathbf{y}}$ yields a vector in the $+\vec{\mathbf{z}}$ direction, $\vec{\mathbf{y}}$ into $\vec{\mathbf{z}}$ yields $+\vec{\mathbf{x}}$, and $\vec{\mathbf{z}}$ into $\vec{\mathbf{x}}$ yields $\vec{\mathbf{y}}$. In other words, crossing any two successive positive directional vectors in the xyz coordinate order of a right-hand coordinate system yields a

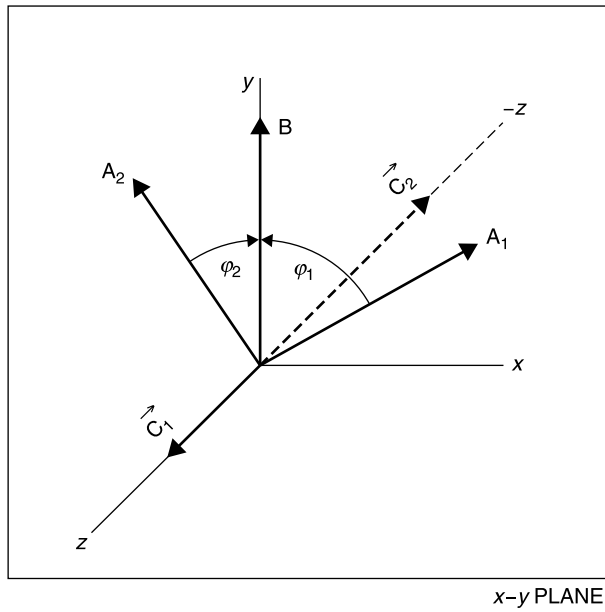


Figure 7.5-2 Direction (sign) of $\vec{\mathbf{C}}$ changes with the sense of rotation of $\vec{\mathbf{A}}$ into $\vec{\mathbf{B}}$.

cross product in the positive direction of the third. It is best to use a right-hand coordinate system to avoid errors in calculating the cross product. Stated in vector notation [6, pp. 88–89],

$$\vec{x} \times \vec{y} = \vec{z} = -\vec{y} \times \vec{x} \tag{7.5-3a}$$

$$\vec{y} \times \vec{z} = \vec{x} = -\vec{z} \times \vec{y} \tag{7.5-3b}$$

$$\vec{z} \times \vec{x} = \vec{y} = -\vec{x} \times \vec{z} \tag{7.5-3c}$$

$$\vec{x} \times \vec{x} = \vec{y} \times \vec{y} = \vec{z} \times \vec{z} = 0 \tag{7.5-3d}$$

where $\vec{x}$, $\vec{y}$, and $\vec{z}$ are unit vectors in the respective positive directions of the x, y, and z coordinate directions. If the vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ are expressed in terms of their rectangular coordinates in this right-hand system,

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = (\mathrm{A}_x\vec{x} + \mathrm{A}_y\vec{y} + \mathrm{A}_z\vec{z}) \times (\mathrm{B}_x\vec{x} + \mathrm{B}_y\vec{y} + \mathrm{B}_z\vec{z}) \tag{7.5-4}$$

where $\mathrm{A}_x, \mathrm{A}_y$, and A_z are the components of $\vec{\mathbf{A}}$ in the $\vec{x}$, $\vec{y}$, and $\vec{z}$ directions, and similarly with $\mathrm{B}_x, \mathrm{B}_y$, and B_z. The cross product of (7.5-4) would be expected to produce nine terms, since each component of $\vec{\mathbf{A}}$ is crossed separately into each component of $\vec{\mathbf{B}}$. However, vector components of $\vec{\mathbf{A}}$ which have the same corresponding directions as components of $\vec{\mathbf{B}}$ produce a zero cross product according to (7.5-3d), so only six terms are nonzero. By observing the cross products

in (7.5-3a,b, and c), (7.5-4) can be expanded explicitly as

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = (\mathrm{A}_y\mathrm{B}_z - \mathrm{A}_z\mathrm{B}_y)\vec{x} + (\mathrm{A}_z\mathrm{B}_x - \mathrm{A}_x\mathrm{B}_z)\vec{y} + (\mathrm{A}_x\mathrm{B}_y - \mathrm{A}_y\mathrm{B}_x)\vec{z} \quad (7.5\text{-}5)$$

A short-hand method for writing and remembering (7.5-5) is the solution to the determinant

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = \begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ \mathrm{A}_x & \mathrm{A}_y & \mathrm{A}_z \\ \mathrm{B}_x & \mathrm{B}_y & \mathrm{B}_z \end{vmatrix} \quad (7.5\text{-}6)$$

Expressions (7.5-5) and (7.5-6) produce the same result. Given the definition of the vector cross product, the force on a moving charge can be written very tersely yet in a manner than conveys all of the relationships between the magnitudes and directions of $\vec{\mathbf{F}}, \vec{\mathrm{v}}$, and $\vec{\mathbf{B}}$.

If a charge q has velocity v *that is at right angles to the magnetic field* $\vec{\mathbf{B}}$*, the force on the moving charge q is given as*

$$\vec{\mathbf{F}}_{\text{magnetic}} = q\vec{\mathrm{v}} \times \vec{\mathbf{B}} \quad (7.5\text{-}7)$$

Some useful conversions for the units of magnetic flux, $\vec{\mathbf{B}}$, are

$$1\ \text{T (Tesla)} = 10{,}000\ \text{G (Gauss)} = 1\frac{\text{Wb (Weber)}}{\text{m}^2}$$
$$= 1\frac{\text{N (Newton)}}{\text{C (Coulomb)} \times \text{m/s}} = 1\frac{\text{N (Newton)}}{\text{Am (Ampere)}} \quad (7.5\text{-}8)$$

For example, an electron moving at 1×10^6 ms/s at right angles to a 1-T magnetic field would experience a force given by

$$\mathrm{F}_{\text{magnetic}} = 1.6 \times 10^{-19}\ \text{C} \times 1 \times 10^6\ \text{m/s} \times 1\frac{\text{N}}{\text{C} \times \text{m/s}} = 1.6 \times 10^{-13}\ \text{N} \quad (7.5\text{-}9)$$

The force of gravity on the same electron is

$$\mathrm{F}_{\text{gravity}} = mg = 9.1 \times 10^{-31}\ \text{kg} \times 9.8\ \text{N/kg} = 8.9 \times 10^{-30}\ \text{N} \quad (7.5\text{-}10)$$

Again, the gravitational force is negligible. Notice that the magnetic force is 1000 times as great as the electric force calculated for the electric field example, given this 1-T field strength. One tesla is a strong magnetic field.

When a charged particle enters at right angles to a magnetic field at a constant velocity, its orbit is a circle [5, Vol. 2, p. 561]. The circular electron orbit is the basis of the magnetron microwave tube [7, p. 41; 8, p. 321]. Even in the 1940s magnetrons were capable of 100 kW of peak power for short pulses of a microsecond or less. These high pulsed power levels were orders of magnitude greater than those achievable with prior tube types and made practical the development of tactical radars during the latter part of World War II.

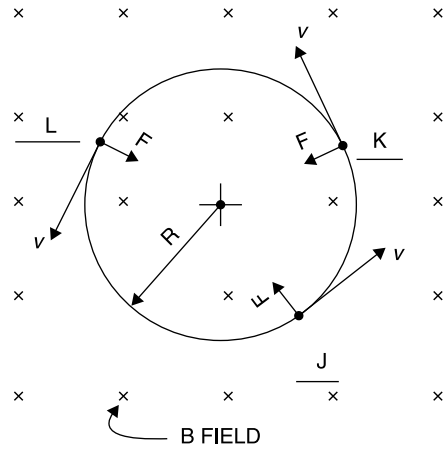


Figure 7.5-3 Positive charge with v speed enters a $\vec{B}$ field (crosses indicate $\vec{B}$ field direction is "into the paper") at point J and has its path turned into a circular orbit (curving to points K and L) with a radius R. The magnetic force $\vec{F}$ is always directed toward the circular orbit's center and is equal and opposite to the particle's centrifugal force.

The centrifugal force of a particle traveling in a circular orbit of radius R (Fig. 7.5-3) is

$$\mathrm{F}_{\text{centrifugal}} = \frac{mv^2}{R} \tag{7.5-11}$$

Equating this to the magnetic force gives [5, Vol. 2, p. 561]

$$q\mathrm{vB} = m\frac{\mathrm{v}^2}{R} \tag{7.5-12}$$

and thus,

$$R = \frac{m\mathrm{v}}{\mathrm{B}q} \tag{7.5-13}$$

In addition to the magnetron tube, this method of orbiting a charged particle also is employed in the cyclotron particle accelerator and in mass spectrographs to analyze ionized samples—the radii of whose orbits are directly proportional to the particle's mass and hence its atomic element number.

As an example, suppose that an electron enters a magnetic field of flux density $\vec{B} = 2 \times 10^{-4}$ T (2 G) traveling in a direction that is at right angles to $\vec{B}$ with a velocity of 1×10^6 m/s:

$$R = \frac{mv}{Bq} = \frac{(9.1 \times 10^{-31}\ \text{kg}) \times (1 \times 10^6\ \text{m/s})}{(2 \times 10^{-4}\ \text{T}) \times (1.6 \times 10^{-19}\ \text{C})} = 2.85\ \text{cm} = 1.12\ \text{in.} \tag{7.5-14}$$

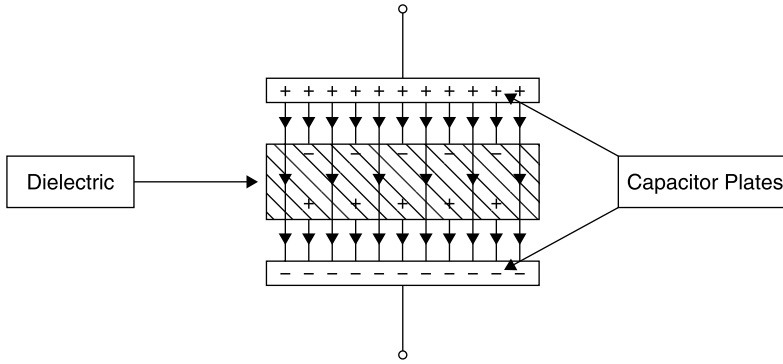


Figure 7.6-1 Induced charges on a dielectric inserted between the plates of a capacitor reduce the electric field within the dielectric.

7.6 ELECTROSTATICS AND GAUSS'S LAW

The magnitude of the electric field $\vec{E}$ depends upon the medium in which it is immersed [5, Vol. 2, p. 478]. For example, when a dielectric material is placed between the charged plates of a capacitor, induced charges in the dielectric cause the $\vec{E}$ field lines to terminate on them, reducing the $\vec{E}$ field within the dielectric if the capacitor is open circuited (Fig. 7.6-1). If a constant voltage is applied the voltage source will supply additional charge to the capacitor to build up to the original $\vec{E}$ field and voltage between the plates.

The electric field at a distance r from a fixed point charge q is given by (7.3-3) and repeated here (Fig. 7.6-2):

$$\vec{E} = \frac{q}{4\pi\varepsilon_0\varepsilon_R r^2}\vec{r} \tag{7.6-1}$$

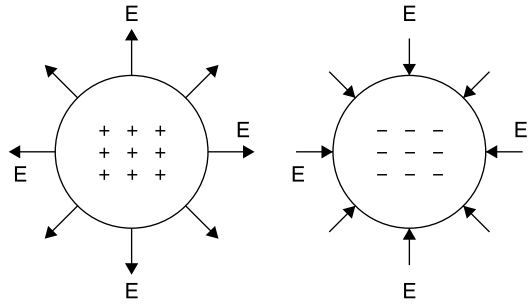


Figure 7.6-2 Coulomb observed that the electric field at a radial distance r from a point charge q is proportional to q and inversely proportional to the square of r. The field direction points away from a positive charge (to repel a like positive charge) and inward toward a negative charge (to attract an opposite, positive charge).

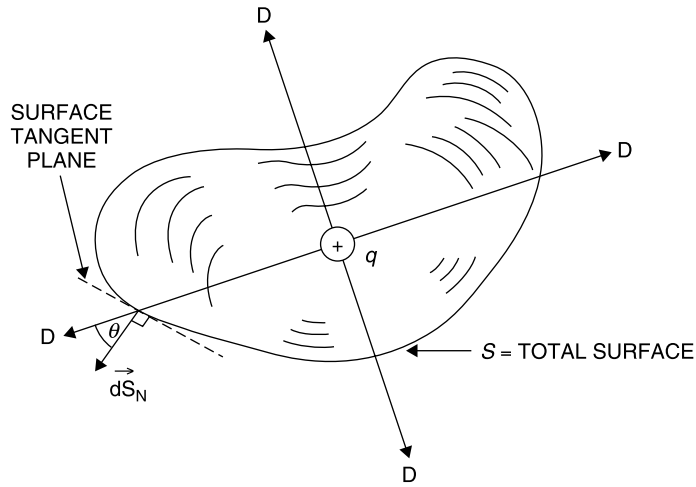


Figure 7.6-3 Arbitrary volume encloses a charge, q. The resulting $\vec{\mathbf{D}}$ vector field exits the volume radially over the closed surface S. Incrementally, the surface is divided into dS areas having a surface normal direction $\overrightarrow{dS}$ at angle θ to the vector field $\vec{\mathbf{D}}$.

This observation can be written without specifying the dielectric environment in which the charge is found by defining a companion vector, *electric flux density* $\vec{\mathbf{D}}$ as

$$\vec{\mathbf{D}} = \varepsilon\vec{\mathbf{E}} = \varepsilon_0\varepsilon_R\vec{\mathbf{E}} = \frac{q}{4\pi r^2}\vec{r} \tag{7.6-2}$$

The reason for defining $\vec{\mathbf{D}}$ is that the effect of induced charge in the dielectric is absorbed in ε_R and therefore the right-hand side (7.6-2) is only required to account for free charge. A more general observation that is suggested by the spherical geometry of Figure 7.6-2 is known as *Gauss's law*.

Gauss's law states that for any closed surface containing a total fixed charge q, the product of the emerging $\vec{\mathbf{D}}$ *field magnitude normal to each infinitesimally small area* $\overrightarrow{dS}_N$*, summed over the entire containing surface, S, is equal to the total charge q contained within the surface* (Fig. 7.6-3). A mathematical expression of this statement is provided in the next section.

7.7 VECTOR DOT PRODUCT AND DIVERGENCE

To describe the mathematical process that Gauss's law requires, we assign a vector, $\overrightarrow{d}$, to the surface that is always normal to and has the magnitude dS of a small area of the surface at each point on the surface. It has a direction defined as positive when directed toward the outside of the volume contained

by the surface. The vector field $\vec{\mathbf{D}}$ (Fig. 7.6-3) resulting from charge q is a function of the x, y, z position on the surface S. Then, everywhere on the surface S we form the product

$$\text{Vector product} = |\vec{\mathbf{D}}|\,|\overrightarrow{\mathrm{dS}}|\cos\theta \tag{7.7-1}$$

where θ is the angle between $\vec{\mathbf{D}}$ and $\overrightarrow{\mathrm{d}}$. Note that when $\vec{\mathbf{D}}$ and $\overrightarrow{\mathrm{dS}}$ point in opposite directions the $\cos\theta$ is negative. In this way, the vector product in (7.7-1) produces both magnitude and sign. The product described by (7.7-1) occurs regularly in physical situations and has a formal vector mathematical definition called the *dot product*.

The vector dot product, $\vec{\mathbf{A}} \cdot \vec{\mathbf{B}}$ *(read as "A dot B") of two vectors* $\vec{\mathbf{A}}$ *and* $\vec{\mathbf{B}}$ *separated by an angle* θ *is formed by multiplying the magnitude of* $\vec{\mathbf{A}}$ *times the magnitude of* $\vec{\mathbf{B}}$ *times the cosine of the angle between them to produce a scalar quantity having the sign of* $\cos\theta$:

$$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = (\mathrm{A}_x\vec{x} + \mathrm{A}_y\vec{y} + \mathrm{A}_z\vec{z}) \cdot (\mathrm{B}_x\vec{x} + \mathrm{B}_y\vec{y} + \mathrm{B}_z\vec{z}) \tag{7.7-2}$$

in which, as previously, $\vec{x}$, $\vec{y}$, and $\vec{z}$ are unit vectors in the positive directions of the x, y, and z axes, respectively, of a right-hand coordinate system. As was true of the cross product, some of the multiplication terms described by (7.7-2) are zero. In particular, when each component of $\vec{\mathbf{A}}$ is dot multiplied by the three components of $\vec{\mathbf{B}}$, only like-directed components produce a nonzero product. Furthermore, since the A_i and B_i components (for $i = x, y, z$) are in exactly the same direction, the cosine of the angle between them is unity. Accordingly, (7.7-2) reduces to

$$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = \mathrm{A}_x\mathrm{B}_x + \mathrm{A}_y\mathrm{B}_y + \mathrm{A}_z\mathrm{B}_z \tag{7.7-3}$$

Note that, unlike the vector cross product which produces a new vector, *the dot product produces a scalar quantity*.

In terms of the dot product, Gauss's law can be stated in integral form as

$$\oint_{\mathrm{S}} \vec{\mathbf{D}} \cdot \overrightarrow{\mathrm{dS}} = \mathrm{q} \tag{7.7-4}$$

Taking the limit as the volume enclosed by S tends to zero, this equation can be used to define the *divergence of the vector* $\vec{\mathbf{D}}$.

The divergence $\nabla \cdot \vec{\mathbf{A}}$ *(pronounced "del dot* $\vec{\mathbf{A}}$*") of a vector field* $\vec{\mathbf{A}}$ *from a closed surface S is equal to the dot product of* $\vec{\mathbf{A}}$ *with a vector* $\overrightarrow{\mathrm{dS}}$ *at the surface having a magnitude equal to the surface area at the location of evaluation, a direction that is orthogonal to the surface, and a sign that is positive when the surface vector points outside the volume enclosed by the surface.*

Written in differential form this is

$$\nabla \cdot \vec{A} \equiv \lim_{\Delta V \to 0} \frac{\oint_S \vec{A} \cdot \vec{dS}}{\Delta V} \tag{7.7-5}$$

The divergence can be calculated for any vector field. Using the divergence, Gauss's law can be written in differential form as

$$\nabla \cdot \vec{D} = \rho \tag{7.7-6}$$

where ρ is the volume *charge density* at the location at which $\vec{D}$ is evaluated. If (7.7-6) is integrated over any given volume [6, p. 71]

$$\int_V \nabla \cdot \vec{D}\, dV = \int_V \rho\, dV \tag{7.7-7}$$

and then, replacing the right side by its equivalent from Gauss's law

$$\int_V \nabla \cdot \vec{D}\, dV = \int_S \vec{D} \cdot \vec{dS} \tag{7.7-8}$$

This result has been obtained from a consideration of the $\vec{D}$ field; however, it is a general result that applies to any vector $\vec{A}$, and is called the *divergence theorem* (and sometimes *Gauss's theorem*).

The divergence theorem: For any continuous vector function of space, $\vec{A}$*, the integral of the divergence of* $\vec{A}$ *over any volume completely enclosed by surface* S *is equal to the integral of* $\vec{A} \cdot \vec{dS}$ *over the whole surface* S.

$$\int_V \nabla \cdot \vec{A}\, dV = \int_S \vec{A} \cdot \vec{dS} \tag{7.7-9}$$

Like the dot product the *divergence produces a scalar result*. The divergence of a vector $\vec{A} = A_x\vec{x} + A_y\vec{y} + A_z\vec{z}$ can be evaluated using

$$\nabla \cdot \vec{A} = \frac{\partial}{\partial x}A_x + \frac{\partial}{\partial y}A_y + \frac{\partial}{\partial z}A_z \tag{7.7-10}$$

7.8 STATIC POTENTIAL FUNCTION AND THE GRADIENT

In an electrostatic region, in which charges are fixed, each coordinate point (x, y, z) has an associated *static potential energy*, $\Phi(x, y, z)$ for a point charge relative to some reference. The potential energy of a test charge in a region of electrostatic potential energy is analogous to a skier who has potential energy

for each elevated location on a mountain relative to the base of the mountain or to some other location defined as a reference at which his potential energy is defined to be zero.

The energy which must be expended to move a unit positive charge from location 1 to some different location 2 is given by [6, p. 73]

$$\Phi_2 - \Phi_1 = -\int_1^2 \vec{\mathbf{E}} \cdot \overrightarrow{\mathrm{dl}} \tag{7.8-1}$$

The negative sign occurs because a positive $\vec{\mathbf{E}}$ field produces a force that tends to move a positive charge from the higher potential to the lower one. To move in the direction opposite to the force field ($\vec{\mathbf{E}}$ field), we must do work (provide energy) rather than recover it. *Only a difference in potential is important*; the zero potential reference can be selected arbitrarily.

In an electrostatic system, having no moving charges (currents) or changing magnetic fields, the potential is the same as voltage. However, we reserve the term *voltage* for later to describe the combined effects of electrostatic charge distributions as well as changing magnetic fields.

The potential energy in a region due to the presence of electrostatic (fixed) charges and their resulting $\vec{\mathbf{E}}$ fields has a single value at every location independent of how that location is approached. Thus, it follows that the integral of $\vec{\mathbf{E}}$ about any closed path in the region must be zero. The same is true of a gravitational field around which any closed path (in which you return to the starting point) produces no change in potential energy. Similarly, there is no net energy change in moving a given charge about a closed path in an electrostatic $\vec{\mathbf{E}}$ field. This means that *the $\vec{\mathbf{E}}$ field is conservative*. Practically, there may be some energy lost to friction (resistance) in moving charge about a system, but *no difference in potential energy results in completing a closed path within a conservative field.*

Mathematically this is stated

$$\oint \vec{\mathbf{E}} \cdot \overrightarrow{\mathrm{dl}} = 0 \tag{7.8-2}$$

Suppose that we wish to determine the potential at point P due to a charge q_1 located at a point Q_1 (Fig. 7.8-1).

Setting the zero potential at infinity, the work done to move a unit test charge from infinity to the point P is equal to the work integral

$$\Phi_1 = -\int_\infty^P \vec{\mathbf{E}} \cos\theta \, \mathrm{dl} = -\int_\infty^P \frac{q_1 \, \mathrm{dr}}{4\pi\varepsilon r^2} = \frac{q_1}{4\pi\varepsilon r_1} \tag{7.8-3}$$

in which the equivalence $\cos\theta \, \mathrm{dl} = \mathrm{dr}$ has been recognized, the value for $\vec{\mathbf{E}}$ has been substituted from (7.8-1), and r_1 is the distance from point Q_1 to P. The

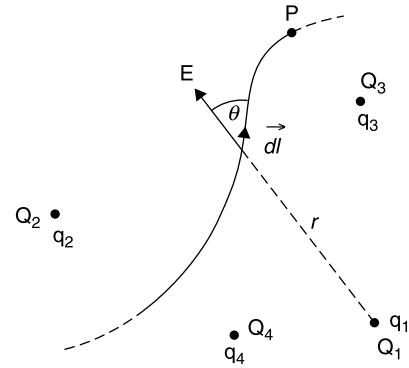


Figure 7.8-1 A work integration path to point P in a field of discrete charges q_1 at position Q_1, q_2 at Q_2, q_3 at Q_3 and q_4 at Q_4 [6, p. 72, with permission].

total potential at P due to a collection of n charges is

$$\Phi = \frac{q_1}{4\pi\varepsilon r_1} + \frac{q_2}{4\pi\varepsilon r_2} + \cdots = \sum_{i=1}^{n} \frac{q_i}{4\pi\varepsilon r_i} \tag{7.8-4}$$

If, instead of discrete charges, we have a charge density distribution,

$$\Phi = \frac{1}{4\pi} \int\!\!\int\!\!\int_V \frac{\rho\, dV}{\varepsilon_r \varepsilon_0 r} \tag{7.8-5}$$

which defines the static potential function Φ. The volume charge density ρ is a function of position. The constant of integration in (7.8-5) is used to set the zero potential point. If this point is made infinitely far away, the constant is zero.

The effect of individual charges or a charge distribution can be taken into account using Φ. This involves only scalar calculations, the advantage gained by defining the scalar potential function. From Φ we will be able to calculate $\vec{E}$ very easily. On the other hand, determining $\vec{E}$ directly requires that we sum the three directional components of the $\vec{E}$ field due to each charge source, or use a three-dimensional integration for a charge density. This direct approach would result in a much more complex calculation.

Once Φ is determined, the $\vec{E}$ field at any location is given by the *gradient* of $\Phi, \nabla\Phi$. *The gradient of a function is the rate of change of the function per unit distance in three mutually orthogonal directions.* In rectangular coordinates it is

$$\nabla\Phi = \vec{x}\frac{\partial\Phi}{\partial x} + \vec{y}\frac{\partial\Phi}{\partial y} + \vec{z}\frac{\partial\Phi}{\partial z} \tag{7.8-6}$$

$$\vec{E} = -\nabla\Phi \tag{7.8-7}$$

The gradient operates on a scalar function and produces a vector function (Fig. 7.8-2). Note that $\vec{E}$ *is the negative of the gradient of* Φ [the reason is given below

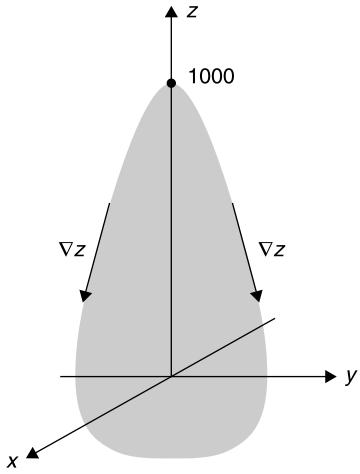


Figure 7.8-2 The ski mountain, with gradient (slope), ∇z.

(7.8-1)]. The gradient can be used with any scalar function; it is not limited to the static potential function defined in this section. For example, consider the function representing an idealized ski mountain:

$$z = 1000 - 100x^2 - 100y^2 \tag{7.8-8}$$

where x, y, and z are in meters, z is the height of the mountain, $\pm x$ are the north–south directions, and $\pm y$ are the east–west directions.

If the mountain in Figure 7.8-2 is truncated in any xy plane [$z =$ constant, the section formed is a circle $x^2 + y^2 = (1000 - z)/100$], and therefore the function of (7.8-7) has cylindrical symmetry with respect to z. As a consequence, the fall line of the mountain is the same in any direction (north, south, east, west, or any intermediate direction). At the very top, 1000 m, the slope of the mountain is zero. A skier might just balance there without being drawn downward. However, a movement of 1 m in any direction would result in a drop of 100 m, and at 1 m the slope is 200 m/m. The slope becomes ever steeper with increasing values of x and/or y, as one descends the mountain. The exact slope and its vector direction at any x, y, z point on the mountain is given by

$$\nabla z = (-200x)\vec{x} + (-200y)\vec{y} \tag{7.8-9}$$

where x and y must satisfy (7.8-7) to be points on the surface of the mountain. Notice that the gradient is independent of the height of the mountain, depending only on the slope at the location (x, y, z) on the mountain's surface.

The ski mountain example, $z = f(x, y)$, was chosen in order to depict the gradient graphically. However, a static potential function, Φ, in general would

be a function of x, y, and z. That is, $\Phi = \Phi(x, y, z)$, and the resultant gradient is three dimensional, as indicated by (7.8-6).

7.9 DIVERGENCE OF THE B FIELD

Previously, we related the electric flux density and the permittivity (dielectric constant) ε, to the electric field $\vec{\mathbf{E}}$ such that $\vec{\mathbf{D}} = \varepsilon\vec{\mathbf{E}}$. Similarly, it is useful to define the magnetic field $\vec{\mathbf{H}}$ that is related to the *magnetic flux density* $\vec{\mathbf{B}}$ by a property of the magnetic medium, the *magnetic permeability* μ. The reason for defining two variables, $\vec{\mathbf{B}}$ and $\vec{\mathbf{H}}$, is that $\vec{\mathbf{B}}$ has zero divergence, while $\vec{\mathbf{H}}$ does not. The two variables allow the "Amperian" currents (spins) of the molecules of magnetic materials to be accounted for by μ_R, and then the right-hand side of Ampere's law does not include these Amperian currents. Ampere's law is discussed in the next section. Finally, the normal component of the $\vec{\mathbf{B}}$ field does not change across a magnetic material boundary, while that of the $\vec{\mathbf{H}}$ field may:

$$\vec{\mathbf{B}} = \mu\vec{\mathbf{H}} \tag{7.9-1}$$

Thus the *electric and magnetic flux densities are* $\vec{\mathbf{D}}$ *and* $\vec{\mathbf{B}}$, respectively, and the *electric and magnetic fields are* $\vec{\mathbf{E}}$ *and* $\vec{\mathbf{H}}$, respectively. For homogeneous (uniform) and isotropic media (wherein the electric and magnetic properties are the same for all directions and locations) both ε and μ are constants.

Unlike the electric flux density $\vec{\mathbf{D}}$, which has as its vector sources discrete charges, q, there are no point sources for the magnetic flux density $\vec{\mathbf{B}}$ (Fig. 7.9-1). Stated another way, *the divergence of* $\vec{\mathbf{B}}$ *is zero:*

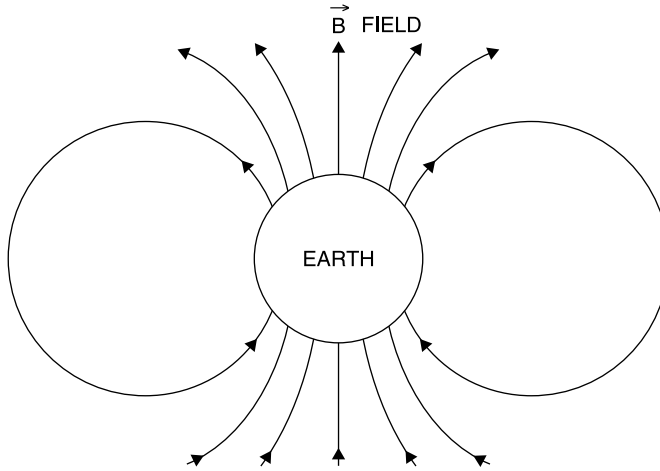


Figure 7.9-1 Divergence of the $\vec{\mathbf{B}}$ field is zero, illustrated using an idealized sketch of the Earth's magnetic field.