Handbook of Coaxial Microwave Measurements

GENERAL RADIO
WEST CONCORD, MASSACHUSETTS 01781

Handbook of Coaxial Microwave Measurements

by
David A. Gray

PRICE: $2.00

Copyright 1968 by General Radio Company

West Concord, Massachusetts, USA
Printed in USA
Form No. 3219-A

Preface

This book is not for microwave engineers; the engineering literature on the subject of microwave measurements is already ample. This book has been written for three groups of readers. First, for students, who it is hoped will find it a readable introduction to laboratory procedures. Second, for technicians who are working with coaxial instruments and components. Third, for scientists and engineers from other fields who must make microwave measurements in the course of their research.

Throughout most of the book we have assumed of the reader only a familiarity with the basic theory of alternating currents, including the representation of ac quantities by complex phasors and the elementary algebra of complex numbers. An exception is the optional Chapter 4, where we have presented some theoretical material. Even here, an acquaintance with the solutions to the one-dimensional wave equation will see the reader through quite adequately.

The efforts of many persons besides the author have gone into the creation of this handbook. Mrs. Gladys J. Carter typed the manuscript (several times), Mrs. Barbara R. Mucciaccio set the text and equations in type, and Mrs. Jane S. Putnam prepared the drawings. Layout was done by Mrs. Wilna I. Tannahill, and editorial supervision was capably performed by Miss Audrey J. Boyan. The entire handbook was read both in draft and in proof by Mr. Douglas M. Woodard of General Radio's Microwave Group. He has made an invaluable contribution to the book by ensuring the accuracy of formulas and numerical examples (but by the same token he cannot escape responsibility for any errors that remain).

D.A.G.

Table of Contents

CHAPTER 1 – INTRODUCTION TO COAXIAL TRANSMISSION LINES

CHAPTER 2 – THE SMITH CHART

CHAPTER 3 – TWO–PORTS AND DISCONTINUITIES

CHAPTER 4 – SOME THEORETICAL BACKGROUND

CHAPTER 5 – BASIC MEASUREMENT METHODS AND PROCEDURES

CHAPTER 1

Introduction to Coaxial Transmission Lines

Traveling Waves

1.1 FIELDS IN COAXIAL LINES

Although an infinite number of electromagnetic field configurations, or **modes**, as they are called, can propagate along a coaxial transmission line, the one we are almost always interested in is the **principal** or **transverse electromagnetic** (**TEM**) mode, because except in rare instances coaxial lines are intended to operate in this mode. The name "transverse electromagnetic" derives from the fact that both the electric and magnetic fields belonging to the TEM mode are entirely normal to the direction of propagation. All the higher modes have, in addition to the transverse fields, components of either the electric or magnetic field in the direction of propagation.

Not only coaxial lines but also parallel-wire lines, strip lines, in fact all transmission lines having two or more conductors, allow propagation of TEM waves. Like the coaxial line, these other multiconductor transmission lines are almost invariably intended to work in the TEM mode, although they too have higher modes of propagation. Hollow waveguides, on the other hand, are transmission lines that have just a single conductor, and they will not support TEM waves. Waveguide transmission must therefore utilize a higher mode.†

Unguided waves in an unbounded medium (that is, free electromagnetic radiation) are transverse electromagnetic and share all the properties that characterize principal mode waves on transmission lines.

Waves of any frequency, from dc upward, can propagate in the principal mode. Higher-mode waves propagate only above certain cutoff frequencies that depend on the cross section of the guiding structure and on the particular mode. The possibility of propagation in the higher modes normally limits the usefulness of a coaxial line to frequencies below the lowest higher-mode cutoff.

Figures 1.1-1 and 1.1-2 show the fields belonging to the principal mode in an ideal, lossless coaxial line. The electric field has radial lines of force which terminate on the conducting surfaces. The magnetic field is tangential; its lines of force are concentric, circular loops around the inner conductor. Both fields

---

†Properly speaking, any transmission line is a waveguide, and we should probably be talking about "coaxial waveguides." To most people, however, "waveguide" still connotes a hollow pipe, and we hope we may be forgiven for using old-fashioned terminology when we talk about coaxial "transmission lines."

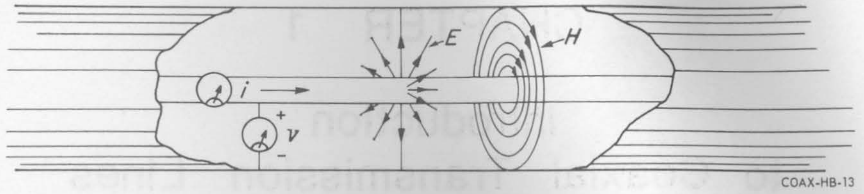


Figure 1.1-1. Electric field (E) and magnetic field (H) belonging to the principal mode in a coaxial line.

are most intense at the surface of the inner conductor and decrease in intensity inversely with increasing radius. The instantaneous magnitude of the electric field at a distance r(meters) from the axis is

$$E = \frac{v}{\log_e \frac{b}{a}} \cdot \frac{1}{r} \quad \text{(volts/meter)} \tag{1.1-1}$$

where v is the instantaneous potential difference across the line in volts and a and b are the radii of the inner and outer conductors in meters. The instantaneous magnitude of the magnetic field is

$$H = \frac{i}{2\pi} \cdot \frac{1}{r} \quad \text{(amperes/meter)} \tag{1.1-2}$$

where i is the instantaneous current in amperes.

Example: Electrical breakdown of air at a pressure of one atmosphere occurs when the electric field intensity is around 10^4 volts/cm. What is the breakdown voltage of standard 9/16-inch 50-ohm coaxial line (outer conductor ID = 0.563 inch, inner conductor OD = 0.244 inch)?

Breakdown will occur where the electric field is strongest, at the surface of the inner conductor. Therefore the breakdown voltage v(breakdown) will be given by (1.1-1) with $r = a$ and $E = E$(breakdown):

$$v(\text{breakdown}) = a \log_e \frac{b}{a} \cdot E(\text{breakdown})$$

The numbers must have the right units before they are plugged into the formula. E(breakdown) = 10^4 volts/cm = 10^4 volts/10^{-2} meter = 10^6 volts/meter. a = ½ × 0.244 inch = ½ × 0.244 × 0.0254 meter = 0.0031 meter.

$$\log_e \frac{b}{a} = \log_e \frac{\tfrac{1}{2} \times 0.563 \text{ inch}}{\tfrac{1}{2} \times 0.244 \text{ inch}} = \log_e \frac{0.563}{0.244} = \log_e 2.32 = 0.842$$

Finally, then, v(breakdown) = 2.61×10^3 volts, or about 2.5 kilovolts.

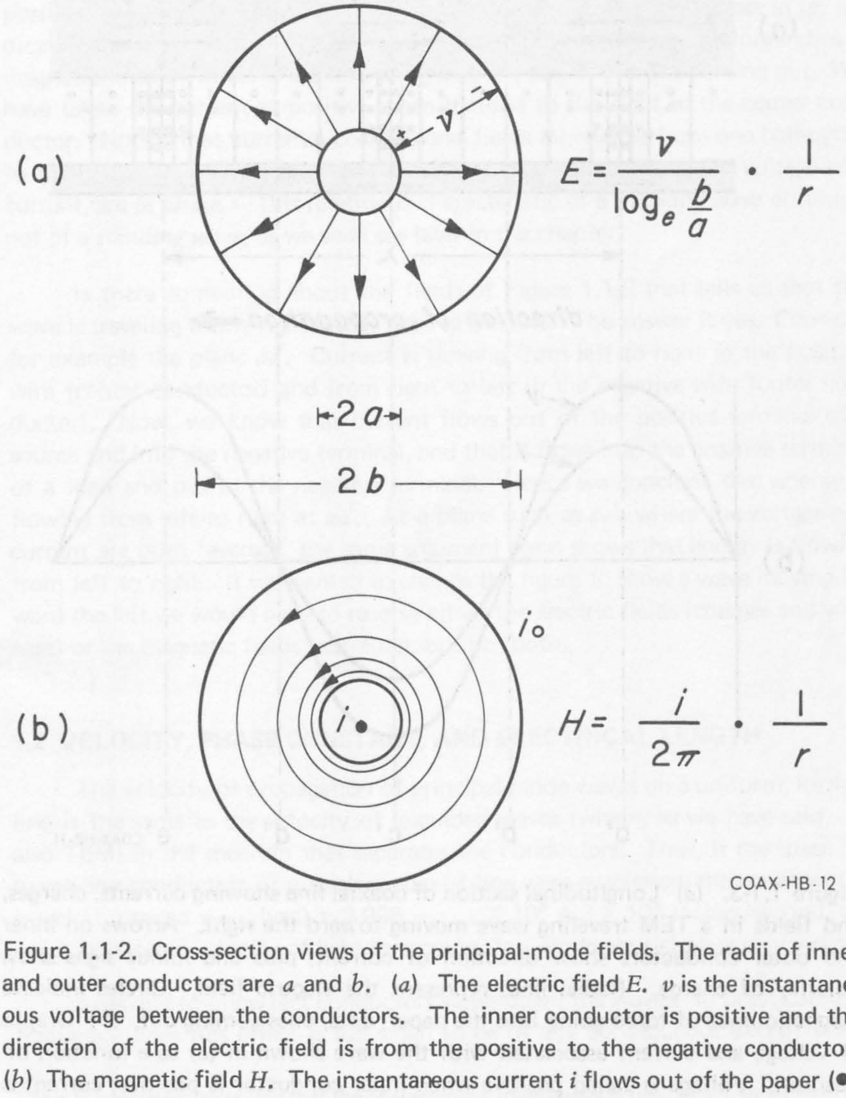


Figure 1.1-2. Cross-section views of the principal-mode fields. The radii of inner and outer conductors are a and b. (*a*) The electric field E. v is the instantaneous voltage between the conductors. The inner conductor is positive and the direction of the electric field is from the positive to the negative conductor. (*b*) The magnetic field H. The instantaneous current i flows out of the paper (●) in the inner conductor and into the paper (○) in the outer conductor.

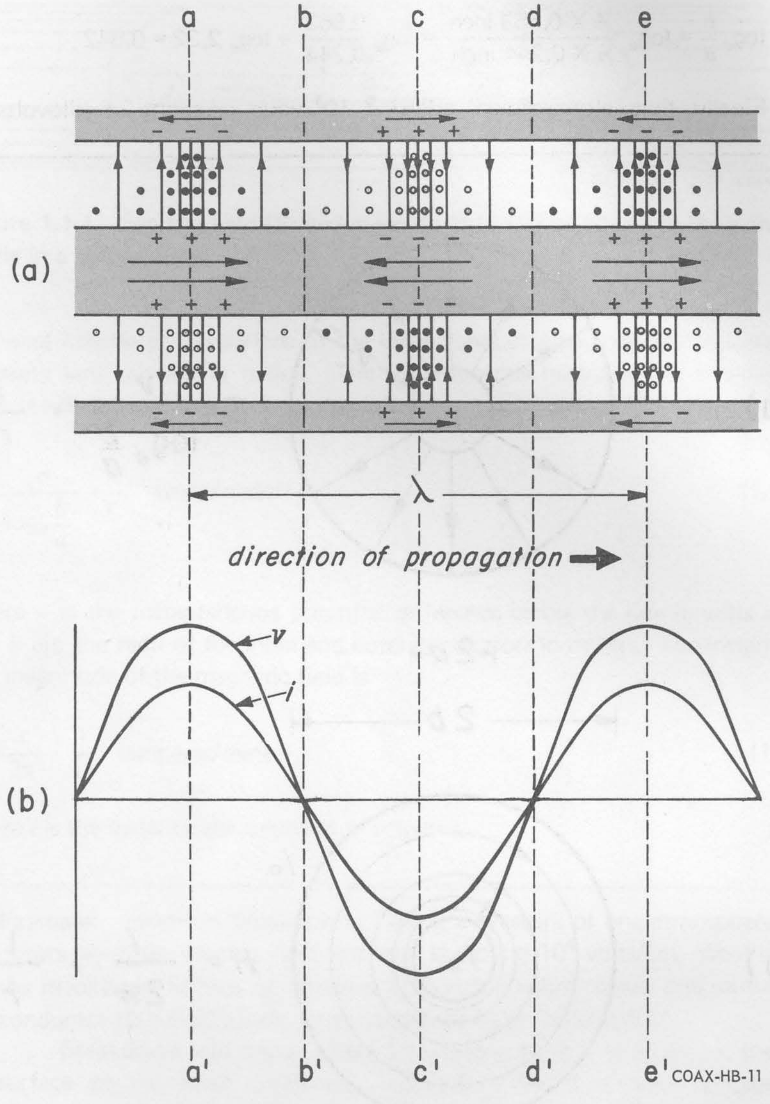


Figure 1.1-3. (*a*) Longitudinal section of coaxial line showing currents, charges, and fields in a TEM traveling wave moving toward the right. Arrows on inner and outer conductors show direction of current; plus and minus signs show polarity of charge. Radial lines represent the electric field. Circles indicate magnetic lines of force going into the paper; dots, ones coming out. (*b*) Graphs of voltage and current associated with the wave shown in (*a*) as a function of position. Voltage is called positive when inner conductor is positive; current is called positive when it flows to the right in the center conductor.

Figure 1.1-3(a) shows in longitudinal section the essential features of a sinusoidal traveling wave that is propagating toward the right along a coaxial line. Immediately below, in (b), graphs show the axial distribution of voltage and current at the same instant. The distance between two planes such as *aa'* and *ee'* that are exactly one spatial cycle apart is the wavelength, λ. The + and − signs in (a) indicate the charge on the conductors and the radial lines are electric lines of force. We have chosen to call the voltage positive when the center conductor is positive, as it is at the plane *aa'*. The arrows drawn on the conductors in (a) indicate current direction. The symbols ○ and ● between the conductors indicate magnetic lines of force; ○ is a line going into the paper, ● is one coming out. We have taken the current as positive when it flows to the right in the center conductor. Notice that currents, charges, and fields all reverse from one half-cycle to the next. Notice too that the electric and magnetic fields, hence voltage and current, are *in phase.*† This relation is characteristic of a *traveling wave* although not of a *standing wave,* as we shall see later in the chapter.

Is there something about the fields of Figure 1.1-3 that tells us that the wave is traveling to the right rather than to the left? The answer is yes. Consider for example the plane *aa'*. Current is flowing from left to right in the positive wire (center conductor) and from right to left in the negative wire (outer conductor). Now, we know that current flows out of the positive terminal of a source and into the negative terminal, and that it flows into the positive terminal of a load and out of the negative terminal. Hence we conclude that energy is flowing from left to right at *aa'*. At a plane such as *cc'*, where the voltage and current are both reversed, the same argument again shows that energy is flowing from left to right. If we wanted to change the figure to show a wave moving toward the left we would have to reverse either the electric fields (charges and voltages) or the magnetic fields (currents), but not both.

1.2 VELOCITY, PHASE CONSTANT, AND ELECTRICAL LENGTH

The velocity of propagation of principal-mode waves on a uniform, lossless line is the same as the velocity of unguided waves (which, as we have said, are also TEM) in the medium that separates the conductors. Thus, if the space between the conductors of a lossless coaxial line were evacuated, the waves would travel at a speed $v_{\text{TEM}}(\text{vac}) = 2.998 - \ldots \times 10^8 \doteq 3 \times 10^8$ meters/second, the much publicized velocity of light in vacuum, for which physicists usually write *c*.

Loss due to imperfect conductors slows down the waves. In practical high-frequency lines this effect is too small to be of any consequence except under

---

† This is equivalent to saying that the characteristic immittance is real, which, as we shall see in Section 1.3, is not quite true of a lossy line.

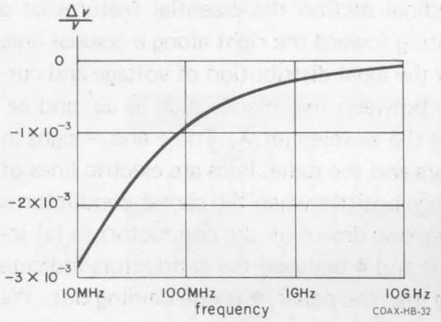


Figure 1.2-1. The effect of conductor loss on the velocity of propagation in General Radio 9/16-inch 50-ohm precision silver air-dielectric line.

(After Zorzy, IEEE Transactions on Instrumentation and Measurement, Vol IM-15, No. 4, December, 1966.)

circumstances of the most exacting precision. The relative decrement $\Delta v/v$ that conductor loss causes in the velocity of TEM waves in General Radio 9/16-inch 50-ohm precision silver air-dielectric line is shown as a function of frequency in Figure 1.2-1.

In an ordinary dielectric the speed v_{TEM}(diel) of TEM waves is less than v_{TEM}(vac). Physicists call the ratio

$$\frac{v_{\mathrm{TEM}}\ \text{(vac)}}{v_{\mathrm{TEM}}\ \text{(diel)}} = n \tag{1.2-1}$$

(which is a dimensionless number greater than unity) the **index of refraction** of the particular material because it is the difference in the velocity of light in two media that causes refraction at an interface. Engineers often describe the reduction in the velocity of waves in a cable due to the presence of a dielectric between the conductors in terms of a number called the **velocity factor**, which is just the reciprocal of the index of refraction.

$$\text{velocity factor} = \frac{v_{\mathrm{TEM}}\ \text{(diel)}}{v_{\mathrm{TEM}}\ \text{(vac)}} < 1 \tag{1.2-2}$$

Example: When we look up the optical index of refraction of polyethylene we find figures that are close to 1.5. Now, the reciprocal of 1.5 is 0.67, which as a matter of fact is a typical velocity factor for a cable filled with solid polyethylene. Actually this agreement is little more than coincidental. One would be naive to expect the velocity of electromagnetic waves in any material medium to be the same at microwave frequencies as it is at optical frequencies, five or six orders of magnitude higher.

The velocity of TEM waves is determined by two properties of the medium according to the relation

$$\nu = \frac{1}{\sqrt{\mu\epsilon}} \qquad \text{(meters/second)} \qquad (1.2\text{-}3)$$

where μ is the medium's **magnetic permeability** (henrys/meter) and ϵ is its **electric permittivity** (farads/meter).[†] The permeability of vacuum, μ(vac), is a defined number exactly equal to $4\pi \times 10^{-7}$ henry/meter. The permittivity of vacuum, ϵ(vac), is an experimental number equal to $8.85 + \ldots \times 10^{-12}$ farad/meter.

In dielectric media the permeability has its vacuum value but ϵ(diel) is always larger than ϵ(vac), often many times larger. The dimensionless ratio ϵ(diel)/ϵ(vac) is called the **relative permittivity** or **dielectric constant** of the material in question and is represented variously by ϵ_r, κ, k, K, and, regrettably, quite often by ϵ.

$$\epsilon_r = \frac{\epsilon(\text{diel})}{\epsilon(\text{vac})} \qquad (1.2\text{-}4)$$

Since the velocity of TEM waves depends inversely on the square root of ϵ, the velocity in a dielectric medium may be written

$$\nu_{\text{TEM}}(\text{diel}) = \frac{\nu_{\text{TEM}}\ (\text{vac})}{\sqrt{\epsilon_r}} \qquad (1.2\text{-}5)$$

Example: The dielectric constant of dry air at one atmosphere and 23 degrees Celsius is 1.00068. What is the velocity factor of an air-dielectric coaxial line?

If we compare (1.2-2) with (1.2-5) we see that the velocity factor is equal to $1/\sqrt{\epsilon_r}$. One can find the square root of a number that is very close to unity simply by taking the first term of a binomial series: $(1 + x)^{\pm 1/2} \doteq 1 \pm \frac{1}{2} x$. Thus the velocity factor of an air-dielectric line is $1 - \frac{1}{2}(0.00068) = 0.99966$, which is so close to unity that the difference between air and vacuum can almost always be ignored.

$$v_{\text{TEM}}\ (\text{air}) \doteq v_{\text{TEM}}\ (\text{vac})$$

[†] The basic system of units used by electrical engineers is the meter-kilogram-second-ampere (mksA) system. The practical electrical units – volts, amperes, watts, farads, ohms, etc – belong to the mksA system. The reader should be wary of two things: First, physicists and chemists frequently continue to use the older Gaussian electrical units, a centimeter-gram-second (cgs) system based on the electrostatic unit (esu) of charge and the electromagnetic unit (emu) of current. Formulas in the Gaussian system have different constants and quantities have different sizes and different units. Second, in practice nobody bothers to stick to a single system anyway. Thus in this book we shall use centimeters and inches as well as meters, degrees and decibels as well as radians and nepers, and so forth.

Example: What is the velocity factor of a flexible cable filled with polyethylene, $\epsilon_r = 2.25$?

velocity factor = $1/\sqrt{2.25} = 0.67$

We have already pointed out that v_{TEM}(diel) varies with frequency, and so, therefore, must the dielectric "constant" ϵ_r. We shall say no more about this except to remark that variations in ϵ_r with frequency are accompanied by high dielectric loss, and low-loss dielectrics such as polyethylene, polystyrene, and Teflon[†] have relative permittivities that are constant with frequency.

The **wavelength** λ of a periodic wave is related to its frequency f and velocity v by the well-known formula

$$\lambda f = v \tag{1.2-6}$$

Since the velocity of TEM waves is the velocity in vacuum divided by $\sqrt{\epsilon_r}$, the wavelength of principal-mode waves on a coaxial line is given by

$$\lambda_{\mathrm{TEM}} = \frac{v_{\mathrm{TEM}}(\mathrm{vac})}{f\sqrt{\epsilon_r}} = \frac{\lambda_{\mathrm{TEM}}(\mathrm{vac})}{\sqrt{\epsilon_r}} \tag{1.2-7}$$

Thus the wavelength in an air-dielectric coaxial line is the same as the free-space wavelength, while that in a solid dielectric line is shorter by the factor $1/\sqrt{\epsilon_r}$. (But this is not true of waves in hollow waveguides, which are not transverse electromagnetic. The phase velocity of non-TEM waves is greater than that of TEM waves, and it depends on the frequency. Therefore the guide wavelength is longer than the free-space wavelength and is not simply proportional to $1/f$.)

The **phase factor** or **phase constant** β tells how rapidly the phase of a sinusoidal traveling wave changes with distance along the line. If we imagine the traveling wave "frozen" at a particular instant of time, β is the amount of phase change per unit distance. Since the phase changes by 2π radians or 360 degrees in one wavelength, we have

$$\left.\begin{aligned} \beta(\text{radians/meter}) &= \frac{2\pi}{\lambda} \\ \beta(\text{degrees/meter}) &= \frac{360}{\lambda} \end{aligned}\right\} \tag{1.2-8}$$

Example: What is the phase constant of waves in a flexible cable whose velocity factor is 0.67 if the frequency is 300 MHz?

[†] Registered trademark of E.I. duPont de Nemours and Company.

The velocity in the cable is 0.67 × 3 × 10^8 meters/second = 2.0 × 10^8 meters/second, so that the wavelength λ is

$$\frac{2.0 \times 10^8 \text{ meters/second}}{3.0 \times 10^8 \text{ second}^{-1}} = 0.67 \text{ meter}$$

and the phase constant β is 360 degrees/0.67 meter = 540 degrees/meter.

We assume that the reader is familiar with the representation of sinusoidally time-varying quantities by phasors or, as they are often called (incorrectly, from the mathematician's point of view), vectors in the complex plane. We will use upper-case V's and I's to denote phasor voltages and currents. Thus an instantaneous voltage $v(t)$ that varies with time according to

$$v(t) = |V|\cos(2\pi ft + \phi) \tag{1.2-9}$$

where $|V|$ is the peak value of $v(t)$, f is the frequency, and ϕ is the phase, will be represented by the phasor

$$V = |V| \measuredangle \phi \tag{1.2-10}$$

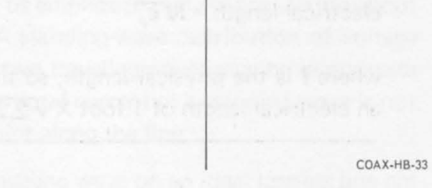


Figure 1.2-2. The phasor $V = |V| \measuredangle \phi$.

The magnitude $|V|$ of the phasor V is equal to the peak value of the time-varying quantity $v(t)$, and the angle ϕ is equal to the phase angle of $v(t)$. The important thing to notice is what happens if we change ϕ. We can see from (1.2-9) that increasing ϕ has the same effect as decreasing t, so that after we have increased ϕ, $v(t)$ will reach any particular value in its cycle at a smaller t, that is, earlier than it did before. *Increasing the phase angle* ϕ, which by convention means rotating the phasor V *counterclockwise*, makes $v(t)$ occur *earlier*.

If V is the phasor that represents the instantaneous voltage $v(t)$ due to a traveling wave on a transmission line, the angle ϕ of V will be found to increase as V is observed at points closer and closer to the source of the wave. This is because the time at which $v(t)$ reaches a particular angle in its cycle becomes progressively earlier at points closer and closer to the source. The rate at which ϕ changes with distance is the phase constant β.

$$\left.\begin{array}{l} \begin{array}{l}\text{phase shift of}\\ \text{traveling wave}\\ \text{in line segment}\\ \text{of length } l\end{array} \quad = \quad \pm\beta l \\ \\ \left.\begin{array}{l} +\ \text{toward}\\ -\ \text{away from}\end{array}\right\}\ \text{source of wave}\end{array}\right\} \qquad (1.2\text{-}11)$$

The terms **electrical length** and **electrical distance** are used in two really quite different senses. One meaning, which applies to a device or a component of a transmission system, is the length of air-dielectric line that has the same delay time as the device in question. Electrical lengths in this sense are measured in units of length: inches, centimeters, etc. The electrical length of a connector with a solid dielectric support bead, for example, will be longer than its physical length because the waves propagate more slowly in the solid dielectric than they do in air.

Example: What is the electrical length, in the sense just defined, of a foot of cable whose dielectric is solid polyethylene ($\epsilon_r = 2.25$)?

From equation 1.2-5 we see that

$$\text{electrical length} = l\sqrt{\epsilon_r} \qquad (1.2\text{-}12)$$

where l is the physical length, so that a foot of the cable in question has an electrical length of 1 foot $\times \sqrt{2.25} = 1.5$ feet.

The second and more common use of "electrical length" or "electrical distance" is to refer to the phase difference βl between two points on a transmission line. Thus one speaks of a section of line that is $\pi/4$ radian or 45 degrees in electrical "length."

Example: A simple way to measure the velocity v of propagation in a cable (at moderate frequencies) is to short both ends of a length of the cable and then to measure two or more resonant frequencies of the shorted line with a wavemeter or loosely coupled generator and indicator. Resonance will occur when $2\beta l$, the electrical round-trip "distance" (that is, phase shift) down the shorted cable and back again, is a multiple of 360 degrees. Thus the resonant frequencies f_{res} will be given by

$$f_{\text{res}} = n \cdot \frac{v}{2l} \quad (n = 1, 2, \ldots)$$

If Δf is the difference between two adjacent resonant frequencies,

$$v = 2l\Delta f \tag{1.2-13}$$

1.3 CHARACTERISTIC IMMITTANCE†

The ratio of voltage V to current I in a traveling wave is a constant, a property of the transmission line called the **characteristic impedance,** Z_c.

$$Z_c = \left(\frac{V}{I}\right)_{\text{traveling wave}} \quad \text{(ohms)} \tag{1.3-1}$$

Its reciprocal is called the **characteristic admittance.**

$$Y_c = \left(\frac{I}{V}\right)_{\text{traveling wave}} \quad (\text{ohms}^{-1}) \tag{1.3-2}$$

So that there is no misunderstanding, let us emphasize that we are talking about a *traveling wave,* not a *standing wave.* A standing-wave distribution of voltage and current is due to the superposition of two traveling waves moving in opposite directions, and the ratio of *total* voltage to *total* current in a standing wave is not constant at all but varies from point to point along the line.

The voltage and current due to a traveling wave on an ideal lossless line are exactly in phase, a fact that we remarked upon in Section 1.1. Thus the characteristic impedance and admittance of such an ideal line—and as a matter of fact for nearly all practical purposes the characteristic impedance and admittance of actual lines as well—are real numbers. One might therefore have preferred to call them characteristic resistance and conductance. The characteristic imped-

---

†**The term "immittance" means "impedance" and/or "admittance."**

ance of a lossless coaxial line with perfectly smooth conducting surfaces is given by

$$\begin{aligned} Z_c &= \frac{1}{2\pi}\sqrt{\frac{\mu(\text{vac})}{\epsilon(\text{vac})}} \cdot \frac{1}{\sqrt{\epsilon_r}} \log_e \frac{b}{a} \\ &= (59.950 \text{ ohms}) \frac{1}{\sqrt{\epsilon_r}} \log_e \frac{b}{a} \\ &= (138.03 \text{ ohms}) \frac{1}{\sqrt{\epsilon_r}} \log_{10} \frac{b}{a} \qquad (1.3\text{-}3) \end{aligned}$$

where ϵ_r is the relative permittivity of the dielectric, and a and b are the radii of the inner and outer conducting surfaces, as in Figure 1.1-2.

Notice that the dimensions of the line enter into equation 1.3-3 only through the ratio b/a, so that the over-all size of the line has nothing to do with Z_c. Fifty-ohm, rigid, air-dielectric coaxial line is manufactured in standard sizes from 7 millimeters to 9 inches in diameter. With a given outer conductor ID, the smaller the inner conductor, the higher the characteristic impedance. Type RG-8A/U and Type RG-11A/U flexible cables, for example, both have a nominal diameter, measured inside the braided copper outer conductor, of 0.284 inch. The inner conductor of the Type 8A/U, which is a 50-ohm cable, consists of seven strands of 0.0206-inch copper wire, whereas that of the Type 11A/U, a 75-ohm cable, consists of seven strands of 0.0159-inch wire.

The appearance of the factor $1/\sqrt{\epsilon_r}$ in (1.3-3) shows that the presence of dielectric material between the conductors lowers the characteristic impedance. The decrease in Z_c is in the same ratio as the decrease in the velocity of propagation.

$$Z_c(\text{solid dielectric}) = \text{velocity factor} \times Z_c(\text{air dielectric}) \qquad (1.3\text{-}4)$$

Equation 1.3-3 is derived under the assumption of an ideal, lossless line, whereas in fact losses and imperfections in the conducting surfaces do influence the characteristic impedance. At high frequencies these effects are very small in lines with solid, smooth conducting surfaces,[†] but they nevertheless can be significant, for example in a precision air-dielectric line that is used as a standard of impedance. Although a quantitative discussion of conductor loss must wait until Chapter 4, this seems like an appropriate place to describe, in a physical way at

---

[†] **Although at low frequencies the influence of finite conductivity on transmission-line characteristic impedances is appreciable; telephone lines, for example, have characteristic impedances with sizeable imaginary components at voice and carrier frequencies.**

least, the effect that imperfect conductors have on the flow of current and the way in which they influence the line's characteristic impedance.

Electromagnetic fields are rapidly attenuated in conducting media, and consequently they penetrate only very small distances into conductors. In a perfect conductor the field would not penetrate at all, and the current that forms the boundary of the field would flow in a surface layer of zero thickness. The attenuation of the field beneath the surface of a real conductor depends on the conductivity of the metal, the frequency, and the geometry of the surface; but at frequencies higher than a few kilohertz, the attenuation in a good conductor is very rapid and the current distribution below the surface can be treated as though it were a uniform layer of very small thickness δ that is virtually independent of surface geometry. In the case of a flat, perfectly smooth, non-ferromagnetic metal surface, the distance δ, called the **skin depth**, is related to the frequency f (hertz) and conductivity σ (ohm^{-1} meter^{-1}) by

$$\delta = \left(503.3 \, \frac{\text{meters}^{1/2}}{\text{henrys}^{1/2}}\right) \cdot \frac{1}{\sqrt{f\sigma}} \quad \text{(meters)} \qquad \frac{1}{\sqrt{f}} \quad \frac{1}{\sqrt{\sigma}} \qquad (1.3\text{-}5)$$

Notice that *larger* skin depths occur with *lower* frequencies and *poorer* conductivities. In copper plate, whose direct-current conductivity is approximately 6 X 10^7 ohm^{-1}meter^{-1}, (1.3-5) gives skin depths of about

8 mm at 60 Hz
0.7 mm at 10 kHz 100^2 .07 1 MHz
0.02 mm at 10 MHz
0.0007 mm at 10 GHz $10{,}000^2$

In a coaxial line with perfect conductors the currents would flow only in infinitely thin layers on the conducting surfaces and the field would stay in the dielectric space between the conductors. But when the conductivities are finite the current flow extends somewhat below the metal surfaces and the field penetrates a little into the metal. One effect of the field penetration is that the magnitude of Z_c is slightly higher than the value that (1.3-3) gives for an ideal line, somewhat as though the conductor separation had increased. Less easy to explain on simple physical grounds is the fact that conductor loss causes a slight phase lag of the electric field behind the magnetic field. This gives rise to a small negative imaginary (capacitive) component in Z_c. If the conductor surfaces are compact and smooth by comparison with dimensions on the order of the skin depth, the real and negative imaginary components of the increment in Z_c are equal.

An idea of the size of the effect we are talking about can be gained from Figure 1.3-1, which shows the increment in Z_c due to conductor loss as a func-

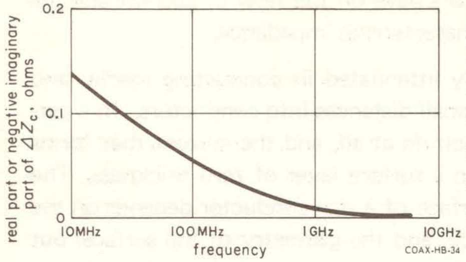


Figure 1.3-1. The effect of conductor loss on the characteristic impedance of General Radio 9/16-inch precision 50-ohm silver air-dielectric line. The increment in Z_c is complex; the real and negative imaginary components of ΔZ_c are equal. *(After Zorzy, loc. cit. Figure 1.2-1.)*

tion of frequency for General Radio 9/16-inch precision 50-ohm silver air-dielectric line.

1.4 ATTENUATION

Losses and gains, when unqualified by the words "voltage" or "current," are comparisons in amounts of *power.* What we call a loss and what we call a gain is just a matter of which way around we want to express the comparison.

$$\text{loss ratio, } P_2 \text{ re } P_1 = \frac{P_1}{P_2} \tag{1.4-1}$$

$$\text{gain ratio, } P_2 \text{ re } P_1 = \frac{P_2}{P_1} \tag{1.4-2}$$

If for example P_1 and P_2 are the powers at the input and output, respectively, of an attenuator, so that $P_1 > P_2$, the loss ratio is greater than unity and the gain ratio is less than unity.

Losses and gains are usually measured by the logarithms of the loss and gain ratios, rather than by these ratios themselves. The **neper** is a unit of loss and gain based on the natural, or Napierian, logarithm, and the **decibel** is a unit based on the common, or Briggsian, logarithm.

$$\text{loss (nepers), } P_2 \text{ re } P_1 = \frac{1}{2} \log_e \frac{P_1}{P_2} \tag{1.4-3}$$

$$\text{loss (dB), } P_2 \text{ re } P_1 = 10 \log_{10} \frac{P_1}{P_2} \tag{1.4-4}$$

Since $\log (1/x) = -\log x$, a gain in nepers or decibels is the negative of the corresponding loss, and vice versa. We can convert nepers and decibels by noting that $\log_{10} x = 0.43429 \log_e x$, so that

$$\left.\begin{aligned} &\text{loss or gain (dB)} = 8.686 \times \text{loss or gain (nepers)} \\ \text{and}\quad &\\ &\text{loss or gain (nepers)} = 0.1151 \times \text{loss or gain (dB)} \end{aligned}\right\} \qquad (1.4\text{-}5)$$

The **neper** is a unit 8.686 times larger than the **decibel**. Because of the natural occurrence of powers of e and Napierian logarithms in transmission theory, nepers are usually used in theoretical work, while the decibel is the practical and laboratory unit.† Nepers thus bear somewhat the same relation to decibels that radians bear to degrees.

Example: What is the number of decibels corresponding to a power ratio of 0.9987?

Ratios that are very close to unity are usually most convenient to deal with when they are expressed as unity plus or minus a very small number. Thus $0.9987 = 1 - 0.0013$. When a ratio is expressed in the form $1 \pm x$, the corresponding number of decibels is equal to $\pm 4.343\,x$, provided x is small.

$$\begin{matrix}\text{number of decibels} \\ \text{in ratio } 1 \pm x\end{matrix} \doteq \pm\, 4.343x \qquad (1.4\text{-}6)$$

Thus the ratio 0.9987 is equivalent to $-4.343 \times 0.0013 = -0.0056$ dB.

We can calculate the ratio of two powers from the corresponding voltage or current ratios provided we also take into account the immittances. Expressed as a function of voltage, the power dissipated in an immittance is $\frac{1}{2}G|V|^2$, where G is the real (conductive) part of the admittance Y. In terms of current, power is equal to $\frac{1}{2}R|I|^2$, where R is the real (resistive) part of the impedance Z. Thus

$$\frac{P_1}{P_2} = \left(\frac{|V_1|}{|V_2|}\right)^2 \frac{G_1}{G_2} = \left(\frac{|I_1|}{|I_2|}\right)^2 \frac{R_1}{R_2} \qquad (1.4\text{-}7)$$

and the number of nepers or decibels in the power ratio P_1/P_2 is given in terms of voltage and current ratios by the formulas

$$\begin{matrix}\text{number of}\left\{\begin{matrix}\text{nepers}\\ \text{decibels}\end{matrix}\right\} \\ \text{in power ratio } P_1/P_2\end{matrix} = \left\{\begin{matrix}\log_e \\ 20\log_{10}\end{matrix}\right\} \frac{|V_1|}{|V_2|} + \left\{\begin{matrix}\frac{1}{2}\log_e \\ 10\log_{10}\end{matrix}\right\} \frac{G_1}{G_2}$$

$$= \left\{\begin{matrix}\log_e \\ 20\log_{10}\end{matrix}\right\} \frac{|I_1|}{|I_2|} + \left\{\begin{matrix}\frac{1}{2}\log_e \\ 10\log_{10}\end{matrix}\right\} \frac{R_1}{R_2} \qquad (1.4\text{-}8)$$

---

† Although the neper is used as a practical unit in the German telephone industry.

One can often ignore the second terms in these formulas, for if $G_1 = G_2$, the conductance term disappears from the first expression on the right of (1.4-8), and if $R_1 = R_2$ the resistance term disappears from the second expression.

$$\begin{array}{l}\text{number of}\left\{\begin{matrix}\text{nepers}\\ \text{decibels}\end{matrix}\right\}\\ \text{in power ratio } P_1/P_2\end{array} = \left\{\begin{matrix}\log_e\\ 20\log_{10}\end{matrix}\right\}\frac{|V_1|}{|V_2|} \qquad (\text{if } G_1 = G_2) \qquad (1.4\text{-}9)$$

$$= \left\{\begin{matrix}\log_e\\ 20\log_{10}\end{matrix}\right\}\frac{|I_1|}{|I_2|} \qquad (\text{if } R_1 = R_2) \qquad (1.4\text{-}10)$$

Textbooks often state as the condition for the validity of (1.4-9) and (1.4-10) that the impedances must be equal. This is incorrect. If what is meant is that $Z_1 = Z_2$, we can see that it is not necessary. If $|Z_1| = |Z_2|$ is meant, it is clearly neither necessary nor sufficient.

The decibel is also used to express voltage and current ratios without regard for the amounts of power involved. Thus

$$\begin{array}{l}\text{number of decibels}\\ \text{in voltage ratio } V_1/V_2\end{array} = 20\log_{10}\frac{|V_1|}{|V_2|} \qquad (1.4\text{-}11)$$

and likewise for currents. Standing-wave ratios, for example, are commonly expressed in decibels.

Attenuation, applied to a transmission line, means the decrease in traveling-wave power in the direction of the wave's propagation. If a section of line has an attenuation expressed in decibels of A(dB), or in nepers of A(nep), the ratio of traveling-wave power leaving to traveling-wave power entering the section, which has to be a number less than unity, is

$$\frac{\begin{array}{l}\text{traveling-wave}\\ \text{power leaving}\end{array}}{\begin{array}{l}\text{traveling-wave}\\ \text{power entering}\end{array}} = 10^{-\frac{1}{10}A(\mathrm{dB})} = e^{-2A(\mathrm{nep})} \qquad (1.4\text{-}12)$$

The corresponding voltage or current ratio is equal to the square root of the power ratio:

$$\frac{\begin{array}{l}\text{traveling-wave}\\ V\text{ (or } I\text{) leaving}\end{array}}{\begin{array}{l}\text{traveling-wave}\\ V\text{ (or } I\text{) entering}\end{array}} = 10^{-\frac{1}{20}A(\mathrm{dB})} = e^{-A(\mathrm{nep})} \qquad (1.4\text{-}13)$$

The **attenuation constant** or **attenuation factor,** α, is the attenuation per unit length of line. Thus a length l(meters) of line has an attenuation of A(dB) = α(dB/m)l or A(nep) = α(nep/m)l. In practice one often finds the attenuation of lines and cables given in decibels per foot or per 100 feet or per mile.

Two kinds of loss are responsible for the attenuation in coaxial lines: loss due to the finite conductivity of conductors and loss due to dielectric relaxation — friction experienced by the alternating polarization in the dielectric.

Conductor loss depends of course on the metal from which the conductors are made or with which they are plated, but it also depends on the frequency, because of the frequency-dependent skin depth. The part of the attenuation constant due to conductor loss, α_{cond}, in an otherwise ideal coaxial line is given by

$$\alpha_{\text{cond}} = Y_{\text{c}} \sqrt{\frac{\mu(\text{vac})}{16\pi}} \sqrt{f} \left(\frac{1}{a\sqrt{\sigma(a)}} + \frac{1}{b\sqrt{\sigma(b)}} \right) \quad \frac{\text{nepers}}{\text{meter}} \qquad (1.4\text{-}14)$$

where Y_{c} is the characteristic admittance, a and b are the radii of the inner and outer conducting surfaces, and $\sigma(a)$ and $\sigma(b)$ are the conductivities of the inner and outer conductors. The formula shows that α_{cond} increases with the square root of the frequency. The first and second terms within the parentheses in (1.4-14) are associated respectively with the inner and outer conductors, and, as one would expect, the first term is likely to be the larger. Notice that small lines have higher conductor loss than larger lines with the same Y_{c}.

Equation 1.4-14 accurately describes the conductor loss in a real coaxial line if suitable values are used for the conductivities $\sigma(a)$ and $\sigma(b)$. Such values are sometimes considerably lower than the dc conductivities of the conductor metals, an effect that is presumably due to the condition of the surface, since the effective conductivity of a rough or porous surface is found to be lower than that of a smooth, compact one and, furthermore, is found to decrease with rising frequency and concomitantly decreasing skin depth. A few examples are given in Table 1.4-1.

The attenuation constant of General Radio 9/16-inch 50-ohm precision silver air-dielectric line is shown as a function of frequency in Figure 1.4-1. The attenuation in air-dielectric lines is due entirely to conductor loss.

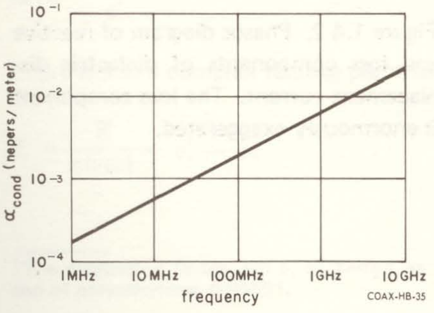


Figure 1.4-1. The attenuation constant of General Radio 9/16-inch 50-ohm precision silver air-dielectric line.
(After Zorzy, loc. cit. Figure 1.2-1.)

TABLE 1.4-1

Surface	Frequency	Conductivity
copper plate	dc	5.9 X 10^7 ohm^{-1} m^{-1}
	2.5 GHz	2
polished brass	dc	1.3
	2.5 GHz	1.3
	8.5 GHz	1.3
silver plate		
dc plated	dc	6.1
	2.5 GHz	5.3
	8.5 GHz	3.1
dc plated with a commercial brightener	2.5 GHz	1.0
	8.5 GHz	0.8
plated with periodically reversed current	2.5 GHz	6.0
	8.5 GHz	6.0

Let us turn to dielectric loss. In an ideal capacitor the dielectric displacement current leads the voltage, and therefore the electric field, by exactly 90 degrees and no power is dissipated. If, on the other hand, the dielectric is lossy, displacement current leads the electric field by less than 90 degrees. This is shown in the phasor diagram of Figure 1.4-2, in which the **loss angle** δ by which the phase difference falls short of 90 degrees is enormously larger than it would

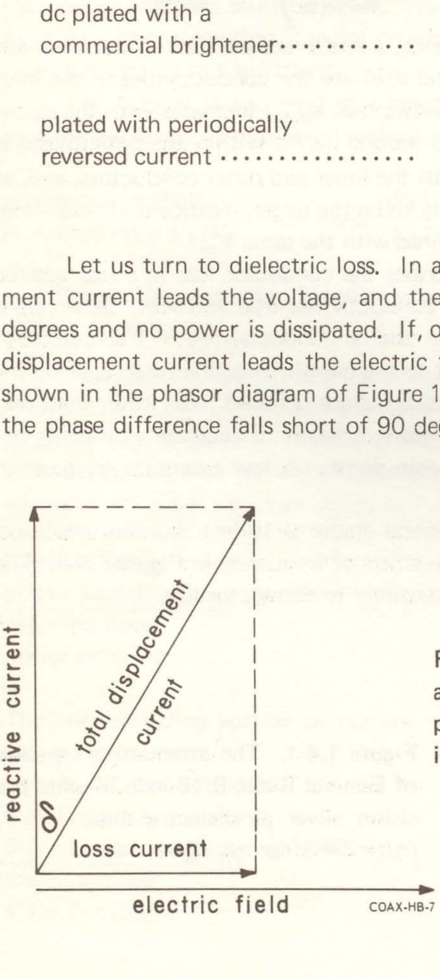


Figure 1.4-2. Phasor diagram of reactive and loss components of dielectric displacement current. The loss component is enormously exaggerated.

be in any practical dielectric.[†] The out-of-phase component of current, or reactive current, dissipates no power; it is associated with energy stored in the dielectric. The in-phase component, or loss current, is associated with dissipation due to dielectric relaxation, a friction-like drag on the dielectric's alternating polarization.

The size of the loss current relative to the reactive current is a measure of the lossiness of the dielectric. Various numbers are used to express this comparison. One of them is the **loss angle** δ, defined in Figure 1.4-2; here are several others:

$$\text{dissipation factor } (D) \text{ or loss tangent} = \tan\delta = \frac{\text{loss current}}{\text{reactive current}} \tag{1.4-15}$$

$$\text{loss factor} = \epsilon_r \tan\delta \tag{1.4-16}$$

$$\text{dielectric } Q = \frac{1}{\text{loss tangent}} = \cot\delta \tag{1.4-17}$$

$$\text{dielectric power factor} = \frac{\text{loss current}}{\text{total displacement current}} = \sin\delta \tag{1.4-18}$$

In a good dielectric the loss current is very small and the total displacement current is practically equal to its reactive component. In this case the power factor and loss tangent are virtually equal.

In theoretical work, use is often made of a **complex permittivity**, $\widetilde{\epsilon} = \epsilon' - j\epsilon''$. The real part ϵ' is the ordinary permittivity and accounts for the reactive component of displacement current. The imaginary part ϵ'' accounts for the loss current; it is a positive number that is very much smaller than ϵ' in a good dielectric. The loss angle δ is minus the angle of the complex number $\widetilde{\epsilon}$:

$$\tan\delta = \frac{\epsilon''}{\epsilon'} \tag{1.4-19}$$

There is also a **complex relative permittivity**, $\widetilde{\epsilon}_r$, defined by

$$\widetilde{\epsilon}_r = \frac{\widetilde{\epsilon}}{\epsilon\,(\text{vac})} = \epsilon'_r - j\epsilon''_r \tag{1.4-20}$$

[†] **The dissipation factor, tan δ, of Teflon at 3 GHz and 22 degrees centigrade is 0.00015, and of polyethylene, 0.00031.**

The real and imaginary parts of $\tilde{\epsilon}_r$ are just the corresponding parts of $\tilde{\epsilon}$, divided by ϵ(vac). The loss factor is minus the imaginary part of the complex relative permittivity:

$$\text{loss factor} = \epsilon_r'' \qquad (1.4\text{-}21)$$

The attenuation in a coaxial line due to dielectric loss is

$$\alpha_{\text{diel}} = \pi \, v_{\text{TEM}}\,(\text{diel})\, f \tan\delta \qquad (\text{nepers/meter}) \qquad (1.4\text{-}22)$$

When we compare this formula with (1.4-14) for the conductor loss we notice two differences. First, α_{diel} does not depend at all on the dimensions of the line. Second, α_{diel} increases proportionally with frequency rather than with the square root of frequency, as α_{cond} does. At frequencies below 10 GHz, losses are due mostly to the conductors rather than the dielectric, even in solid dielectric cables.

1.5 DISTRIBUTED CIRCUIT MODEL

Transmission lines are very often represented by the immensely useful distributed circuit model, which is capable of describing the propagation not only of TEM waves but also, with appropriate definitions of current and voltage, of dominant-mode waves in hollow waveguides.

Figure 1.5-1 shows symbolically an elementary length Δx of line with its associated inductance $l\Delta x$, capacitance $c\Delta x$, resistance $r\Delta x$ and conductance $g\Delta x$. The model is justified in the following way. The magnetic field between the line's conductors links the circuit formed by generator, line, and termination, and hence is represented by series inductance per unit length of line. The electric field fills the dielectric space between the conductors and thus gives rise in the model to parallel capacitance per unit length. Conductor loss is accounted for by adding resistance in series with inductance, and dielectric loss by shunting the capacitance with conductance. We will write these parameters with lower-

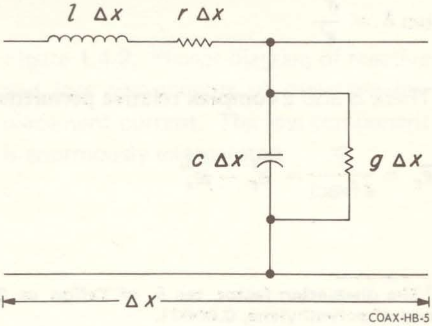


Figure 1.5-1. Distributed parameters of an elementary length of transmission line.

case letters as a reminder that each is a quantity per unit length of line. Thus l, c, r, and g are respectively the series inductance (henrys/meter), shunt capacitance (farads/meter), series resistance (ohms/meter), and shunt conductance (ohms^{-1}/meter). Perhaps we should emphasize that these parameters are linearly distributed, not lumped into coils, capacitors, etc that are periodically disposed along the line, as Figure 1.5-1 might misleadingly imply. Any length Δx of line, no matter how short, contains series inductance equal to $l\Delta x$, shunt capacitance equal to $c\Delta x$ and so forth.

The inductance per unit length of a lossless coaxial line is

$$l = \frac{\mu(\text{vac})}{2\pi} \log_e \frac{b}{a} \qquad \text{(henrys/meter)} \qquad (1.5\text{-}1)$$

and the capacitance per unit length is

$$c = 2\pi\, \epsilon(\text{vac}) \frac{\epsilon_r}{\log_e \dfrac{b}{a}} \qquad \text{(farads/meter)} \qquad (1.5\text{-}2)$$

Both these parameters are independent of frequency, except insofar as ϵ_r may be a function of frequency. If losses and non-ideal conducting surfaces are taken account of, one obtains expressions for the dissipative parameters r and g and also for an additional component of l. These quantities are all frequency-dependent. (We shall discuss the theory of the distributed-circuit model in detail in Chapter 4.)

In the zero-loss approximation the characteristic impedance is given in terms of the distributed parameters by

$$Z_c = \sqrt{\frac{l}{c}} \qquad \text{(ohms)} \qquad (1.5\text{-}3)$$

and the velocity of propagation by

$$v = 1/\sqrt{lc} \qquad \text{(meters/second)} \qquad (1.5\text{-}4)$$

Example: What is the capacitance per foot of a 50-ohm cable with solid polyethylene ($\epsilon_r = 2.25$) dielectric?

From (1.5-3) and (1.5-4) we have

$$c = \frac{1}{v} \cdot \frac{1}{Z_c}$$

The velocity is given by (1.2-5):

$$v = \frac{v(\text{vac})}{\sqrt{\epsilon_r}} = (3 \times 10^8 \text{ m/s})/\sqrt{2.25} = 2 \times 10^8 \text{ m/s}.$$

Thus

$$c = \frac{1}{2 \times 10^8 \text{ m/s}} \cdot \frac{1}{50 \text{ ohms}} = 10^{-10} \text{ farad/meter} =$$

100 pF/meter = 100 pF/3.28 feet = 30.5 pF/foot.

Example: RG-71A/U is a low-capacitance cable with a dielectric of air-spaced polyethylene which gives a velocity factor of 0.84. The capacitance is 13.5 pF/foot. What is the characteristic impedance?

$$Z_c = \frac{1}{v} \cdot \frac{1}{c}$$

The capacitance per meter is 13.5 pF/0.305 meter = 44.5 pF/meter. Therefore

$$Z_c = \frac{1}{0.84 \times 3 \times 10^8 \text{ m/s}} \cdot \frac{1}{44.5 \times 10^{-12} \text{ F/m}} = 89 \text{ ohms.}$$

1.6 HIGHER MODES

We said in Section 1.1 that there are, in addition to the principal or TEM mode, infinitely many higher modes (or waveguide modes) that can propagate on a coaxial line at sufficiently high frequencies. Let us recapitulate the ways in which TEM and higher-mode waves differ. 1) Both the electric and magnetic fields of TEM waves are perpendicular to the direction of propagation. Higher-mode waves also have a field component in the direction of propagation. 2) A transmission line that is to transmit TEM waves must have two or more conductors (the cross section of its conducting surfaces must be a multiply-connected curve). Higher-mode waves can propagate on any kind of transmission line, including single-conductor (simply connected) structures such as hollow waveguides. 3) TEM waves may have any frequency; higher-mode waves can propagate only above certain cutoff frequencies that depend on the particular mode and the cross section of the transmission line. 4) The velocity of TEM waves is independent of frequency, while velocities of waves belonging to the higher modes are frequency-dependent.

The importance of higher modes in coaxial lines is that the onset of waveguide propagation sets an upper limit to the coaxial line's normal useful frequency range. This is so because there is no practical way to prevent the higher modes from interfering with propagation in the principal mode, since any discontinuity in the coaxial systems is likely to couple the TEM fields with those of higher modes.

The coaxial waveguide mode with the lowest cutoff frequency is the H_{11} (or TE_{11}) mode, whose fields are shown in Figure 1.6-1. The cutoff frequency of the H_{11} mode is given approximately by

$$f_{\text{cutoff}} \doteq \frac{v_{\text{TEM}}}{\pi(a+b)} \tag{1.6-1}$$

where v_{TEM} is the velocity of TEM waves in the medium that fills the space between the conductors. One can see from (1.6-1) that f_{cutoff} is the frequency at which the mean circumference of the conductors is approximately equal to a wavelength. If we take as an example standard 9/16-inch 50-ohm air-dielectric line (a = 0.122 inch, b = 0.281 inch), equation 1.6-1 gives a cutoff frequency of about 9.4 GHz. If this same line is now filled with polystyrene, whose dielectric constant is approximately 2.5, the cutoff frequency is reduced by a factor of $1/\sqrt{2.5}$ to about 6 GHz.

The phase velocity[†] of non-TEM waves is higher than that of TEM waves; it is infinite at the cutoff frequency and approaches the TEM velocity as the frequency gets higher[††].

$$\frac{v_{\text{higher mode}}}{v_{\text{TEM}}} = \frac{1}{\sqrt{1-\left(\frac{f_{\text{cutoff}}}{f}\right)^2}} \tag{1.6-2}$$

[†] Phase velocity is the velocity of propagation of any given point of an infinitely long sinusoidal traveling wave. When phase velocity is constant with frequency (as it is in the case of TEM waves) sinusoidal waves, pulses, and modulation envelopes all travel at the same speed and there is no ambiguity when the term "velocity" is used without qualification. But when phase velocity changes with frequency, as it does in the case of higher-mode waves, pulses and modulation envelopes travel more slowly than sine waves and become distorted. There is then said to be "dispersion" and one must distinguish between phase velocity and the velocity of, say, the center of a pulse.

[††] In case the reader thinks this statement conflicts with relativity theory: it doesn't.

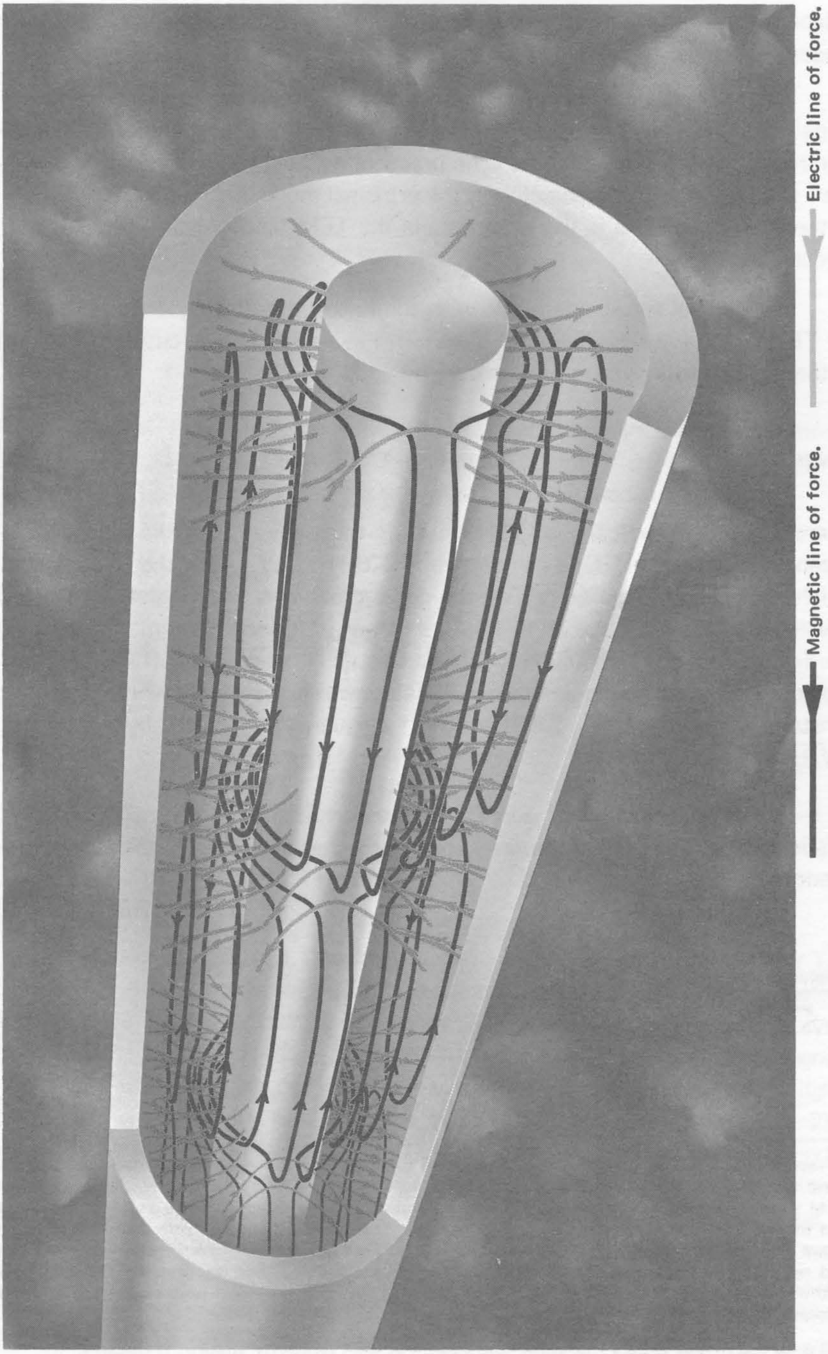


Figure 1.6-1. Fields of the H_{11} coaxial mode.

Below its cutoff frequency a waveguide mode is nonpropagating. The phase constant is zero (there is no change in phase from one place to another) and the fields are rapidly attenuated, the more so the lower the frequency.

$$\alpha\left(\begin{matrix}\text{higher mode}\\ \text{below cutoff}\end{matrix}\right) = \frac{2\pi f_{\text{cutoff}}}{v_{\text{TEM}}}\sqrt{1-\left(\frac{f}{f_{\text{cutoff}}}\right)^2} \tag{1.6-3}$$

This is not an attenuation due to dissipation, like the attenuation of the TEM mode that we discussed in Section 1.4, but due to reflection from a waveguide that is too small to allow the wave to propagate. Below cutoff, too, the electric and magnetic fields are in phase quadrature—the characteristic impedance is reactive. In the propagating region above cutoff the fields are in phase, as they are in the TEM mode.

As an illustration of the kind of difficulty that waveguide modes may cause, we cite the H_{11}-mode resonance of a dielectric support bead in an air-dielectric line. If we consider the section containing the bead as a length of solid-dielectric line, then, as we saw above, the H_{11}-mode cutoff frequency will be considerably lower in this section than in the rest of the line where the dielectric is air. One might not expect to observe any effect due to H_{11}-mode propagation in the bead because, at frequencies that are below H_{11} cutoff in the empty line, the bead is very short compared with the H_{11} wavelength in the bead. But this argument ignores the fact that the bead is terminated on both sides by lengths of air-filled line, which present inductive reactances to the H_{11} waves in the bead at frequencies below H_{11} cutoff in the empty line. Thus, in the frequency range above H_{11} cutoff in the bead but still below cutoff in the air-filled line, it is possible for resonance to occur in the short section of solid-dielectric line with its two inductive terminations. Such resonances have been observed† as narrow peaks in the insertion loss of the bead.

---

†J. F. Gilmore, "TE_{11}-Mode Resonance in Precision Coaxial Connectors," *General Radio Experimenter*, August 1966.

Standing Waves

In the preceding sections we have been talking about sinusoidal traveling TEM waves in coaxial transmission lines, and we have introduced the parameters that describe their propagation: the volocity v (or the phase constant β), the characteristic impedance Z_c, and the attenuation constant α.[†]

A pure traveling wave can exist only on a section of line that is terminated at the receiving, or load, end by a device that reflects no energy back toward the generator. Since in practice there are no perfectly reflectionless terminations, there are always two traveling waves at any point on a transmission line, a **forward** (or **incident**) wave propagating from the generator toward the load and a **reflected** wave propagating back toward the generator. It is the interference of the forward and reflected waves—constructive here, destructive a quarter wavelength away—that produces the distribution of fields along the line that is called a **standing wave.**

1.7 THE REFLECTED WAVE

Any discontinuity in the uniform construction of the transmission line generates reflections. Thus, not only the terminating load but also connectors, junctions, bends, probes, holes, transitions, tuning screws, support beads, and so on are all sources of reflected waves. In Chapter 3 we shall have something to say about the reflections contributed by individual discontinuities, but for the present we shall consider the simple situation, depicted in Figure 1.7-1, in which a uniform line is terminated in a load which is the only source of reflections.

Before going on to talk about the generation of a reflected wave, we must stop for a moment and discuss the lumped impedance that we show at the end of the line in Figure 1.7-1. Of course this is just a convenient fiction that we use to represent the actual state of affairs at the end of the line. One might think that this goes without saying, since one is so used to seeing a one-port device represented at low frequencies by a lumped impedance equal to the impedance that the device presents at its terminals. A transistor is shown schematically with a

---

[†] Let us point out here that, although the primary concern of this book is coaxial lines, nearly everything we shall have to say in the rest of the book is applicable to all kinds of transmission lines whether they work in the TEM mode or not. The reason is that non-TEM as well as TEM waves are described by an appropriately defined "voltage" that is proportional to the electric field, a "current" that is proportional to the magnetic field, a characteristic impedance Z_C, a phase constant β, and an attenuation constant α. Thus the reader who is also interested in waveguides will find the material in the remainder of this chapter and in succeeding chapters relevant.

zig-zag line in its collector circuit marked "load," although the load actually might be a loudspeaker. So far as the transistor is concerned, all that matters is the ratio of voltage to current at the load terminals, and the effects of the loudspeaker and the acoustical circuit of the cabinet and its environment are of no interest except insofar as they affect this ratio.

But at microwave frequencies the situation is not quite so simple. We may still represent a one-port device with a lumped impedance equal to the ratio of voltage to current at its terminals *provided terminals are chosen at which it is possible to define a voltage and a current.* But surely there is no problem in talking about voltage and current anywhere we please! This is a prejudice one acquires as a result of a low-frequency upbringing, and it is not true. *In microwave circuits the concepts of voltage and current are meaningful only in a transmission line and only when a single mode is propagating.* For this reason two strict conditions must be met before we can talk about the impedance of a microwave one-port. First, the device must have a piece of transmission line sticking out of it. Second, somewhere in this transmission line, far enough from the physical termination that the TEM fields are not distorted, a transverse **reference plane** or **terminal plane** t must be established. The plane t defines the device's port or "terminals," and the ratio of voltage to current at t is what we shall mean by the device's impedance. The value of the impedance will, as we shall see in the next section, depend on the location that is chosen for the terminal plane.

In practice, the piece of transmission line may be provided by the device's connector. The terminal plane might then be specified at the outer surface of a bead supporting the connector's inner conductor. Alternatively, in a connector that makes a butt contact, a mating contact surface can provide the reference plane. However the terminal plane may be defined, it is important for the reader to realize that the impedance of a microwave one-port is a meaningless number

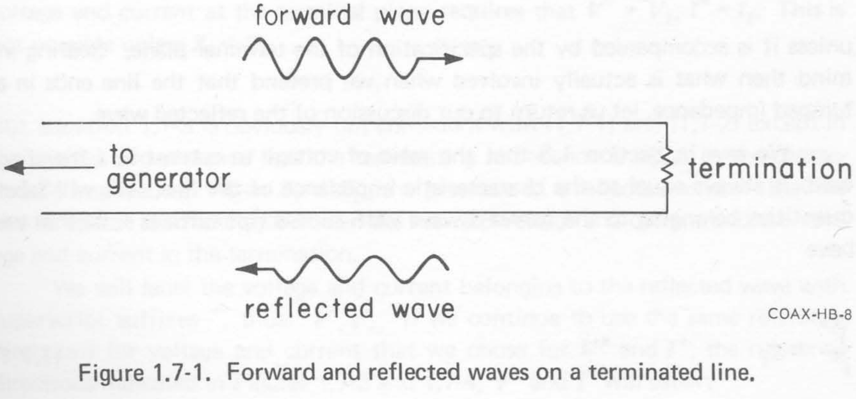


Figure 1.7-1. Forward and reflected waves on a terminated line.

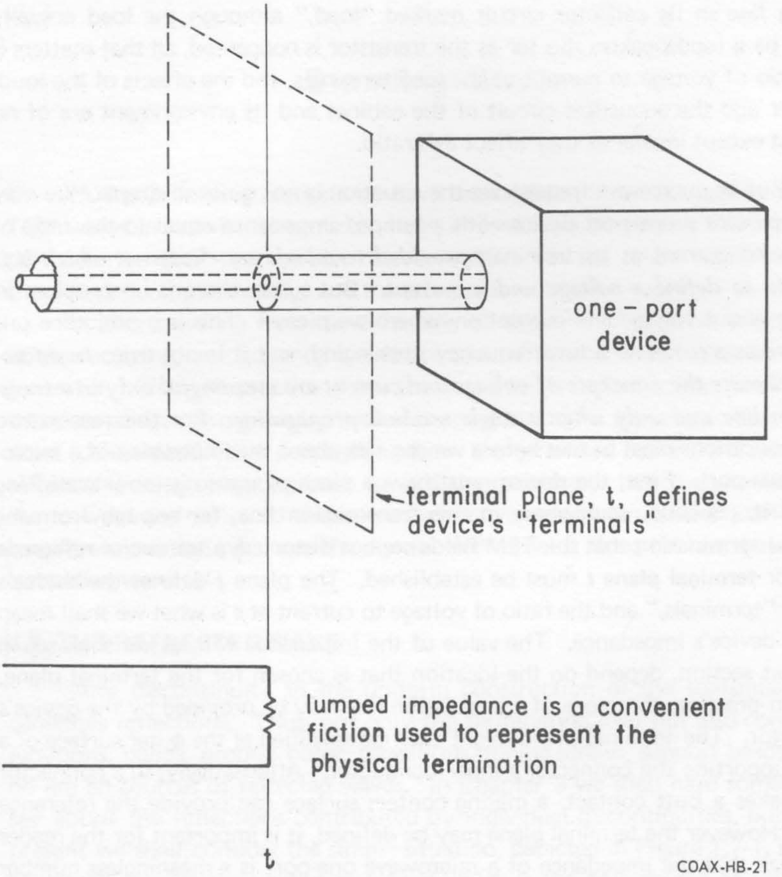


Figure 1.7-2. The presence of the physical termination is accounted for by means of a hypothetical lumped immittance at the terminal plane.

unless it is accompanied by the specification of the terminal plane. Bearing in mind then what is actually involved when we pretend that the line ends in a lumped impedance, let us return to our discussion of the reflected wave.

We saw in Section 1.3 that the ratio of voltage to current in a traveling wave is always equal to the characteristic impedance of the line. We will label quantities belonging to the forward wave with superscript suffixes +, so that we have

$$\frac{V^+}{I^+} = Z_c \tag{1.7-1}$$

But at the same time, the total voltage and total current in the termination, which we will label with subscript suffixes t, have to satisfy

$$\frac{V_t}{I_t} = Z_t \qquad (1.7\text{-}2)$$

where Z_t is the terminating impedance. Suppose for a moment that the forward wave is the only wave on the line. Since voltage and current have to be continuous across the terminal plane, we would then have

$$\left.\begin{aligned} V_t &= V^+ \\ I_t &= I^+ \end{aligned}\right\} \text{ (only a forward wave on the line)} \qquad (1.7\text{-}3)$$

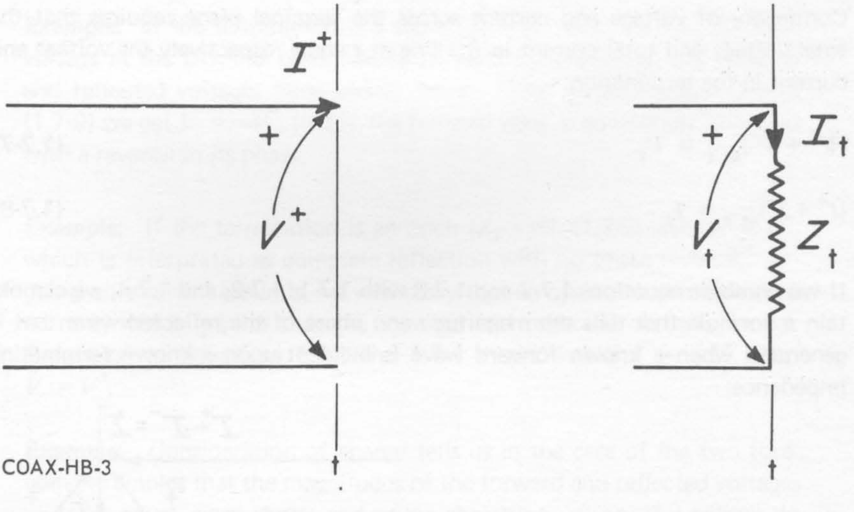


Figure 1.7-3. If the incident wave is the only wave on the line, continuity of voltage and current at the terminal plane requires that $V^+ = V_t$, $I^+ = I_t$. This is not possible unless $Z_t = Z_c$.

But equation 1.7-3 is obviously not consistent with (1.7-1) and (1.7-2) except in the special circumstance that the terminating impedance is equal to the characteristic impedance. When $Z_t \neq Z_c$, the presence of a reflected wave on the line makes up the discrepancy between the forward voltage and current and the voltage and current in the termination.

We will label the voltage and current belonging to the reflected wave with superscript suffixes $^-$, thus: V^-, I^-. If we continue to use the same reference directions for voltage and current that we chose for V^+ and I^+, the reference directions indicated in Figures 1.7-3 and 1.7-4, V^- and I^- will satisfy

$$\frac{V^-}{I^-} = -Z_c \tag{1.7-4}$$

This is the same equation as (1.7-1) except for the minus sign in front of Z_c, which arises because V^- and I^- belong to a wave that travels away from the termination rather than toward it.

The *total* voltage and *total* current on the line are the sums of the forward and reflected voltages and forward and reflected currents:

$$V = V^+ + V^- \tag{1.7-5}$$

$$I = I^+ + I^- \tag{1.7-6}$$

Continuity of voltage and current across the terminal plane requires that the total voltage and total current in the line at t equal respectively the voltage and current in the termination:

$$(V^+ + V^-)_{\text{at } t} = V_t \tag{1.7-7}$$

$$(I^+ + I^-)_{\text{at } t} = I_t \tag{1.7-8}$$

If we combine equations 1.7-7 and 1.7-8 with 1.7-1, 1.7-2, and 1.7-4, we can obtain a formula that tells the magnitude and phase of the reflected wave that is generated when a known forward wave is incident upon a known terminating impedance:

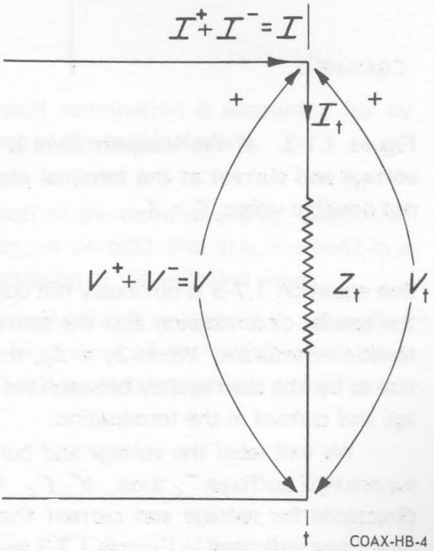


Figure 1.7-4. When $Z_t \neq Z_c$, there is a reflected wave on the line and the *total* voltage and *total* current on the line are equal respectively to the voltage and current in the termination.

$$V^-(\text{at } t) = \frac{\dfrac{Z_t}{Z_c} - 1}{\dfrac{Z_t}{Z_c} + 1} \cdot V^+(\text{at } t) \tag{1.7-9}$$

Example: We have seen that if the terminating impedance equals the characteristic impedance of the line, a forward wave alone satisfies the requirements of voltage and current continuity at the terminal plane. When $Z_t = Z_c$, (1.7-9) gives $V^- = 0$. A terminating impedance that equals the characteristic impedance is called a **reflectionless termination.**[†]

Example: If the termination is a short ($Z_t = 0$), we know that the total voltage at the terminal plane has to be zero; consequently, the forward and reflected voltages must cancel there. If we substitute $Z_t = 0$ into (1.7-9) we get $V^- = -V^+$, that is, the forward wave is completely reflected with a reversal in its phase.

Example: If the termination is an open ($Z_t = \infty$), (1.7-9) gives $V^- = V^+$, which is interpreted as complete reflection with no phase reversal. Arguing on physical grounds, we would say that an open circuit means zero current, which implies that the forward and reflected currents cancel. Reference to equations 1.7-1 and 1.7-4 shows that if $I^- = -I^+$, then $V^- = V^+$.

Example: Consideration of energy tells us in the case of the two foregoing examples that the magnitudes of the forward and reflected voltages must be equal, since shorts and opens absorb no power. But neither do reactances absorb power. If $Z_t = jX_t$, one can show quite easily that

$$\left| \frac{\dfrac{jX_t}{Z_c} - 1}{\dfrac{jX_t}{Z_c} + 1} \right| = 1$$

so that (1.7-9) gives $|V^+| = |V^-|$.

[†] Or very often a "matched" termination. However, we shall avoid the term "matched" because it is used with several different meanings.

We will now introduce several quantities that are used to express the magnitude and phase or just the magnitude of the reflection. The ratio of the reflected to the incident voltages is called the **reflection coefficient**. We shall represent it with a Γ, although ρ is often used.

$$\Gamma = \frac{V^-}{V^+} \tag{1.7-10}$$

Equation 1.7-9 gives the reflection coefficient at the termination:

$$\Gamma \text{ (at } t) = \frac{\dfrac{Z_t}{Z_c} - 1}{\dfrac{Z_t}{Z_c} + 1} \tag{1.7-11}$$

(We shall have a good deal more to say about this extremely important formula in the next section and in Chapter 2.) Reflection coefficients, like immittances, are ratios of phasors and are consequently complex quantities. We shall use θ for the angle of Γ.

$$\Gamma = |\Gamma| \measuredangle \theta \tag{1.7-12}$$

θ is the angle by which the *reflected* voltage *leads* the *incident* voltage. The magnitude of Γ can have values from zero, which corresponds to a reflectionless termination, to unity, which corresponds to a totally reflecting termination, that is, an open, a short, or a pure reactance. The relation between the reflection coefficient and the forward and reflected currents is

$$\Gamma = -\frac{I^-}{I^+} \tag{1.7-13}$$

as one can see by comparing (1.7-10) with (1.7-1) and (1.7-4).

We shall give a proper definition of **standing-wave ratio (SWR)** in Section 1.9, but for completeness we must mention it here since it is one of the commonest ways of describing the *magnitude* of the reflection. The standing-wave ratio r (S and σ are also used) is related to the magnitude $|\Gamma|$ of the reflection coefficient by

$$r = \frac{1 + |\Gamma|}{1 - |\Gamma|} \qquad |\Gamma| = \frac{r - 1}{r + 1} \tag{1.7-14}$$

Since, as we shall see in Section 1.9, the standing-wave ratio is a voltage ratio, it is expressed in decibels by

$$r(\text{dB}) = 20 \log_{10} r \tag{1.7-15}$$

One can show that r(dB) and $|\Gamma|$ are related by

$$r(\text{dB}) = 8.686 \times 2 \tanh^{-1} |\Gamma| \tag{1.7-16}$$

where $\tanh^{-1}$ is the inverse hyperbolic tangent. The standing-wave ratio can have values from unity (0 dB) for a reflectionless termination to infinity (∞ dB) for a totally reflecting one.

Return loss, which we shall designate with an R, compares the power in the reflected wave with that in the forward wave. It is the number of decibels between the amount of power in the forward wave and the amount of power in the reflected wave.

$$R(\text{dB}) = 10 \log_{10} \frac{\text{incident power}}{\text{reflected power}} = 10 \log_{10} \frac{|V^+|^2}{|V^-|^2} = 20 \log_{10} \frac{1}{|\Gamma|} \tag{1.7-17}$$

If the termination is totally reflecting the return loss is zero; a reflectionless termination has infinite return loss. The return loss and standing-wave ratio are related by

$$r = \text{ctnh} \left(\frac{1}{2} \frac{R(\text{dB})}{8.686} \right) \tag{1.7-18}$$

where ctnh means the hyperbolic cotangent.

Reflection loss refers to the loss, due to reflection, in power absorbed by the load:[†]

$$\text{reflection loss (dB)} = 10 \log_{10} \frac{\text{incident power}}{\text{power absorbed by termination}}$$

$$= 10 \log_{10} \frac{|V^+|^2}{|V^+|^2 - |V^-|^2} = 10 \log_{10} \frac{1}{1 - |\Gamma|^2} \tag{1.7-19}$$

Zero reflection loss occurs when the load is nonreflecting.

[†] The term "transmission loss" has been used for this quantity, but our term "reflection loss" is more usual. So many different things have been named "transmission loss" that we shall steer entirely clear of the term.

A graphical comparison of the four quantities—reflection coefficient, standing-wave ratio, return loss, and reflection loss—that describe the magnitude of a reflection is presented in Figure 1.7-5. The reader may be surprised to note that a reflection coefficient as high as 0.45 results in a reflection loss of only 1 dB.

A special case of considerable practical interest is that in which the reflection is very small. The following approximations can often be used when the standing-wave ratio is less than about 1.1.

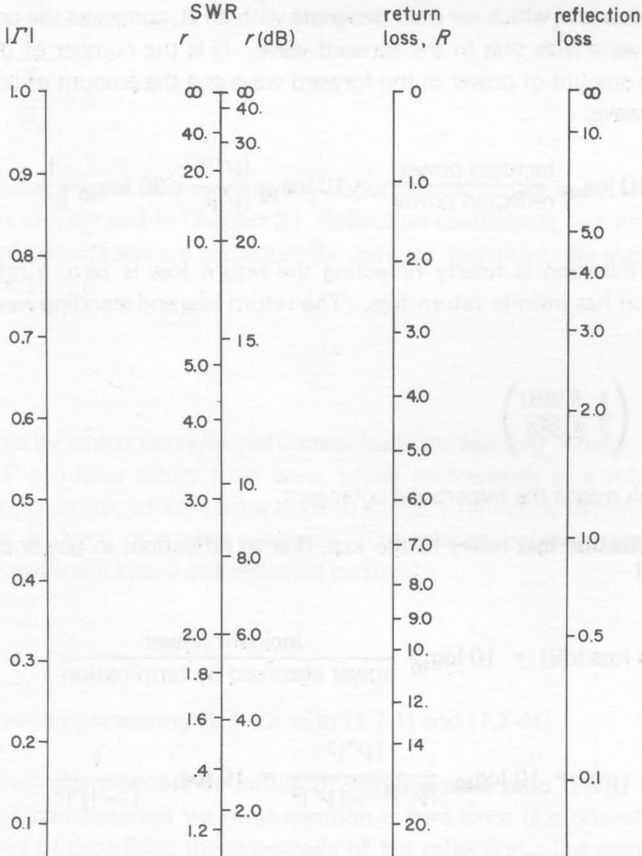


Figure 1.7-5. Graphical comparison of the magnitude |Γ| of the reflection coefficient, the standing-wave ratio, the return loss, and the reflection loss.

$$|\Gamma| \doteq \frac{1}{2}(r-1) \tag{1.7-20}$$

$$r \doteq 1 + 2\,|\Gamma| \tag{1.7-21}$$

$$r(\mathrm{dB}) \doteq 8.686\,(r-1) \tag{1.7-22}$$

$$r \doteq 1 + 0.1151 \times r(\mathrm{dB}) \tag{1.7-23}$$

$$R(\mathrm{dB}) \doteq -20 \log_{10} \frac{1}{2}(r-1) \tag{1.7-24}$$

$$r \doteq 1 + 2 \times 10^{-\frac{1}{20} R(\mathrm{dB})} \tag{1.7-25}$$

$$\text{ref. loss(dB)} \doteq 8.686 \times \frac{1}{2} |\Gamma|^2 \tag{1.7-26}$$

$$\text{ref. loss(dB)} \doteq 8.686 \times \frac{1}{8} (r-1)^2 \tag{1.7-27}$$

$$\text{ref. loss(dB)} \doteq 8.686 \times \frac{1}{2}\, 10^{-\frac{1}{10} R(\mathrm{dB})} \tag{1.7-28}$$

$$R(\mathrm{dB}) \doteq -10 \log_{10} [\text{ref. loss (dB)}] - 6.378 \tag{1.7-29}$$

1.8 IMMITTANCE AND REFLECTION COEFFICIENT

Although we have talked about immittances and reflection coefficients only in connection with the line's termination, it should be clear from the arbitrariness of the way in which the terminal plane is defined that these quantities are equally meaningful at any other reference plane anywhere on the line.

Let us assume that a terminal plane t has already been agreed upon. We will specify the location of any other reference plane by giving its (physical) distance w from t toward the generator. The total voltage and total current at w

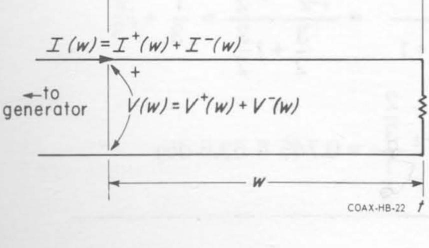


Figure 1.8-1. Total voltage $V(w)$, total current $I(w)$, forward and reflected voltages $V^+(w)$ and $V^-(w)$, and forward and reflected currents $I^+(w)$ and $I^-(w)$ at a reference plane located a distance w from the terminal plane.

will be written $V(w)$ and $I(w)$, and the forward and reflected voltages and currents at w will be written $V^+(w)$, $V^-(w)$, $I^+(w)$, and $I^-(w)$. The impedance $Z(w)$ that we see at the plane w when we look toward the load is defined as the ratio of total voltage to total current at w:

$$Z(w) = \frac{V(w)}{I(w)} \tag{1.8-1}$$

Similarly, the reflection coefficient $\Gamma(w)$ that we see at the plane w when we look toward the load is the ratio of the reflected voltage to the forward voltage at w:

$$\Gamma(w) = \frac{V^-(w)}{V^+(w)} \tag{1.8-2}$$

Mathematically, $Z(w)$ and $\Gamma(w)$ each conveys exactly the same information. If we know one of them we can calculate the other. Equation 1.7-11, which gives the terminating reflection coefficient as a function of the terminating impedance, is obviously quite general, and we now rewrite it so as to show that it is valid at any reference plane w.

$$\Gamma(w) = \frac{\dfrac{Z(w)}{Z_c} - 1}{\dfrac{Z(w)}{Z_c} + 1} \tag{1.8-3}$$

The companion formula for $Z(w)$ as a function of $\Gamma(w)$ is

$$\frac{Z(w)}{Z_c} = \frac{1 + \Gamma(w)}{1 - \Gamma(w)} \tag{1.8-4}$$

Because of the utility of both the impedance and reflection-coefficient concepts, the transformation expressed in (1.8-3 and -4) figures prominently in microwave theory.

Example: What is the reflection coefficient at a reference plane of a 50-ohm line where the impedance is 25 + *j*75 ohms?

$$\Gamma = \frac{\dfrac{(25 + j75)\text{ ohms}}{50\text{ ohms}} - 1}{\dfrac{(25 + j75)\text{ ohms}}{50\text{ ohms}} + 1} = \frac{\left(\dfrac{1}{2} + j\dfrac{3}{2}\right) - 1}{\left(\dfrac{1}{2} + j\dfrac{3}{2}\right) + 1} = \frac{-\dfrac{1}{2} + j\dfrac{3}{2}}{\dfrac{3}{2} + j\dfrac{3}{2}} = \frac{1}{3} + j\frac{2}{3}$$

$$= \sqrt{\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2} \measuredangle \tan^{-1} \frac{\dfrac{2}{3}}{\dfrac{1}{3}} = 0.745 \measuredangle 63.5\text{ deg}$$

Let us interrupt the discussion for a moment to introduce some labor-saving notation. The reader has probably noticed that wherever impedances have occurred in our formulas they have been divided by the characteristic impedance of the line. As a matter of fact, whenever immittances turn up in transmission-line formulas they are always divided by the corresponding characteristic immittance, and we can tidy up such formulas by writing them in terms of **normalized impedances** and **admittances**, which we shall distinguish from the ordinary, or unnormalized, quantities with bars:

$$\bar{Z} = \frac{Z}{Z_c} \qquad \bar{Y} = \frac{Y}{Y_c} \tag{1.8-5}$$

Notice that $\bar{Z}$'s and $\bar{Y}$'s are not impedances and admittances at all; they are dimensionless ratios. In terms of the normalized impedance, equations 1.8-3 and 1.8-4 are

$$\Gamma(w) = \frac{\bar{Z}(w) - 1}{\bar{Z}(w) + 1} \tag{1.8-6}$$

and

$$\bar{Z}(w) = \frac{1 + \Gamma(w)}{1 - \Gamma(w)} \tag{1.8-7}$$

Example: An inductive impedance corresponds to a reflection coefficient that lies in the upper half of the complex plane, that is, to one whose angle θ has a positive value between zero and 180 degrees. We can see why this is so by referring to Figure 1.8-2. If $\bar{Z}$ is inductive, its imaginary

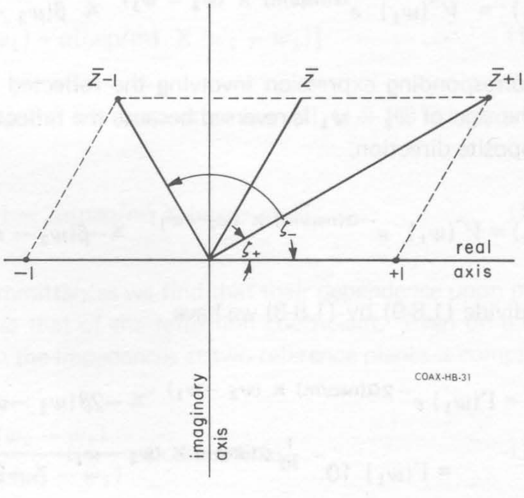


Figure 1.8-2.

part is positive and it lies in the first quadrant of the complex $\bar{Z}$-plane. It is obvious from an inspection of the figure that, no matter where $\bar{Z}$ falls within the first quadrant, 1) the angle ζ_- of the complex number $\bar{Z} - 1$ will always lie between 0 and 180 degrees, 2) the angle ζ_+ of the number $\bar{Z} + 1$ will always lie between 0 and 90 degrees, and 3) ζ_- will always be larger than ζ_+. We conclude therefore that $\zeta_- - \zeta_+$, which is equal to the angle θ of the reflection coefficient Γ, must be between 0 and 180 degrees.

By assuming that $\bar{Z}$ has a negative imaginary part, so that it falls in the fourth quadrant, the reader will be able to show that capacitive impedances correspond to reflection coefficients that lie in the lower half of the complex plane, that is, have negative angles between 0 and -180 degrees.

One of the most important properties of the reflection coefficient is the mathematically simple way in which it changes with position on the line. Suppose we know the reflection coefficient at one reference plane, w_1 say, and wish to calculate it at another, w_2. As we move from w_1 to w_2 we observe that the forward voltage V^+ changes in magnitude by a factor $e^{\alpha(\text{nep/m}) \times (w_2 - w_1)}$. If $w_2 - w_1$ is a positive quantity the magnitude increases, for we are moving toward the generator; if $w_2 - w_1$ is negative the magnitude decreases, for we are moving toward the load. Along with the change in its amplitude, V^+ experiences a phase change between w_1 and w_2 equal to $\beta(w_2 - w_1)$, positive toward the generator, negative toward the load.

We can express the change in V^+ from w_1 to w_2 both in amplitude and in phase by writing

$$V^+(w_2) = V^+(w_1)\; e^{\alpha(\text{nep/m}) \times (w_2 - w_1)} \;\measuredangle\; \beta(w_2 - w_1) \tag{1.8-8}$$

The corresponding expression involving the reflected wave is the same except that the sign of $w_2 - w_1$ is reversed because the reflected wave is propagating in the opposite direction.

$$V^-(w_2) = V^-(w_1)\; e^{-\alpha(\text{nep/m}) \times (w_2 - w_1)} \;\measuredangle\! -\beta(w_2 - w_1) \tag{1.8-9}$$

If we divide (1.8-9) by (1.8-8) we have

$$\begin{aligned} \Gamma(w_2) &= \Gamma(w_1)\; e^{-2\alpha(\text{nep/m}) \times (w_2 - w_1)} \;\measuredangle\! -2\beta(w_2 - w_1) \\ &= \Gamma(w_1)\; 10^{-\frac{1}{10}\alpha(\text{dB/m}) \times (w_2 - w_1)} \;\measuredangle\! -2\beta(w_2 - w_1) \end{aligned} \tag{1.8-10}$$

which is the desired relation between the reflection coefficient at w_1 and that at w_2. Note that the angle of Γ changes in the *negative (clockwise)* sense and the magnitude ***diminishes toward the generator.*** Also note the factor 2. The angle of Γ changes with position on the line *twice* as fast as the phase of a traveling wave, and the magnitude of Γ varies as the *power*, rather than the voltage, of a traveling wave.

One can almost always neglect the attenuation of air-dielectric lines. To the extent that this approximation is valid, the magnitude of the reflection coefficient is constant everywhere on the line while the angle changes with distance at a rate 2β, in the negative sense (clockwise) toward the generator.

$$\Gamma(w_2) = \Gamma(w_1) \cdot 1 \;\measuredangle{-2\beta(w_2 - w_1)} \quad \text{(lossless line)} \qquad (1.8\text{-}11)$$

The standing-wave ratio and return loss, defined in the preceding section, may be used to express the magnitude of the reflection at any point on a transmission line as well as at the termination. On a lossless line they are both constant. On lossy lines the SWR gets smaller and the return loss larger as we get farther from the load. The return loss is affected by the line's attenuation in a particularly simple way: as we move away from the load the return loss increases by just twice the added line attenuation. The relation is expressed by

$$R(w_2) = R(w_1) + 2\alpha \times (w_2 - w_1), \qquad (1.8\text{-}12)$$

where the R's are in decibels if α is in decibels/meter and in nepers if α is in nepers/meter. The corresponding formula in terms of standing-wave ratios is considerably more complicated:

$$r(w_2) = \text{ctnh}\;[\text{ctnh}^{-1}\, r(w_1) + \alpha(\text{nep/m}) \times (w_2 - w_1)] \qquad (1.8\text{-}13)$$

If the reflection is small and if the attenuation is small, (1.8-13) is approximated by

$$r(w_2) - 1 = [r(w_1) - 1]\;[1 - 2\alpha(\text{nap/m}) \times (w_2 - w_1)] \qquad (1.8\text{-}14)$$

When we turn to immittances we find that their dependence upon position is not nearly so simple as that of the reflection coefficient. Even on a lossless line, the relation between the impedances at two reference planes is complicated:

$$\overline{Z}\,(w_2) = \frac{\overline{Z}\,(w_1) + j\tan\beta\;(w_2 - w_1)}{1 + j\overline{Z}(w_1)\;\tan\beta\;(w_2 - w_1)} \quad \text{(lossless line)} \qquad (1.8\text{-}15)$$

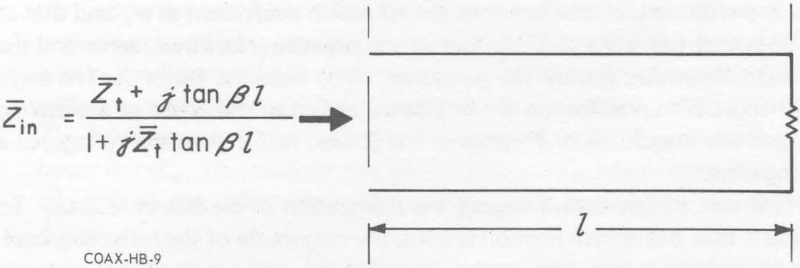


Figure 1.8-3. A transmission-line section of length l transforms the normalized terminal impedance $\overline{Z}_t$ into a normalized input impedance $\overline{Z}_{in}$.

Equation 1.8-15 expresses the impedance-transforming property of a piece of transmission line. If we assume that $w_2 > w_1$, so that the plane w_1 is nearer than w_2 to the load, then one way of looking at (1.8-15) is to regard the length $w_2 - w_1$ of line as a transformer which sees an impedance $\overline{Z}(w_1)$ connected to its output and presents a transformed impedance $\overline{Z}(w_2)$ at its input. Let us rewrite (1.8-15) so as to emphasize this transformer point of view. If a length l of lossless line is terminated in an impedance $\overline{Z}_t$, (1.8-15) shows that its input impedance is

$$\overline{Z}_{in} = \frac{\overline{Z}_t + j \tan\beta l}{1 + j\,\overline{Z}_t \tan\beta l} \tag{1.8-16}$$

One can see from (1.8-15) or (1.8-16) that the transmission line is a different kind of transformer from the low-frequency sort that consists of two coupled coils. For one thing, the transmission line's "turns ratio" is in general a complex number. For another, the "turns ratio" is not fixed; it depends on the load impedance and also on the frequency. Unfortunately there is no microwave equivalent to the low-frequency transformer with its fixed turns ratio, and this makes the problem of broadband impedance matching a difficult one at microwave frequencies.

Equations 1.8-15 and 1.8-16 are hard to use for computation, and the most practical way of performing transmission-line impedance calculations is provided by the Smith chart, the subject of the next chapter. But we can learn quite a lot about the impedance-transforming property of a piece of line by looking at (1.8-16) in a few interesting special cases.

To begin with, if $\overline{Z}_t = 1$, that is, if $Z_t = Z_c$, (1.8-16) gives $\overline{Z}_{in}(l) = 1$, or $Z_{in}(l) = Z_c$ for any length l of line. The impedance anywhere on a reflectionless line is equal to Z_c.

When l is a half wavelength (or any multiple of a half wavelength), $\beta l = 180$ degrees (or a multiple of 180 degrees), the tangents in (1.8-16) are zero, and we have

$$\bar{Z}_{\text{in}}(l = \frac{\lambda}{2}) = \bar{Z}_t \tag{1.8-17}$$

A half-wave lossless line is thus a one-to-one transformer.

When the line length is a quarter wavelength (or an odd multiple of a quarter wavelength), the tangents in (1.8-16) are infinite. The formula nevertheless gives a definite value for $\bar{Z}_{\text{in}}$, which we can find by taking the limit:

$$\bar{Z}_{\text{in}}(l = \frac{\lambda}{4}) = \underset{\beta l \to \pi/2}{\text{limit}} \frac{\bar{Z}_t + j\tan\beta l}{1 + j\bar{Z}_t\tan\beta l} = \frac{1}{\bar{Z}_t} \tag{1.8-18}$$

If we write this result with the unnormalized impedances we have

$$Z_{\text{in}}(l = \frac{\lambda}{4}) = \frac{Z_c^2}{Z_t} \tag{1.8-19}$$

which shows that a quarter-wave line transforms Z_t into its geometric extreme with respect to the characteristic impedance; small terminal impedances become large input impedances and vice versa

Open- and short-circuited line sections, sometimes called stubs, are of considerable practical importance. The input impedance of a shorted stub can be found by putting $\bar{Z}_t = 0$ in (1.8-16), which gives

$$\bar{Z}_{\text{in}}\,(\text{shorted stub}) = j\tan\beta l \tag{1.8-20}$$

Thus a lossless shorted stub looks like a reactance whose value and sign depend on the length. The behavior of the shorted stub is summarized in Table 1.8-1.

The open stub has an input impedance given by

$$\bar{Z}_{\text{in}}\,(\text{open stub}) = \underset{\bar{Z}_t \to \infty}{\text{limit}} \frac{\bar{Z}_t + j\tan\beta l}{1 + j\bar{Z}_t\tan\beta l} = \frac{1}{j\tan\beta l} \tag{1.8-21}$$

and therefore behaves in just the opposite way from the shorted stub (Table 1.8-2).

TABLE 1.8-1

The impedance of the shorted stub.

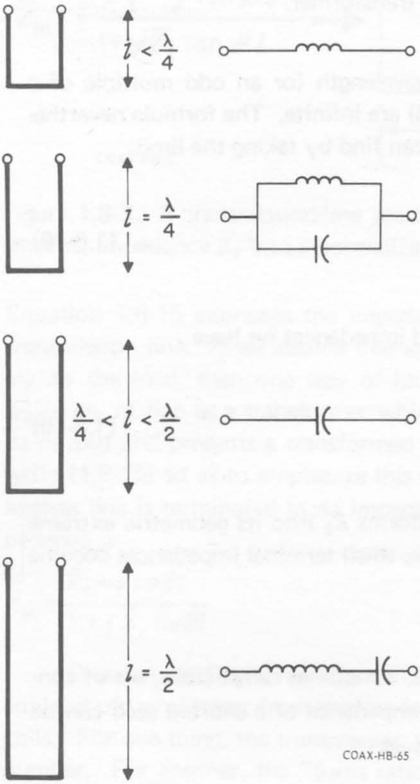
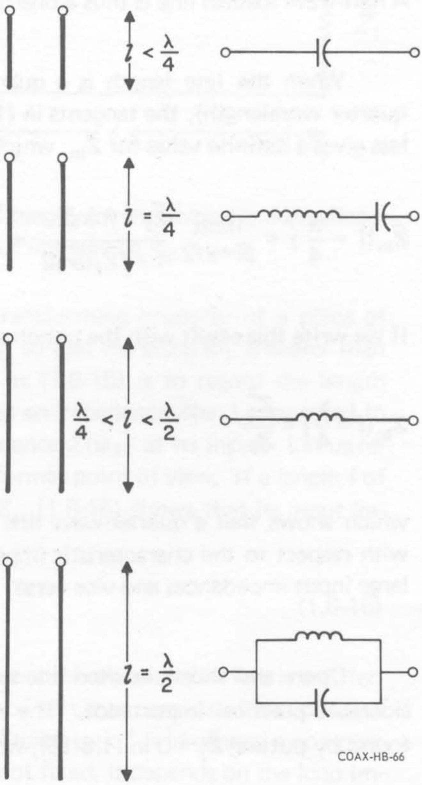


TABLE 1.8-2

The impedance of the open stub.



1.9 STANDING WAVES

The distribution of total voltage and total current on the transmission line, the **standing wave,** is the interference pattern formed by the superposition of the forward and reflected waves. The magnitude of the voltage standing wave due to a totally reflecting termination on a line with a relatively large amount of attenuation per wavelength is shown in Figure 1.9-1. The high loss makes very apparent the fact that the undulations become shallower with increasing distance from the termination as the reflected wave becomes more attenuated and the forward wave less so. In this book, though, the standing waves that we shall be concerned with are those on slotted lines, which can almost always be regarded as lossless. Our discussion will therefore be confined to lossless lines.

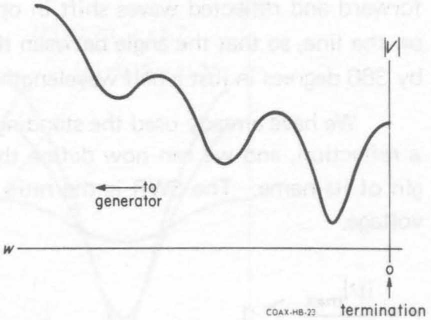


Figure 1.9-1. Voltage standing wave on a very lossy line.

Standing waves on lossless lines are periodic—the maxima are all equal and the minima are all equal. Furthermore, the voltage maxima and minima occur at points on the line where the forward and reflected waves are respectively exactly in- and out-of-phase. Thus

$$|V|_{max} = |V^+| + |V^-| \tag{1.9-1}$$

and

$$|V|_{min} = |V^+| - |V^-| \tag{1.9-2}$$

Note that we do not have to put a w in parentheses after V^+ and V^- because only the magnitudes of these quantities are involved in (1.9-1) and (1.9-2), and the magnitudes of the forward and reflected waves do not change from point to point on a lossless line.

The length of a single period of the standing wave—the distance between adjacent minima or adjacent maxima—is a half wavelength, that is to say, half the wavelength of a traveling wave. The reason for this is that the phases of the

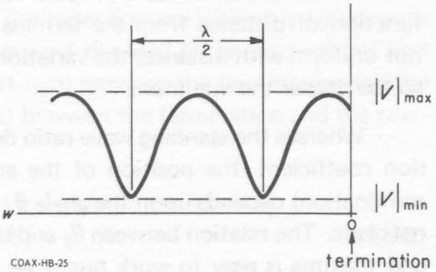


Figure 1.9-2. Voltage standing wave on a lossless line.

forward and reflected waves shift in opposite directions with changing position on the line, so that the angle between the forward and reflected voltages changes by 360 degrees in just a half wavelength.

We have already used the standing-wave ratio r to express the magnitude of a reflection, and we can now define this quantity in a way that explains the origin of its name. The SWR is the ratio of maximum to minimum standing-wave voltage.

$$r = \frac{|V|_{\text{max}}}{|V|_{\text{min}}} \geq 1 \tag{1.9-3}$$

If we substitute (1.9-1) and (1.9-2) in (1.9-3) and note that $|V^-|/|V^+| = |\Gamma|$, the magnitude of the reflection coefficient, we get

$$r = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{1.9-4}$$

This formula, which relates the SWR to the reflection coefficient, is just (1.7-14), with which we defined the SWR in terms of $|\Gamma|$ in Section 1.7.

Figure 1.9-3(a) shows the shapes of the standing-wave distributions of voltage corresponding to three different amounts of reflection. The magnitude of the forward wave is the same in each of the three graphs. When $|\Gamma|$ is small, $|V^-|$ is small compared with $|V^+|$, there is not much difference between $|V|_{\text{max}}$ and $|V|_{\text{min}}$, and r is not much larger than unity. As the reflection grows larger the standing wave becomes more pronounced. When $|\Gamma| = 1$, so that $|V^-|$ and $|V^+|$ are equal, $|V|_{\text{min}} = 0$, $|V|_{\text{max}} = 2\,|V^+|$, and $r = \infty$. It is important to notice that the minima are always sharper than the maxima. This feature disappears as the standing wave becomes very shallow, but at the opposite extreme, when $|\Gamma| = 1$ ($r = \infty$), the minima are cusps.

The phase of the standing-wave voltage is shown (relative to the phase at the terminal plane) in (b) of Figure 1.9-3. The phase of V is an ever-increasing function of distance from the termination. Notice that the change in phase is not uniform with distance; the variation is most rapid near the minima, the more so the deeper the minimum.

Whereas the standing wave ratio depends upon the *magnitude* of the reflection coefficient, the position of the standing wave on the line (relative to the termination) depends upon the *angle* θ_t of the reflection coefficient at the terminal plane. The relation between θ_t and the positions of the standing-wave minima and maxima is easy to work out if we remember two things: 1) At a standing-

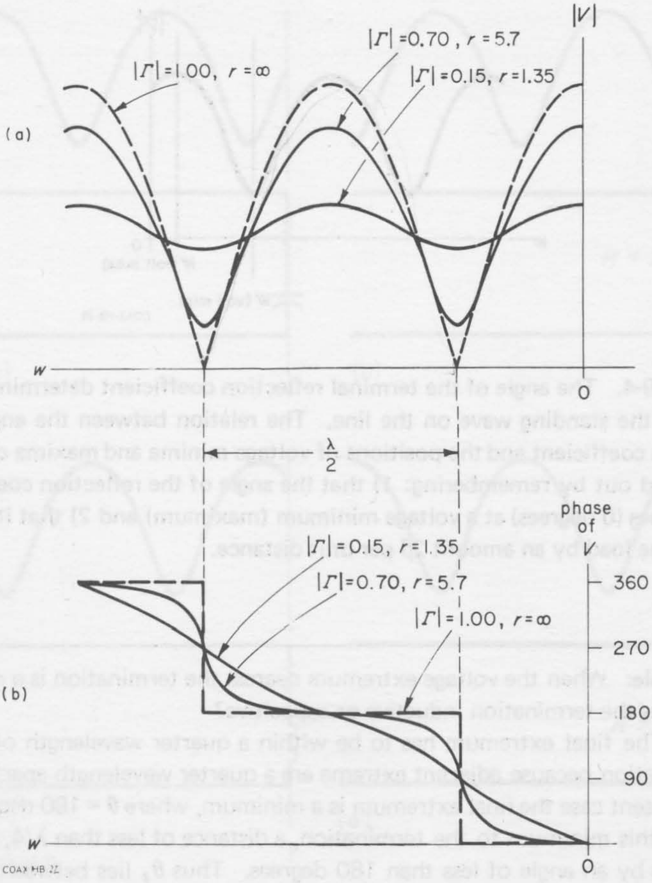


Figure 1.9-3. Voltage standing waves for three different amounts of reflection. (*a*) Magnitudes and (*b*) phases of the total voltage on the line.

wave minimum (maximum) the forward and reflected waves are exactly out-of-phase (in-phase), that is, the angle θ of the reflection coefficient is 180 degrees (0 degrees). 2) The angle θ *increases toward the load* by an amount per unit distance of 2β. Thus the distance w(volt min) between the termination and the nearest voltage minimum, or w(volt max) between the termination and the nearest maximum, is related to θ_t by

$$\begin{Bmatrix}180\text{ deg}\\ 0\text{ deg}\end{Bmatrix} + 2\beta \begin{Bmatrix}w(\text{volt min})\\ w(\text{volt max})\end{Bmatrix} = \theta_t \tag{1.9-5}$$

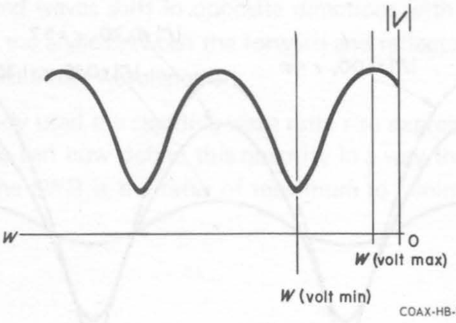


Figure 1.9-4. The angle of the terminal reflection coefficient determines the position of the standing wave on the line. The relation between the angle of the reflection coefficient and the positions of voltage minima and maxima can always be worked out by remembering: 1) that the angle of the reflection coefficient is 180 degrees (0 degrees) at a voltage minimum (maximum) and 2) that it increases toward the load by an amount 2β per unit distance.

Example: When the voltage extremum nearest the termination is a minimum, is the termination inductive or capacitive?

The final extremum has to be within a quarter wavelength of the termination because adjacent extrema are a quarter wavelength apart. In the present case the final extremum is a minimum, where $\theta = 180$ degrees. From this minimum to the termination, a distance of less than $\lambda/4$, θ increases by an angle of less than 180 degrees. Thus θ_t lies between 180 and 360 degrees (or 0 and -180 degrees). We saw in Section 1.8 that reflection coefficients with angles in the lower half of the complex plane correspond to capacitive reactances.

If the extremum nearest the load is a minimum, the load is capacitive; if it is a maximum, the load is inductive. If a minimum falls at the load, the load is a resistance less than Z_c; if a maximum falls at the load, the load is a resistance greater than Z_c.

Example: A minimum is observed at a distance of 0.40λ from the terminal plane. This means that there is a maximum 0.15λ from the termination. (One always measures the location of minima rather than maxima.) What is θ_t?

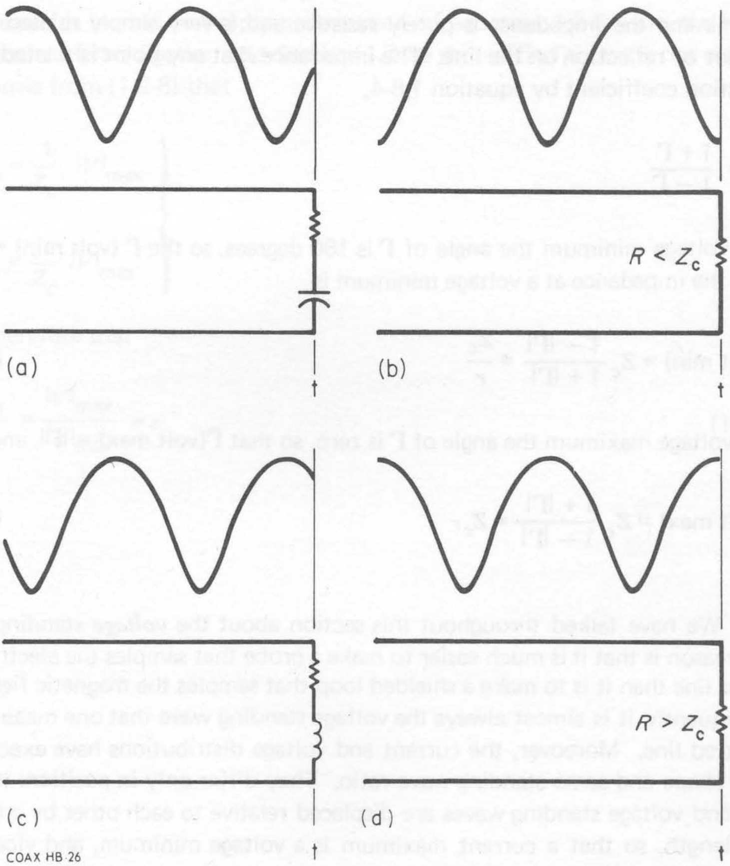


Figure 1.9-5. (*a*) If the voltage extremum nearest the termination is a minimum, the termination is capacitive. (*b*) If the minimum is *at* the termination, the terminating impedance is resistive and smaller than the characteristic impedance. (*c*) A maximum falls nearest an inductive termination, and (*d*) a maximum falls *at* a resistive termination that is larger than the characteristic impedance.

From the final maximum, where θ is 0 degrees, to the termination, an electrical "distance" of 0.15 wavelength X 360 degrees/wavelength = 54 degrees, θ increases by 2 X 54 degrees = 108 degrees. Thus θ_t = 108 degrees.

The impedance, as we have seen, is in general a complex number and varies in a complicated way with position on the line. But at standing wave maxima and minima the impedance is purely resistive and is very simply related to the amount of reflection on the line. The impedance Z at any point is related to the reflection coefficient by equation 1.8-4,

$$\frac{Z}{Z_c} = \frac{1+\Gamma}{1-\Gamma}$$

At a voltage minimum the angle of Γ is 180 degrees, so the Γ (volt min) $= -|\Gamma|$. Thus the impedance at a voltage minimum is

$$Z(\text{volt min}) = Z_c \frac{1-|\Gamma|}{1+|\Gamma|} = \frac{Z_c}{r} \tag{1.9-6}$$

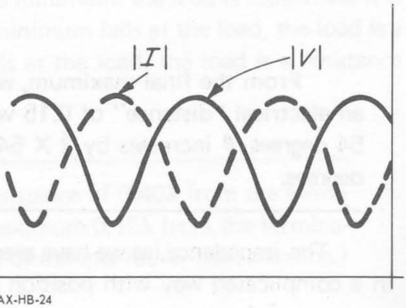
At a voltage maximum the angle of Γ is zero, so that $\Gamma(\text{volt max}) = |\Gamma|$, and

$$Z(\text{volt max}) = Z_c \frac{1+|\Gamma|}{1-|\Gamma|} = Z_c r \tag{1.9-7}$$

We have talked throughout this section about the *voltage* standing wave. The reason is that it is much easier to make a probe that samples the electric field in the line than it is to make a shielded loop that samples the magnetic field, and consequently it is almost always the voltage standing wave that one measures on a slotted line. Moreover, the current and voltage distributions have exactly the same shape and same standing-wave ratio. They differ only in position; the current and voltage standing waves are displaced relative to each other by a quarter wavelength, so that a current maximum is a voltage minimum, and vice versa. Mathematically, the connection between the current and voltage standing waves is

$$|I(w)| = \frac{1}{Z_c} \; |V(w \pm \frac{\lambda}{2})| \tag{1.9-8}$$

Figure 1.9-6. The current and voltage standing waves are displaced relative to each other by a quarter wavelength.



The reader is invited to supply a derivation of equation 1.9-8; with the help of the results of this Section and Section 1.7 he should find that it is not difficult. It follows from (1.9-8) that

$$\left.\begin{aligned} |I|_{\text{max}} &= \frac{1}{Z_c}\,|V|_{\text{max}} \\ &\text{and} \\ |I|_{\text{min}} &= \frac{1}{Z_c}\,|V|_{\text{min}} \end{aligned}\right\} \tag{1.9-9}$$

and therefore that

$$\frac{|I|_{\text{max}}}{|I|_{\text{min}}} = \frac{|V|_{\text{max}}}{|V|_{\text{min}}} = r \tag{1.9-10}$$

CHAPTER 2

The Smith Chart

Transmission-line calculations that one frequently has to make, sometimes over and over again, would be extremely laborious if they had to be done by computation from the formulas given in the previous chapter. The Smith chart provides a quick and powerful graphical method for performing many of these calculations.

2.1 THE REFLECTION-COEFFICIENT PLANE

Figure 2.1-1 shows how a complex number $z = x + jy$ is represented as a point on the complex plane. The real part x is set off on the horizontal real axis, positive toward the right, and the imaginary part y is set off on the vertical imaginary axis, positive upward. We may also express the complex number z by giving its polar coordinates. These are the magnitude (or modulus) $r = \sqrt{x^2 + y^2}$ and angle (or argument or amplitude) $\theta = \tan^{-1} y/x$.

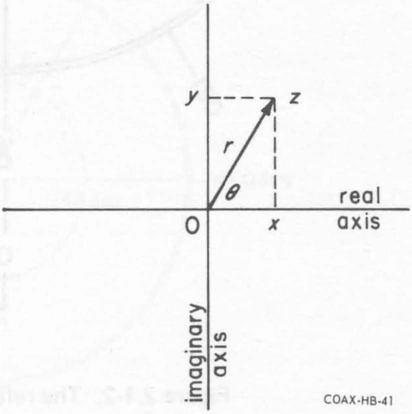
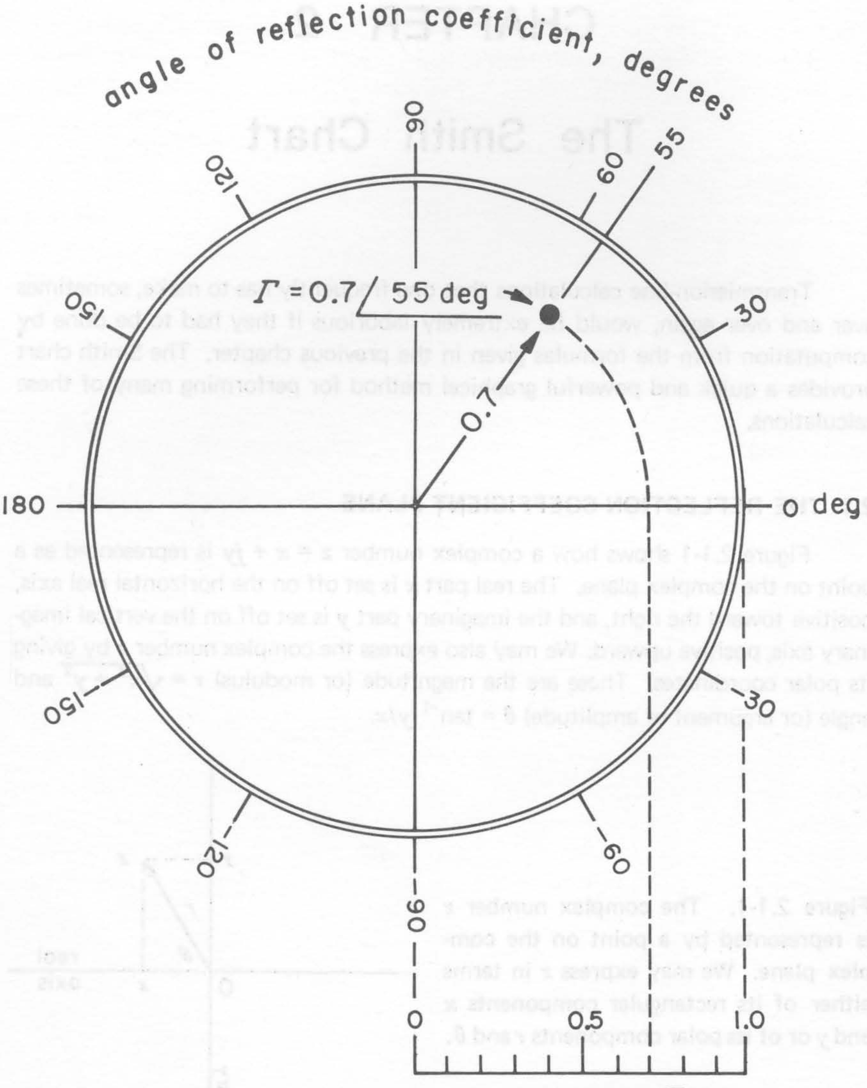


Figure 2.1-1. The complex number z is represented by a point on the complex plane. We may express z in terms either of its rectangular components x and y or of its polar components r and θ.

The particular complex numbers that we wish to plot are reflection coefficients, and the particular plane whose points represent reflection coefficients we shall call the reflection-coefficient plane. We know that the magnitude of a reflection coefficient cannot be greater than unity, at least as long as the load is passive; therefore the part of the reflection-coefficient plane that we shall be con-



COAX-HB-37

Figure 2.1-2. The reflection-coefficient chart.

cerned with is that part lying within a circle of unit radius about the origin. Figure 2.1-2 is a chart of this circular region. Since it is usually most convenient to work with the polar form of a reflection coefficient, the chart includes a radial scale and peripheral degree circle so that values of the magnitude $|\Gamma|$ and angle θ can be located with a straightedge and dividers.

Example: A lossless, air-dielectric line is excited at 600 MHz. The reflection coefficient at reference plane b is $0.5 \measuredangle 60$ deg. How do we find the point on the reflection-coefficient chart corresponding to plane a, 10 centimeters toward the generator? the point corresponding to plane c, 10 centimeters toward the load?

In an air-dielectric line the velocity of propagation is 3×10^{10} cm/s, so that at 600 MHz the wavelength is 3×10^{10} cm-s^{-1}/6×10^{8} s^{-1} = 50 cm,

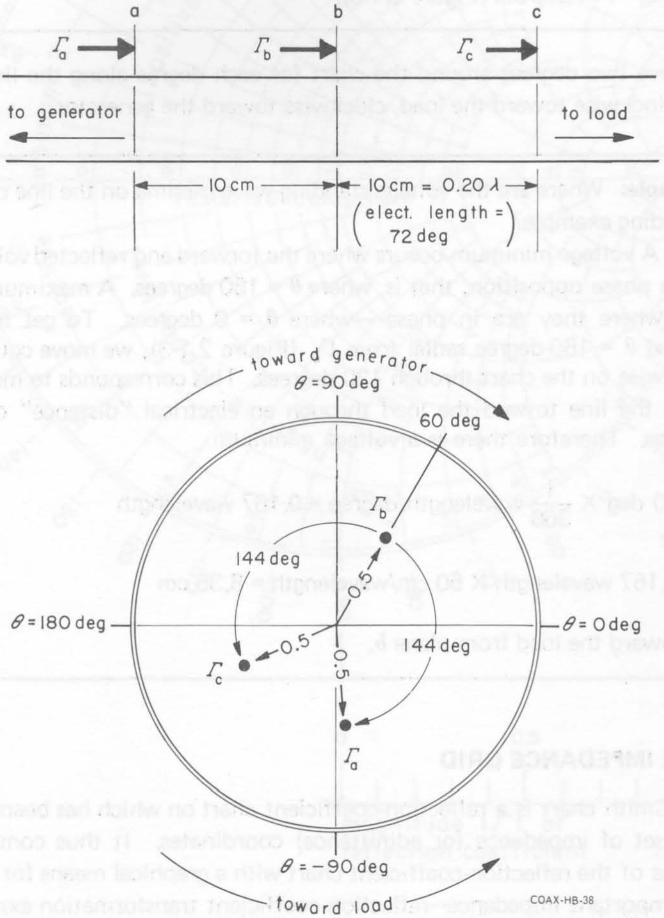


Figure 2.1-3. The angle θ of the reflection coefficient changes with position on the line twice as fast as the phase of a traveling wave. θ increases toward the load.

and a distance of 10 cm, or (10 cm)/(50 cm/wavelength) = 0.20 wavelength, shifts the phase of a traveling wave through an angle of 360 deg/wavelength X 0.20 wavelength = 72 degrees.

We saw in Section 1.8 that $|\Gamma|$ is everywhere the same on a lossless line, while θ changes with position twice as fast as the phase of a traveling wave. The change in θ is positive (counterclockwise) in the direction of the load. Thus to find the points Γ_a and Γ_c on the reflection-coefficient chart, we start at Γ_b and move clockwise and counterclockwise respectively in circular arcs about the chart's center through angles of 2 X 72 = 144 degrees (Figure 2.1-3).

Move two degrees around the chart for each degree along the line. Move counterclockwise toward the load, clockwise toward the generator.

Example: Where are the voltage standing-wave minima on the line of the preceding example?

A voltage minimum occurs where the forward and reflected voltages are in phase opposition, that is, where θ = 180 degrees. A maximum occurs where they are in phase—where θ = 0 degrees. To get to the nearest θ = 180-degree radial from Γ_b (Figure 2.1-3), we move counterclockwise on the chart through 120 degrees. This corresponds to moving along the line toward the load through an electrical "distance" of 60 degrees. Therefore there is a voltage minimum

$$60 \text{ deg} \times \frac{1}{360} \text{ wavelength/degree} = 0.167 \text{ wavelength}$$

or

$$0.167 \text{ wavelength} \times 50 \text{ cm/wavelength} = 8.35 \text{ cm}$$

toward the load from plane b.

2.2 THE IMPEDANCE GRID

A Smith chart is a reflection-coefficient chart on which has been superimposed a set of impedance (or admittance) coordinates. It thus combines the properties of the reflection-coefficient chart with a graphical means for performing the important impedance–reflection-coefficient transformation expressed in equations 1.8-3 and 4.

Figure 2.2-1 shows a Smith chart with a **normalized** impedance grid. The loci of constant $\bar{R}$, the resistive component of $\bar{Z} = Z/Z_c$, and constant $\bar{X}$, the reactive component, are sets of mutually orthogonal circles, as shown in Figure

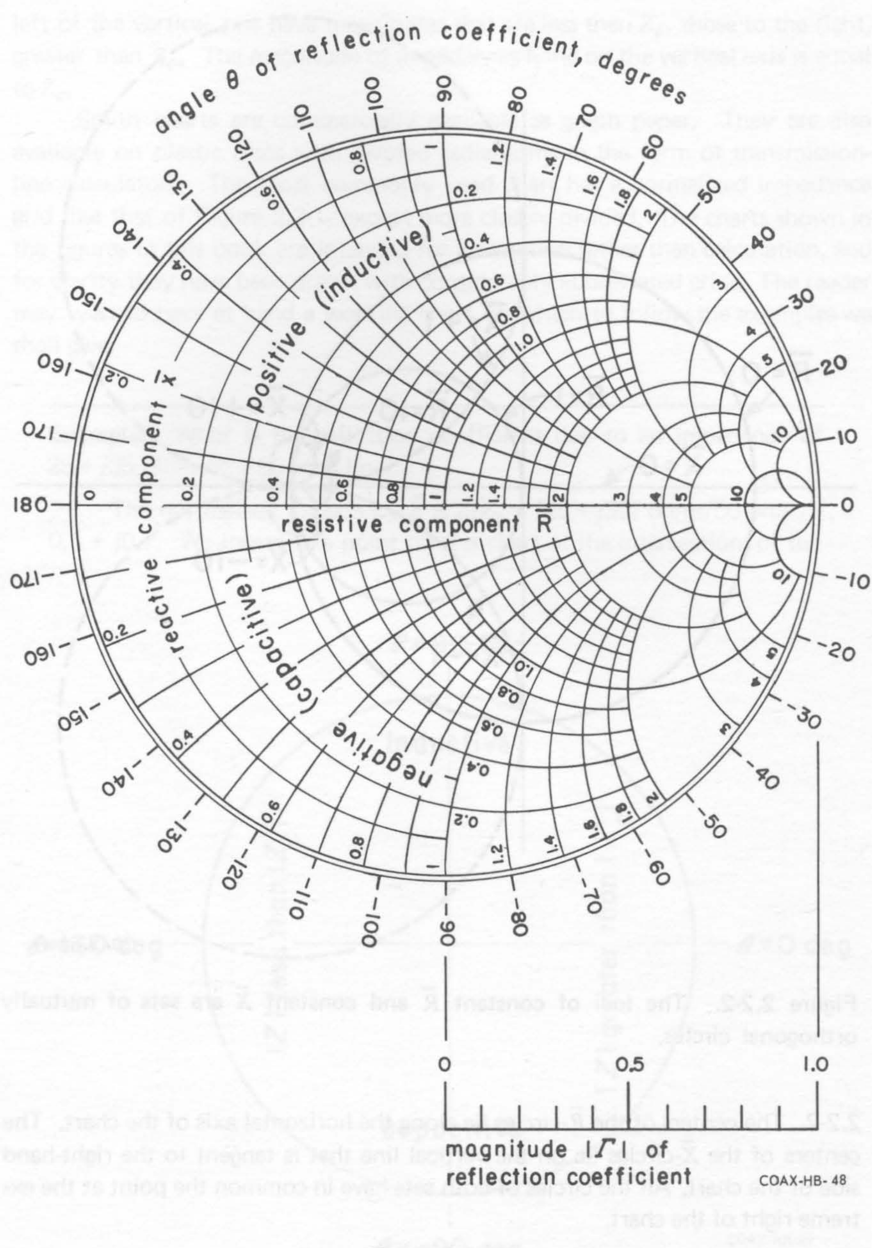


Figure 2.2-1.
The Smith Chart.
It is a reflection-coefficient chart with a superimposed grid of impedance coordinates. The chart shown here has a normalized impedance grid.

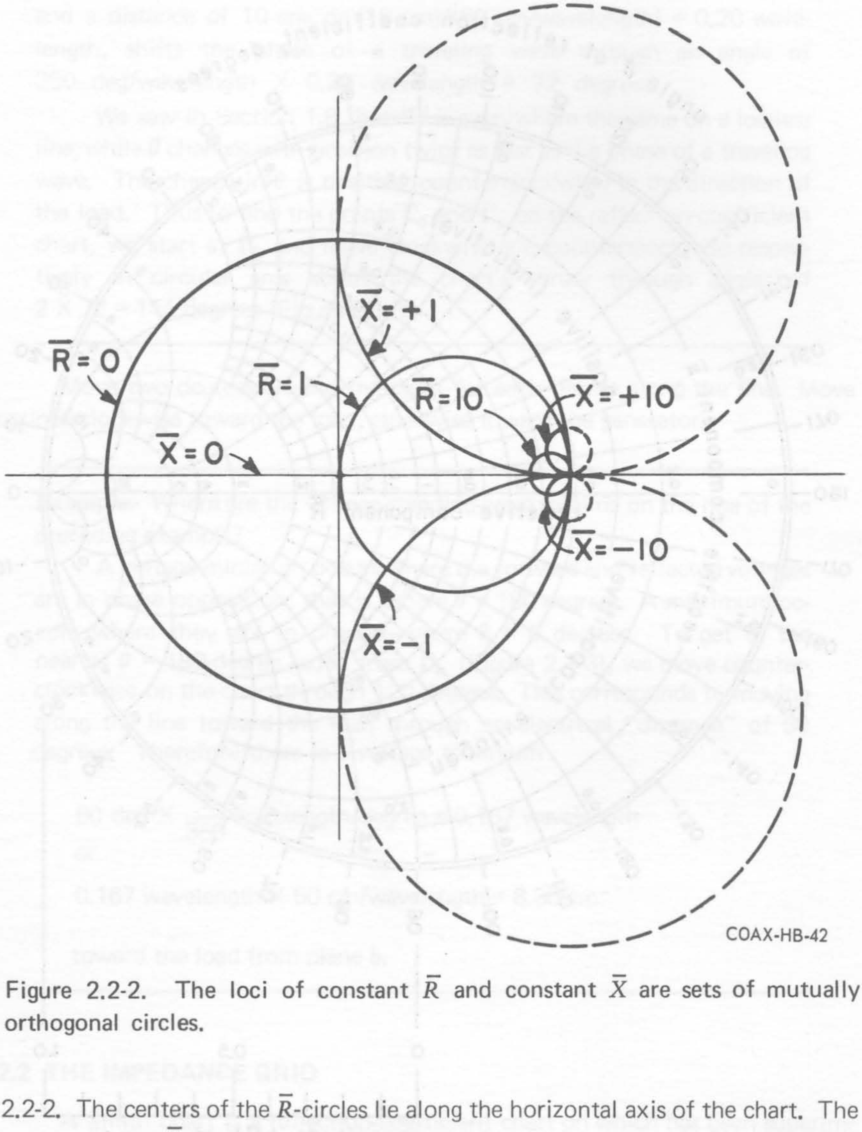


Figure 2.2-2. The loci of constant $\overline{R}$ and constant $\overline{X}$ are sets of mutually orthogonal circles.

2.2-2. The centers of the $\overline{R}$-circles lie along the horizontal axis of the chart. The centers of the $\overline{X}$-circles lie on the vertical line that is tangent to the right-hand side of the chart. All the circles of both sets have in common the point at the extreme right of the chart.

The nature of the impedance in different regions of the chart is indicated in Figure 2.2-3. Points below the horizontal axis correspond to impedances with capacitive reactive components, points above to those with inductive components. Impedances lying on the horizontal axis are resistive. Impedances to the left of the vertical axis have magnitudes that are less than Z_c; those to the right, greater than Z_c. The magnitude of impedances lying on the vertical axis is equal to Z_c.

Smith charts are commercially available as graph paper. They are also available on plastic discs with pivoted radial arms in the form of transmission-line calculators. The most commonly used chart has a normalized impedance grid like that of Figure 2.2-1, except more closely divided. The charts shown in the figures in this book are intended for illustration rather than calculation, and for clarity they have been drawn with considerably abbreviated grids. The reader may wish to have at hand a working chart on which to follow the examples we shall give.

Example: What is the reflection coefficient due to an impedance of 25 + j35 ohms on a 50-ohm line?

The normalized impedance $\bar{Z}$ is Z/Z_c = (25 + j35) ohms/50 ohms = 0.5 + j0.7. We locate this point on the chart at the intersection of the

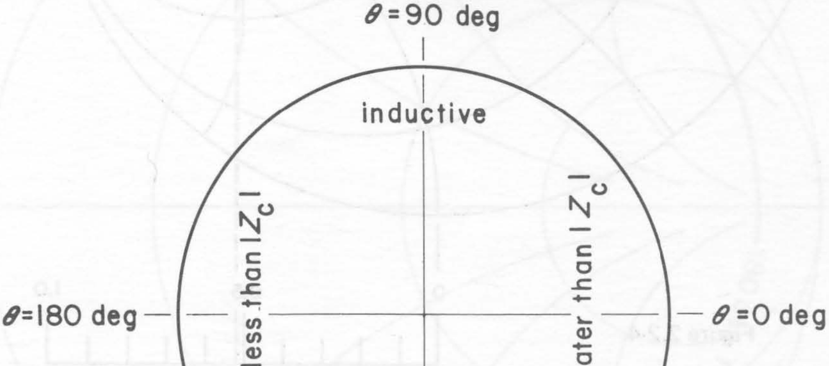


Figure 2.2-3. The nature of the impedance in different regions of the Smith chart.

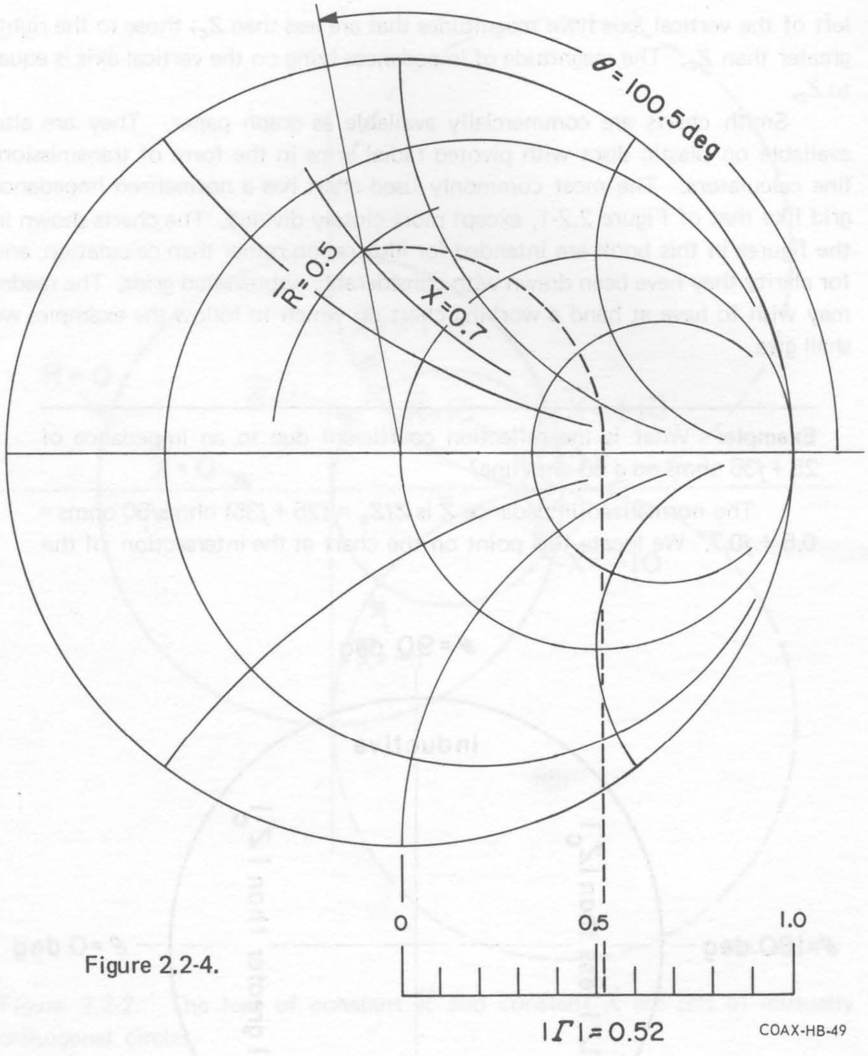


Figure 2.2-4.

$\overline{R} = 0.5$ and $\overline{X} = +0.7$ circles (Figure 2.2-4) and find that it corresponds to a reflection coefficient of 0.52 ∡100.5 deg. We leave it an an exercise for the reader to convince himself of the Smith chart's utility by computing this result with equation 1.8-3.

Example: A quarter-wavelength 50-ohm lossless line is terminated in an impedance $Z_t = 20 + j100$ ohms. What is its input impedance?

We enter the chart (Figure 2.2-5) at the normalized terminal impedance, $\overline{Z}_t = 0.4 + j2.0$, and then move clockwise (toward the generator) in

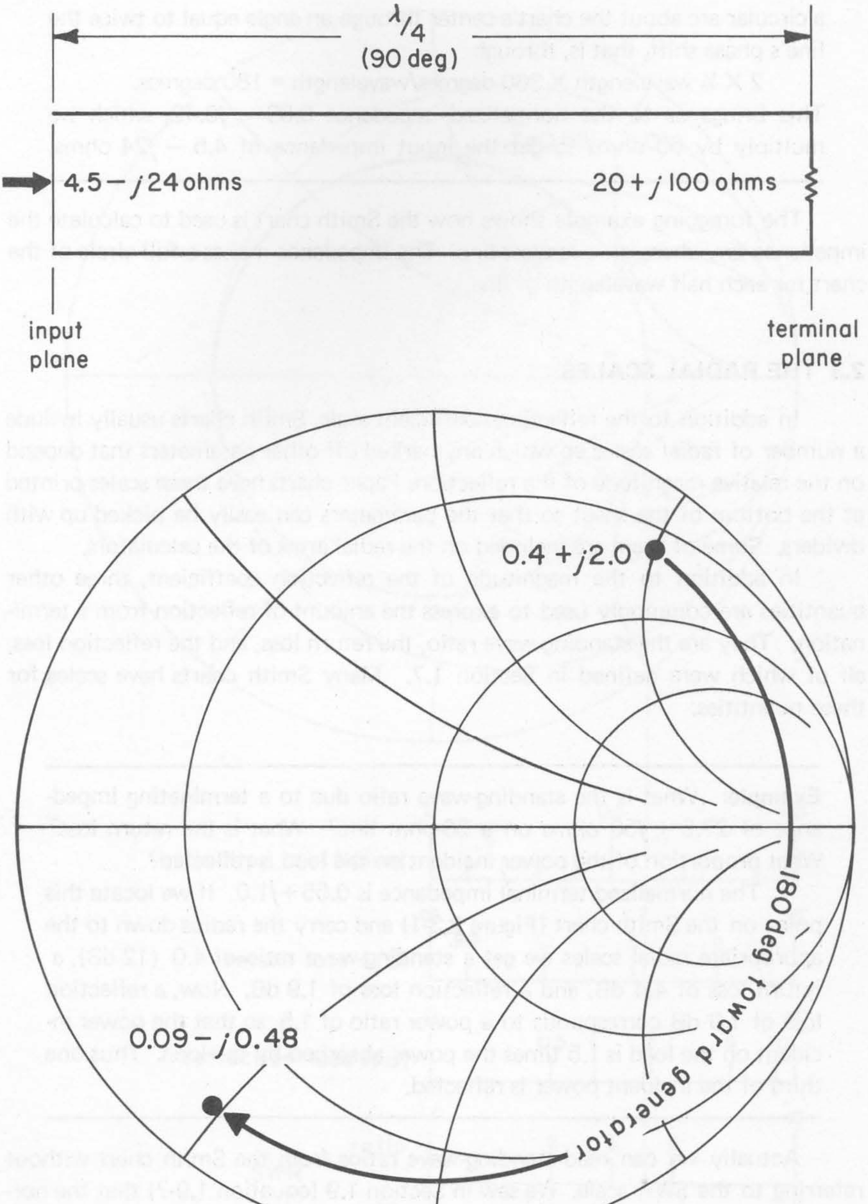


Figure 2.2-5. A terminal impedance of $20 + j100$ ohms is transformed by a quarter-wave line into an input impedance of $4.5 - j24$ ohms.

a circular arc about the chart's center through an angle equal to twice the line's phase shift, that is, through

$$2 \times \tfrac{1}{4} \text{ wavelength} \times 360 \text{ degrees/wavelength} = 180 \text{ degrees.}$$

This brings us to the normalized impedance $0.09 - j0.48$, which we multiply by 50 ohms to get the input impedance of $4.5 - j24$ ohms.

The foregoing example shows how the Smith chart is used to calculate the impedance anywhere on a lossless line. The impedance makes a full circle of the chart for each half wavelength of line.

2.3 THE RADIAL SCALES

In addition to the reflection-coefficient scale, Smith charts usually include a number of radial scales on which are marked off other parameters that depend on the relative magnitude of the reflection. Paper charts have these scales printed at the bottom of the sheet so that the parameters can easily be picked up with dividers. Some of them are included on the radial arms of the calculators.

In addition to the magnitude of the reflection coefficient, three other quantities are commonly used to express the amount of reflection from a termination. They are the standing-wave ratio, the return loss, and the reflection loss, all of which were defined in Section 1.7. Many Smith charts have scales for these quantities.

Example: What is the standing-wave ratio due to a terminating impedance of $27.5 + j50$ ohms on a 50-ohm line? What is the return loss? What proportion of the power incident on the load is reflected?

The normalized terminal impedance is $0.55 + j1.0$. If we locate this point on the Smith chart (Figure 2.3-1) and carry the radius down to the appropriate radial scales we get a standing-wave ratio of 4.0 (12 dB), a return loss of 4.4 dB, and a reflection loss of 1.9 dB. Now, a reflection loss of 1.9 dB corresponds to a power ratio of 1.5, so that the power incident on the load is 1.5 times the power absorbed by the load. Thus one third of the incident power is reflected.

Actually we can read standing-wave ratios from the Smith chart without referring to the SWR scale. We saw in Section 1.9 (equation 1.9-7) that the normalized impedance at a voltage maximum is equal to the standing-wave ratio. Thus we can read SWR's from the normalized-resistance scale along the $\theta = 0$ radial.

The scale marked "attenuation: 1-dB steps" (or "transmission loss: 1-dB steps") facilitates taking into account the effect of the line's attenuation. As we

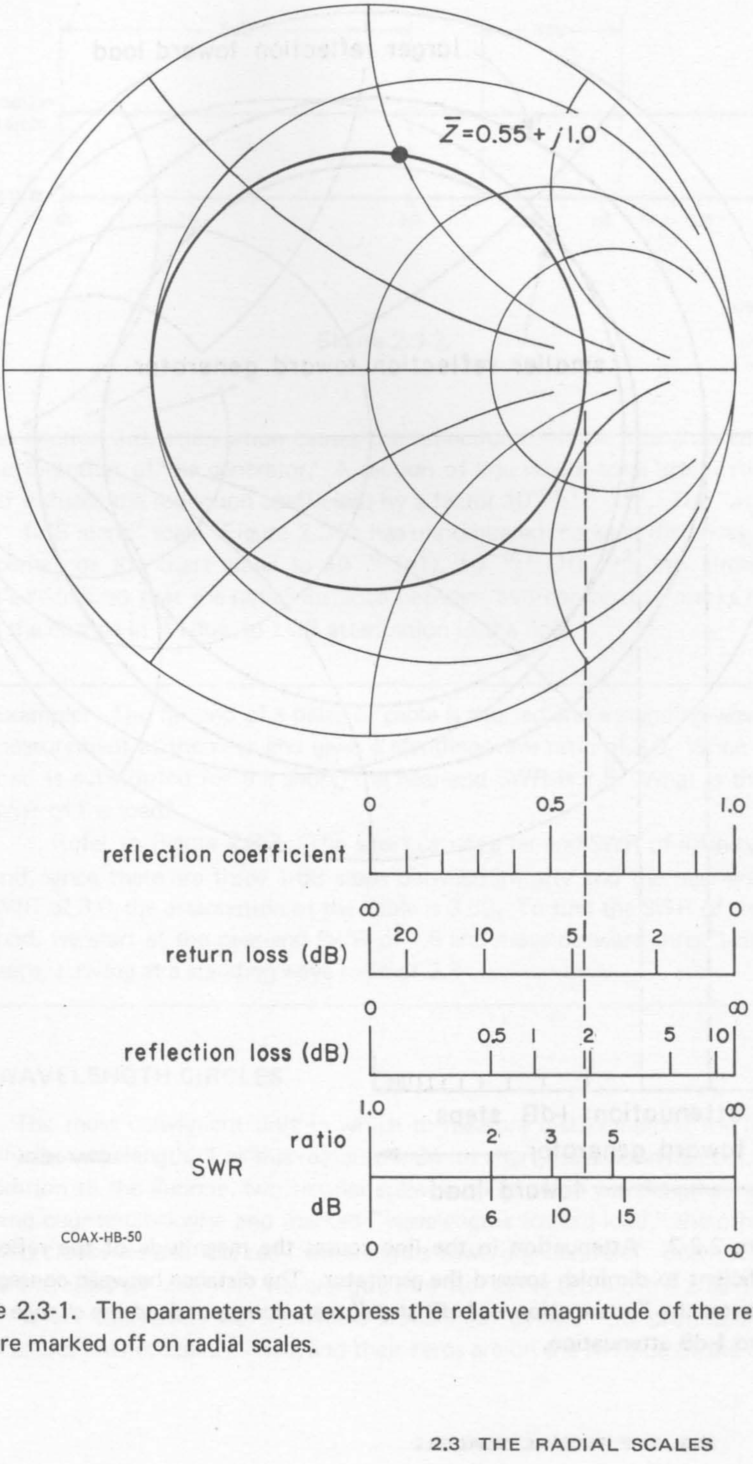


Figure 2.3-1. The parameters that express the relative magnitude of the reflection are marked off on radial scales.

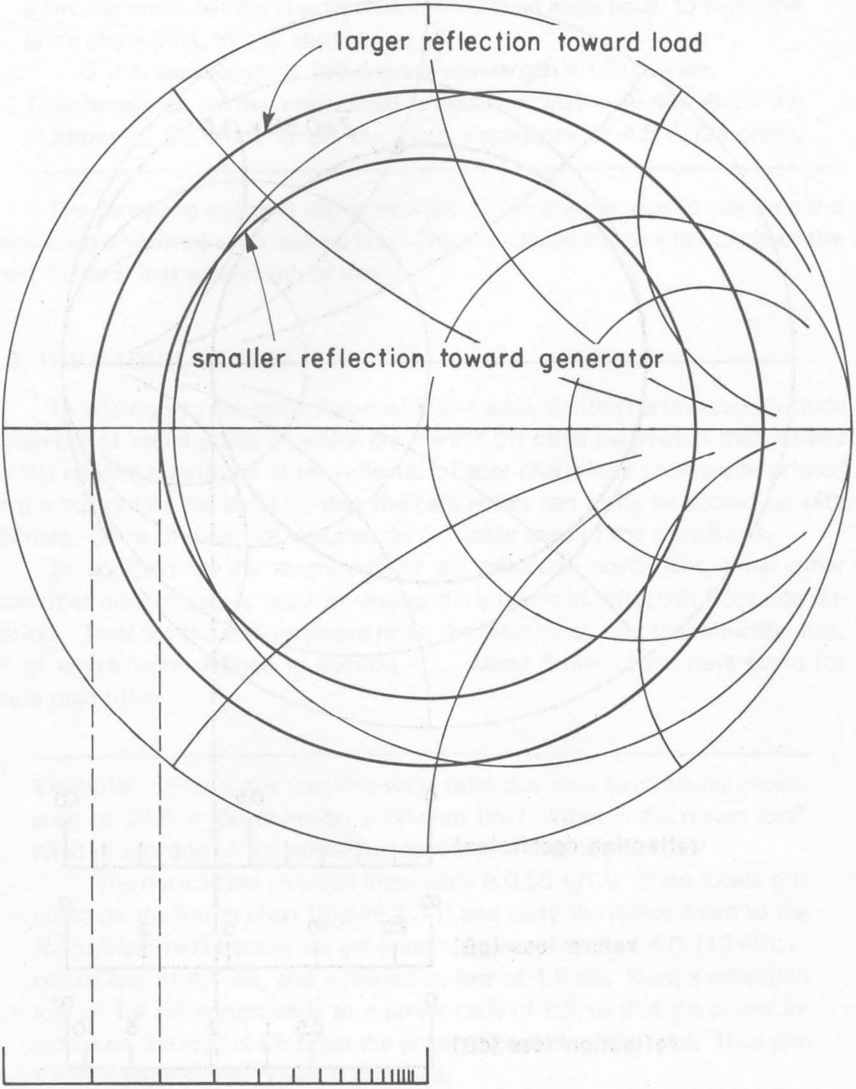


Figure 2.3-2. Attenuation in the line causes the magnitude of the reflection coefficient to diminish toward the generator. The distance between consecutive marks on the "attenuation – 1-dB steps" scale corresponds to the change in $|\Gamma|$ due to 1-dB attenuation.

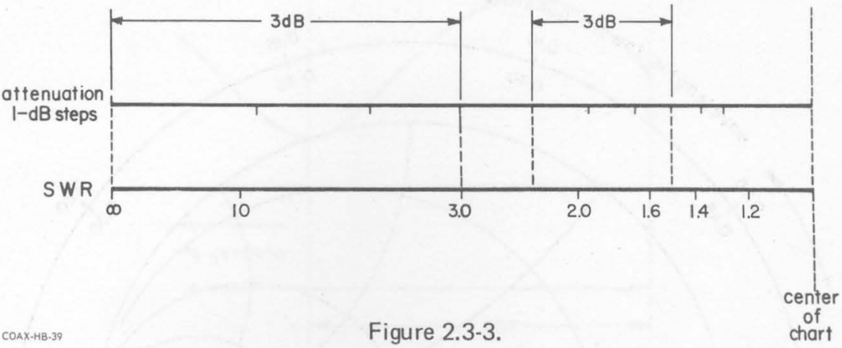


Figure 2.3-3.

saw in Section 1.8, attenuation causes the reflection coefficient to grow smaller in the direction of the generator. A section of line whose total attenuation is A(dB) reduces the reflection coefficient by a factor $10^{-0.1A\,(\mathrm{dB})}$. The "attenuation: 1-dB steps" scale (Figure 2.3-2) has unnumbered marks at distances from the center of the chart equal to 10^{-0} (=1), $10^{-0.1}$, $10^{-0.2}$, etc times the chart's radius, so that the radial distance between two consecutive marks represents the change in $|\Gamma|$ due to 1-dB attenuation in the line.

Example: The far end of a piece of cable is shorted and a standing-wave measurement at the near end gives a standing-wave ratio of 3.0. When a load is substituted for the short, the near-end SWR is 1.5. What is the SWR of the load?

Refer to Figure 2.3-3. The short causes a far-end SWR of infinity, and, since there are three 1-dB steps between infinity and the near-end SWR of 3.0, the attenuation of the cable is 3 dB. To find the SWR of the load, we start at the near-end SWR of 1.5 and move outward three 1-dB steps, arriving at a standing-wave ratio of 2.3.

2.4 WAVELENGTH CIRCLES

The most convenient unit in which to measure distance along the line is usually the wavelength. For this reason the Smith chart has around its periphery, in addition to the θ-circle, two circular scales marked off in wavelengths, one increasing counterclockwise and marked "wavelengths toward load," the other increasing clockwise and marked "wavelengths toward generator." Each of these scales increases by one half wavelength in a full circle around the chart. The Smith chart calculators have movable wavelength circles. On printed charts, these circles are necessarily fixed, and their zeros are on the left side of the chart,

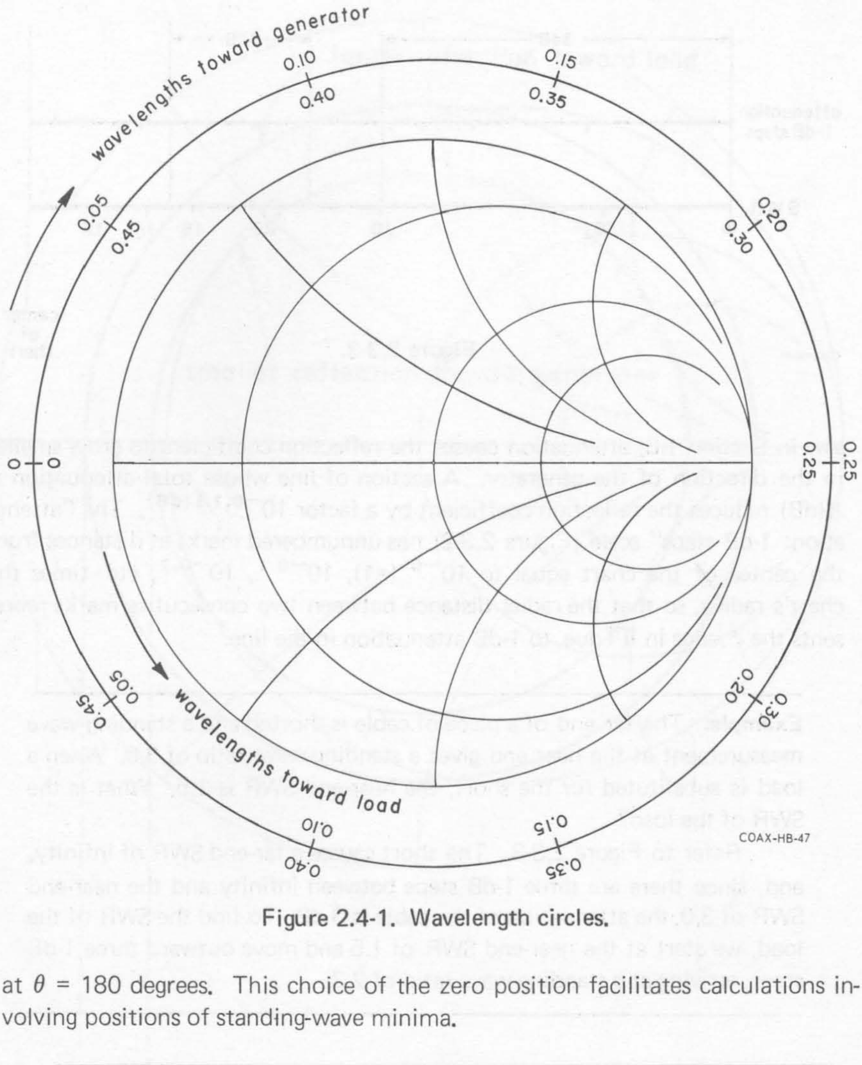


Figure 2.4-1. Wavelength circles.

at θ = 180 degrees. This choice of the zero position facilitates calculations involving positions of standing-wave minima.

Example: A voltage standing-wave minimum is found 0.30 wavelength from the termination of a lossless line. What is the angle θ_t of the terminal reflection coefficient? If the standing-wave ratio is 2.0, what is the terminal impedance?

Refer to Figure 2.4-2. A voltage minimum occurs where the reflection coefficient is 180 degrees. If we start at the θ = 180-degree radial and go around the chart 0.30 wavelength toward the load we find that θ_t = 36 degrees. If the SWR is 2.0, the normalized terminal impedance is 1.57 + j0.70.

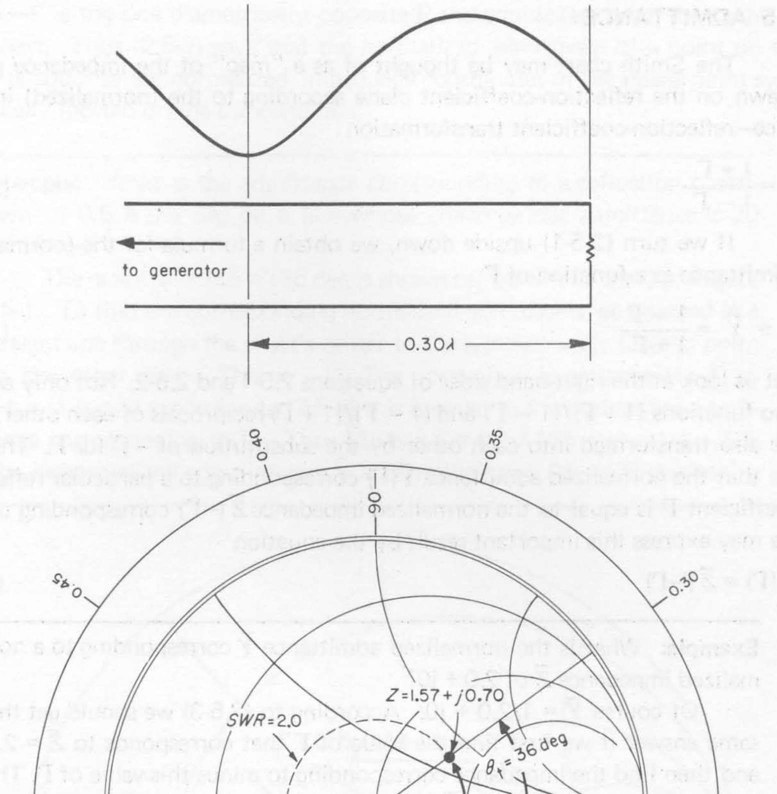


Figure 2.4-2. Using the Smith chart to determine terminal impedance from SWR and position of voltage minimum.

2.5 ADMITTANCE

The Smith chart may be thought of as a "map" of the impedance plane, drawn on the reflection-coefficient plane according to the (normalized) impedance—reflection-coefficient transformation

$$\bar{Z} = \frac{1+\Gamma}{1-\Gamma} \tag{2.5-1}$$

If we turn (2.5-1) upside down, we obtain a formula for the (normalized) admittance as a function of Γ:

$$\frac{1}{\bar{Z}} = \bar{Y} = \frac{1-\Gamma}{1+\Gamma} \tag{2.5-2}$$

Let us look at the right-hand sides of equations 2.5-1 and 2.5-2. Not only are the two functions $(1+\Gamma)/(1-\Gamma)$ and $(1-\Gamma)/(1+\Gamma)$ reciprocals of each other, they are also transformed into each other by the substitution of $-\Gamma$ for Γ. Thus we see that the normalized admittance $\bar{Y}(\Gamma)$ corresponding to a particular reflection coefficient Γ is equal to the normalized impedance $\bar{Z}(-\Gamma)$ corresponding to $-\Gamma$. We may express this important result by the equation

$$\bar{Y}(\Gamma) = \bar{Z}(-\Gamma) \tag{2.5-3}$$

Example: What is the normalized admittance $\bar{Y}$ corresponding to a normalized impedance $\bar{Z}$ of 2.0 + j0?

Of course $\bar{Y} = 1/2.0 + j0$. According to (2.5-3) we should get this same answer if we first find the value of Γ that corresponds to $\bar{Z} = 2.0$ and then find the impedance corresponding to minus this value of Γ. The value of Γ corresponding to $\bar{Z} = 2.0$ is

$$\Gamma(\text{when } \bar{Z} = 2.0) = \left(\frac{\bar{Z}-1}{\bar{Z}+1}\right)_{\bar{Z}=2.0} = \frac{2.0-1}{2.0+1} = \frac{1}{3.0}$$

The value of $\bar{Z}$ corresponding to $\Gamma = -1/3.0$ is

$$\bar{Z}(\text{when } \Gamma = -1/3.0) = \left(\frac{1+\Gamma}{1-\Gamma}\right)_{\Gamma=-1/3.0} = \frac{1-\dfrac{1}{3.0}}{1+\dfrac{1}{3.0}} = \frac{1}{2.0}$$

We note that we can derive (2.5-3) in another way. From the point Γ on the Smith chart one reaches the point $-\Gamma$ by going around the chart a quarter wavelength in either direction. Equation 2.5-3 follows from the fact that, as we saw in Section 1.8 (equation 1.8-15), normalized impedances a quarter wavelength apart on a lossless line are reciprocal.

Equation 2.5-3 shows us how to use a Smith chart with a normalized impedance grid to make conversions between impedance and admittance. The point $-\Gamma$ is the one diametrically opposite Γ and equidistant from the center of the chart. Thus (2.5-3) says that the normalized admittance of a point on the chart is equal to the normalized impedance coordinates of the image point symmetrically located on the opposite side.

Example: What is the admittance corresponding to a reflection coefficient of 0.5 $\measuredangle$135 deg on a line whose characteristic admittance is 20 millimhos?

The point $\Gamma = 0.5 \measuredangle 135$ deg is shown on the Smith chart of Figure 2.5-1. To find the corresponding normalized admittance we proceed in a straight line through the chart's center to the symmetrically located point on the other side. This is $-\Gamma$. The normalized impedance at $-\Gamma$ is $1.4 - j1.3$, and according to (2.5-3) this is equal to the normalized admittance corresponding to Γ. Thus $\overline{Y}$(when $\Gamma = 0.5 \measuredangle 135$ deg) = $1.4 - j1.3$. The unnormalized admittance is $\overline{Y} \times 20$ millimhos = $28 - j26$ millimhos.

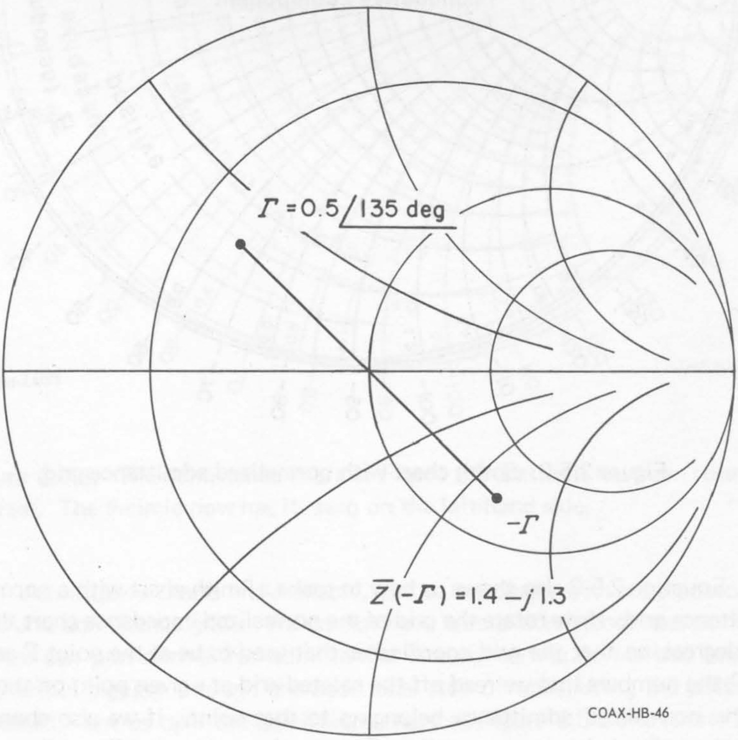


Figure 2.5-1. The normalized admittance corresponding to the point Γ is equal to the normalized impedance corresponding to the point $-\Gamma$.

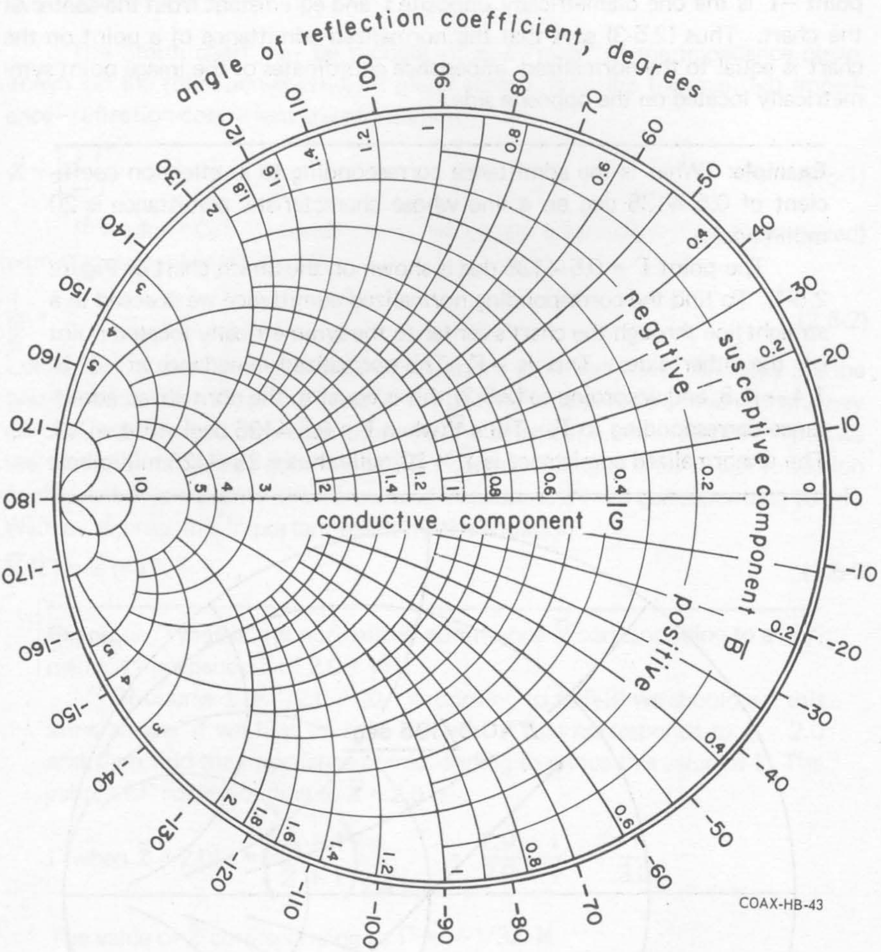


Figure 2.5-2. Smith chart with normalized admittance grid.

Equation 2.5-3 also shows us how to make a Smith chart with a normalized admittance grid. If we rotate the grid of the normalized impedance chart through 180 degrees, so that the grid coordinates that used to be at the point Γ are now at $-\Gamma$, the numbers that we read off the rotated grid at a given point on the chart are the normalized admittance belonging to that point. If we also change the labels from "resistive component" to "conductive component" and from "reactive component" to "susceptive component" we have the admittance chart shown in Figure 2.5-2.

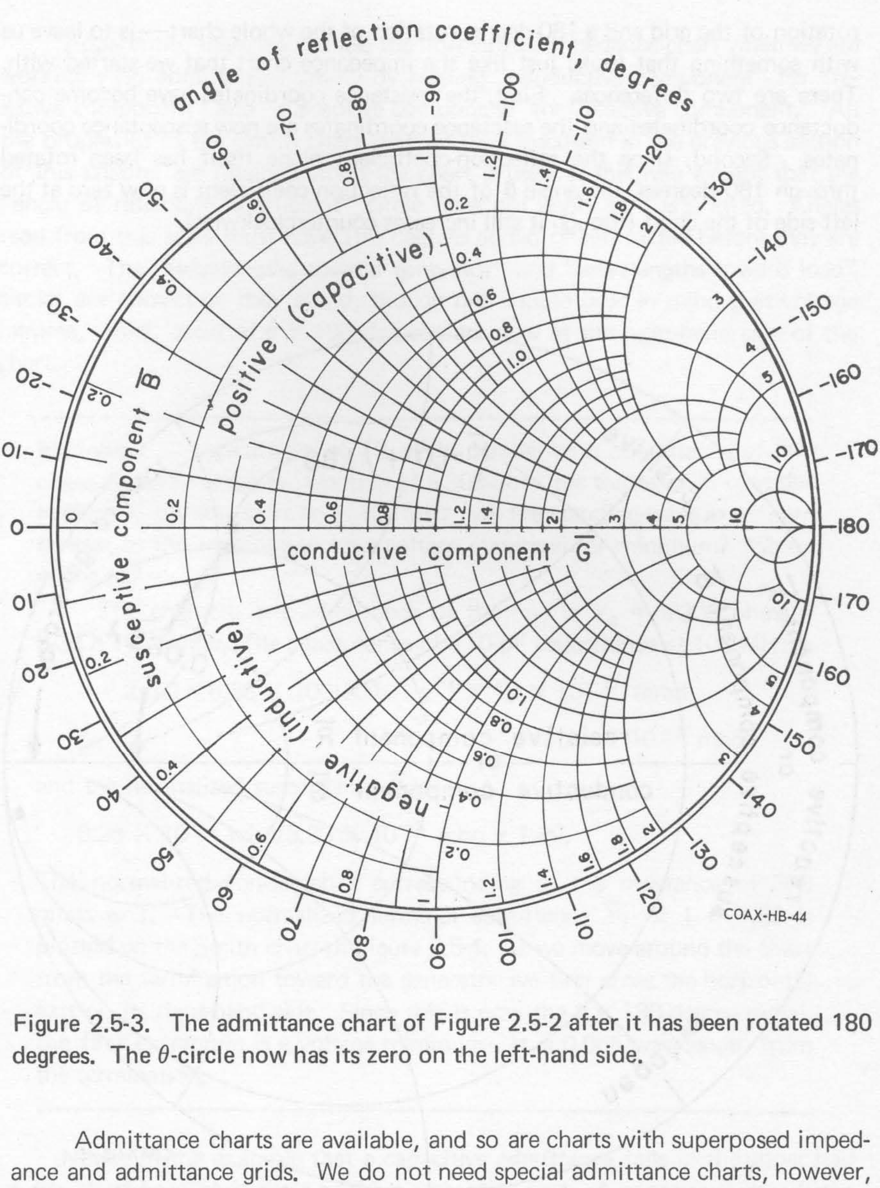


Figure 2.5-3. The admittance chart of Figure 2.5-2 after it has been rotated 180 degrees. The θ-circle now has its zero on the left-hand side.

Admittance charts are available, and so are charts with superposed impedance and admittance grids. We do not need special admittance charts, however, for we can plot normalized admittances directly on the normalized impedance grid without going through the additional step of transferring the point across the chart. Let us take the admittance chart of Figure 2.5-2 and rotate it—the whole chart this time, not just the grid—through 180 degrees. The result of this rotation is shown in Figure 2.5-3. Now, if we compare Figure 2.5-3 with Figure 2.2-1 we see that the combined effect of the two transformations—a 180-degree rotation of the grid and a 180-degree rotation of the whole chart—is to leave us with something that looks just like the impedance chart that we started with. There are two differences. First, the resistance coordinates have become conductance coordinates and the reactance coordinates are now susceptance coordinates. Second, since the reflection-coefficient plane itself has been rotated through 180 degrees, the angle θ of the reflection coefficient is now zero at the left side of the chart (though it still increases counterclockwise).

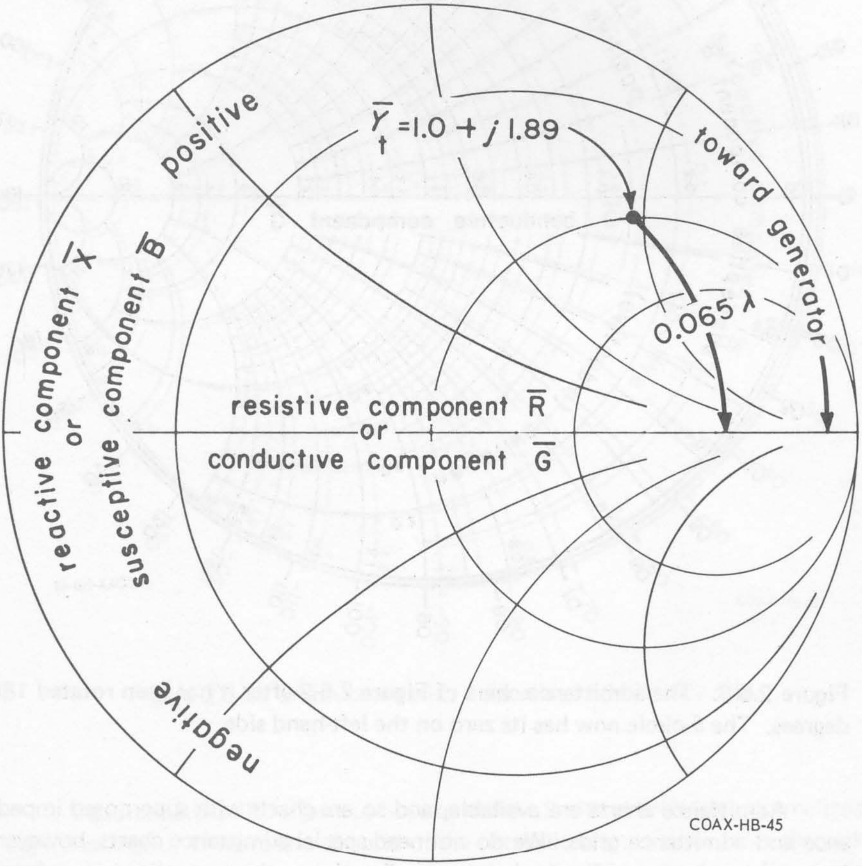


Figure 2.5-4. One can use the normalized impedance chart for admittances simply by reading "conductive component" for "resistive component" and "susceptive component" for "reactive component" and remembering that 180 degrees must be added to readings on the "angle of reflection coefficient" circle.

Apparently, then, we can use the normalized impedance chart when we are working with admittances simply by reading "conductive component" for "resistive component" and "susceptive component" for "reactive component." All the properties of the Smith chart that we have discussed in the previous section of this chapter are retained when the chart is used in this way except that the "angle of reflection coefficient" circle does not apply as it is printed. Angles read from this scale must have 180 degrees added or subtracted before they are correct. The "wavelengths toward generator" and "wavelengths toward load" circles are correct as they stand, though one should bear in mind that voltage minima, which occur at θ = 180 degrees, are now at the right-hand side of the chart.

Example: A capacitance of 10 pF in parallel with a resistance of 300 ohms constitutes the termination of a 300-ohm line that we will consider lossless. If the line is driven at 100 MHz, will the standing-wave extremum nearest to the termination be a voltage maximum or minimum? Where will it be?

The characteristic admittance of the line is Y_c = 1/300 ohms = 3.33×10^{-3} mho. The susceptance of a 10-pF capacitance at 100 MHz is

$$B = 2\pi fC = 6.28 \times 100 \times 10^{6}\ \mathrm{s}^{-1} \times 10 \times 10^{-12}\ \text{farad}$$

$$= 6.28 \times 10^{-3}\ \text{mho},$$

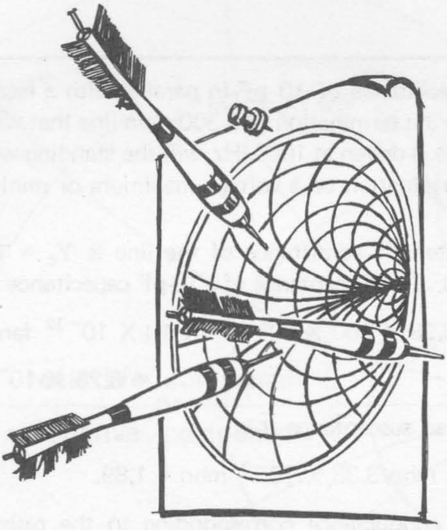
and the normalized susceptance $\overline{B}$ is

$$6.28 \times 10^{-3}\ \text{mho}/3.33 \times 10^{-3}\ \text{mho} = 1.89.$$

The normalized conductance corresponding to the resistance of 300 ohms is 1. The normalized terminal admittance $\overline{Y}_t$ of 1 + j1.89 is plotted on the Smith chart of Figure 2.5-4. As we move around the chart from the termination toward the generator we first cross the horizontal axis on its right-hand side. Since this is now the θ = 180-degree radial, the first extremum is a voltage minimum. It is 0.065 wavelength from the termination.

Notice in the example that a capacitive *admittance* falls in the upper half of the chart because it has a positive susceptive part. A capacitive *impedance,* which has a negative reactive part, would fall in the lower half.

In this chapter we have given the reader only a sketchy introduction to the most commonly used kind of Smith chart. We have not discussed the many different varieties of the chart that are in use, nor the many kinds of calculation that can be done with the chart's help. We leave it to the reader to instruct himself as the need arises.



CHAPTER 3

Two–Ports and Discontinuities

3.1 THE SCATTERING PARAMETERS

We saw in Section 1.7 that we may characterize a one-port device by 1) choosing a convenient reference plane in the associated transmission line and 2) specifying the reflection coefficient or immittance that the device presents at this reference plane. Such a characterization ignores what is actually going on inside the termination, but it allows us to predict the effect that the termination will have on the system to which it is connected. A two-port presents us with an analogous situation. We are often not concerned with the details of wave propagation inside the device itself; we simply want to know what the effect will be of inserting the two-port into the microwave system.

Now, while a single reference plane and a single complex number – a reflection coefficient or immittance – completely characterize a one-port, two reference planes and two or three or four complex numbers are needed for a complete representation of a two-port. The reader is undoubtedly familiar with some of the many sets of two-port parameters, the *y*- or *h*-parameters, for example, used in transistor circuit design at lower frequencies. But of all the two-port representations, by far the most useful at microwave frequencies is the set of four numbers called scattering parameters, or *s*-parameters. Scattering parameters were invented in 1937 by a physicist, who used them to solve a problem in nuclear physics. When physicists went to work on microwave problems during the World War II development of radar, they brought the *s*-parameters with them into electrical engineering.

Figure 3.1-1 shows a two-port "black box" with two transmission lines sticking out of it; reference planes t_1 and t_2, located in these lines, define ports 1 and 2. In the most general possible case there will be both an incident and an outgoing (scattered) wave at each port. We have written V_1^+ and V_1^- to stand for the incident and outgoing voltages at port 1, and V_2^+ and V_2^- for those at port 2.

Whereas the outgoing wave at a passive termination is due entirely to reflection of energy from the incident wave, this is not generally true at the ports of a two-port. The outgoing wave at port 1, for example, can be due partly to reflection of energy that is incident at port 1 and partly to transmission through

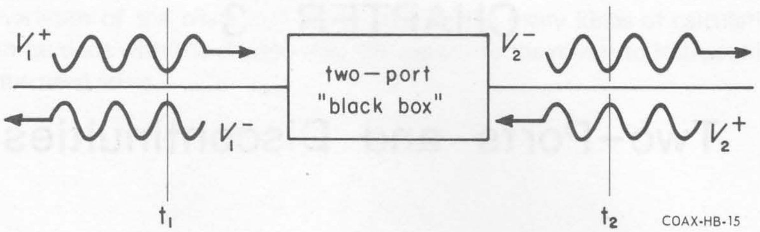


Figure 3.1-1. In general there can be both an incident and a scattered wave at each port.

the two-port of energy that is incident at port 2. For this reason we have to use not just a single coefficient but two coefficients to describe the generation of an outgoing wave at port 1:

$$V_1^- = s_{11} V_1^+ + s_{12} V_2^+ \tag{3.1-1}$$

The coefficient s_{11} accounts for reflection of some of the incident wave at port 1 and s_{12} accounts for transmission through the two-port of some of the incident wave at port 2. Likewise two more coefficients describe the generation of V_2^-:

$$V_2^- = s_{21} V_1^+ + s_{22} V_2^+ \tag{3.1-2}$$

Here s_{21} accounts for transmission of V_1^+ and s_{22} for reflection of V_2^+. The numbers s_{11}, s_{12}, s_{21}, and s_{22} are the scattering parameters.[†] Like the Γ's that describe reflections from one-ports, the scattering parameters are complex numbers, ratios of incident and outgoing voltages. As a matter of fact, it is quite appropriate to think of the s-parameters as a generalization of the notion of reflection coefficient.

Let us look a little more closely at the physical meaning of each of the scattering coefficients. First we consider the reflection parameters s_{11} and s_{22}. Suppose a reflectionless load terminates port 2, as shown in Figure 3.1-2. The in-

---

[†] Many engineers choose to use "normalized" in- and out-going voltages when they define the scattering parameters:

$$(V_1^{\pm})_{\text{normalized}} = \frac{1}{\sqrt{Z_{c1}}}\, V_1^{\pm} \qquad (V_2^{\pm})_{\text{normalized}} = \frac{1}{\sqrt{Z_{c2}}}\, V_2^{\pm}$$

where Z_{c1} and Z_{c2} are the characteristic impedances of the lines in which ports 1 and 2 are located. As the reader can see by looking at (3.1-1) and (3.1-2), this alternative choice does not affect s_{11} and s_{22}, but it does make a difference in s_{12} and s_{21}:

$$(s_{12})_{\text{normalized}} = \sqrt{\frac{Z_{c2}}{Z_{c1}}} \cdot s_{12} \qquad (s_{21})_{\text{normalized}} = \sqrt{\frac{Z_{c1}}{Z_{c2}}} \cdot s_{21}$$

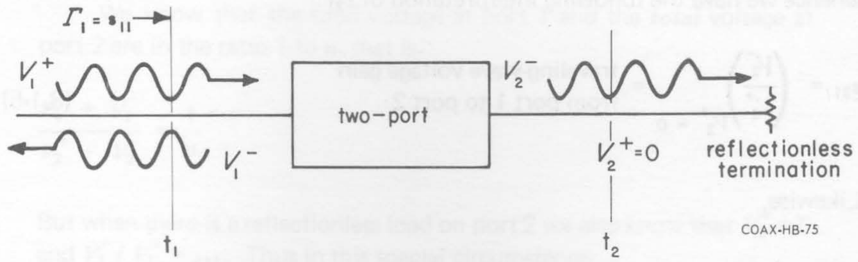


Figure 3.1-2. When port 2 sees a reflectionless termination, the reflection coefficient Γ_1 presented by port 1 is equal to s_{11}.

cident wave at port 2 is just the wave reflected from the termination, which in the present case is zero. When $V_2^+ = 0$, equation 3.1-1 becomes

$$V_1^- = s_{11} V_1^+ \qquad (V_2^+ = 0)$$

But V_1^-/V_1^+ is the reflection coefficient Γ_1 that we measure when we look into port 1, so that we have

$$s_{11} = \Gamma_1 \text{ (with reflectionless load on port 2)} \qquad (3.1\text{-}3)$$

The coefficient s_{11} is the reflection coefficient presented by port 1 when port 2 has a reflectionless termination. The same argument would obviously be valid if we put the reflectionless load on port 1 instead of port 2, so that

$$s_{22} = \Gamma_2 \text{ (with reflectionless load on port 1)} \qquad (3.1\text{-}4)$$

Since s_{11} and s_{22} are reflection coefficients measured at one port when there is no incident wave at the other, they represent reflections that are intrinsic to the two-port.

Let us once again put a reflectionless load on port 2. When there is no incident wave at port 2, the outgoing wave there is due entirely to transmission through the two-port of energy incident at port 1, and the ratio V_2^-/V_1^+ that we measure under these circumstances is the voltage gain that a traveling wave experiences as it traverses the two-port from port 1 to port 2. If $V_2^+ = 0$, equation 3.1-2 becomes

$$V_2^- = s_{21} V_1^+ \qquad (V_2^+ = 0)$$

whence we have the following interpretation of s_{21}:

$$s_{21} = \left(\frac{V_2^-}{V_1^+}\right)_{V_2^+ = 0} = \begin{array}{l}\text{traveling-wave voltage gain} \\ \text{from port 1 to port 2}\end{array} \qquad (3.1\text{-}5)$$

Likewise,

$$s_{12} = \left(\frac{V_1^-}{V_2^+}\right)_{V_1^+ = 0} = \begin{array}{l}\text{traveling-wave voltage gain} \\ \text{from port 2 to port 1}\end{array} \qquad (3.1\text{-}6)$$

Example: What are the scattering parameters of the 1-to-n turns-ratio ideal transformer shown in Figure 3.1-3?

Let the characteristic impedances of the lines in which the ports are located both be Z_c. If we put a reflectionless load on port 2, the impedance connected to the right-hand winding is Z_c, and the impedance that we measure when we look into port 1 is Z_c/n^2. Therefore the reflection coefficient presented by port 1 is

$$\Gamma_1 \,(\text{reflectionless load on port 2}) = \frac{\dfrac{1}{n^2} - 1}{\dfrac{1}{n^2} + 1} = s_{11}$$

Likewise

$$\Gamma_2 \,(\text{reflectionless load on port 1}) = \frac{n^2 - 1}{n^2 + 1} = s_{22} = -s_{11}$$

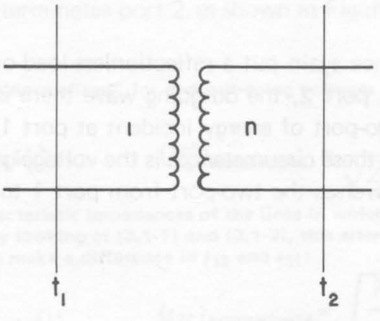


Figure 3.1-3.

We know that the *total* voltage at port 1 and the *total* voltage at port 2 are in the ratio 1 to n, that is

$$\frac{V_1^+ + V_1^-}{V_2^+ + V_2^-} = \frac{1}{n}$$

But when there is a reflectionless load on port 2 we also know that $V_2^+ = 0$ and $V_1^- / V_1^+ = s_{11}$. Thus in this special circumstance

$$\frac{V_1^+(1 + s_{11})}{V_2^-} = \frac{1}{n} \qquad (V_2^+ = 0)$$

or

$$\left(\frac{V_2^-}{V_1^+}\right)_{V_2^+ = 0} = n(1 + s_{11}) = \frac{2n}{1 + n^2} = s_{21}$$

In the other direction we of course have

$$s_{12} = \frac{2\,\frac{1}{n}}{1 + \frac{1}{n^2}} = \frac{2n}{1 + n^2}$$

Example: Let t_1 and t_2 be two reference planes in a uniform transmission line (Figure 3.1-4). The length l of line between these two planes may be regarded as a two-port. What are its s-parameters?

If the reflection coefficient that we see when we look toward the right at t_2 is zero, it will also be zero at t_1. Thus $s_{11} = 0$, and we can see

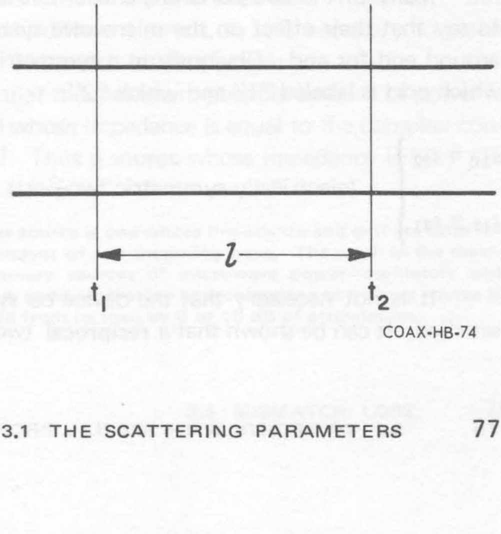


Figure 3.1-4. A section of transmission line may be regarded as a two-port with $s_{11} = s_{22} = 0$ and $s_{12} = s_{21} = 10^{-\frac{1}{20}\alpha(\text{dB/m})l}$ $\measuredangle -\beta l$.

likewise that $s_{22} = 0$. The voltage of a traveling wave is multiplied by a factor $10^{-\frac{1}{20}\alpha(\mathrm{dB/m})l}$ as the wave traverses the line segment in either direction, and its phase is shifted through an angle $-\beta l$. Therefore $s_{12} = s_{21} = 10^{-\frac{1}{20}\alpha(\mathrm{dB/m})l} \measuredangle -\beta l$.

3.2 TWO-PORTS WITH SPECIAL PROPERTIES

In the most general possible case, all four of a two-port's *s*-parameters are different, independent numbers, and we must measure or calculate each of them in order to obtain a complete description of the device. But virtually any two-port we are likely to encounter in practice will have one or more properties that tell us, even before we look at the particular device, simplifying relations among its *s*-parameters.

For instance, if the two-port is **passive**, that is, if it does not contain transistors or other sources of microwave energy, we know that we cannot get out of it energy that we do not put in. This means, for one thing, that the reflected component of the outgoing wave at either port cannot be larger in magnitude than the incident wave at that port, which we can state compactly by writing

$$|s_{11}| \leqslant 1, \quad |s_{22}| \leqslant 1 \quad \text{(passive two-port)} \tag{3.2-1}$$

It means, for another, that a transmitted wave cannot be power-amplified by the two-port. Quantitatively,

$$\frac{Z_{c1}}{Z_{c2}}\,|s_{21}|^2 \leqslant 1 \qquad \frac{Z_{c2}}{Z_{c1}}\,|s_{12}|^2 \leqslant 1 \quad \text{(passive two-port)} \tag{3.2-2}$$

where Z_{c1} and Z_{c2} are the characteristic impedances of ports 1 and 2.

Many of the two-ports one encounters are **electrically symmetric**, which is to say that their effect on the microwave system is unaltered if they are turned around end for end. Obviously in a symmetric two-port it makes no difference which port is labeled "1" and which "2".

$$\left.\begin{aligned} s_{11} &= s_{22} \qquad (3.2\text{-}3)\\ s_{12} &= s_{21} \qquad (3.2\text{-}4)\end{aligned}\right\} \quad \text{(electrically symmetric two-port)}$$

It is not necessary that the device be symmetric in order to have (3.2-4) satisfied. It can be shown that a **reciprocal** two-port has the property

$$Z_{c2}s_{12} = Z_{c1}s_{21} \quad \text{(reciprocal two-port).} \tag{3.2-5}$$

We will not give a definition of reciprocity, but, generally speaking, a device is reciprocal if 1) it is linear, i.e., its parameters do not change with the magnitudes of the fields, and 2) it contains only isotropic media—ones that have the same properties in all directions. Thus transmission lines, connectors, tuners, directional couplers, filters, attenuators, in fact any structure made of ordinary dielectrics and conductors is reciprocal. Amplifiers, mixers and ferrite isolators are non-reciprocal.

Another important class of two-ports comprises the ones that are **lossless.** By lossless we mean that they do not dissipate any energy internally. This is not the same as saying that they have zero insertion loss, since, as we shall see in Section 3.4, insertion loss can be due to reflection as well as dissipation. Sections of air-dielectric line, connectors, and tuners are devices that can be considered lossless for most purposes. In a lossless device the total incident power has to be equal to the total outgoing power. This constraint leads to the following relations:

$$\left.\begin{aligned} |s_{11}| &= |s_{22}| \\ Z_{c2}|s_{12}| = Z_{c1}|s_{21}| &= \sqrt{Z_{c1}Z_{c2}(1 - |s_{11}|^2)} \\ \arg s_{12} + \arg s_{21} &= \arg s_{11} + \arg s_{22} \pm 1, 3, 5, \cdots \times 180 \text{ deg} \end{aligned}\right\} \text{(lossless 2-port)} \qquad \begin{aligned} &(3.2\text{-}6)\\ &(3.2\text{-}7)\\ &(3.2\text{-}8) \end{aligned}$$

Notice the magnitude signs. Without them, (3.2-6) is the same as (3.2-3) and the first part of (3.2-7) is the same as (3.2-5).

3.3 MISMATCH LOSS

The term "mismatch loss" is used when one talks about the transfer of power across a single junction. Mismatch loss measures the ratio by which the power transferred from the source side to the load side of the junction falls short of the amount that would be transferred if the two sides were "matched." But the term "match" is used in several different senses, so it is necessary to make clear which "matched" condition "mismatch loss" refers to.

It is a well-known theorem that the maximum possible amount of power is extracted from a source by a load whose impedance is equal to the complex conjugate of the source impedance.† Thus a source whose impedance is 50 + *j*25

---

†Provided the source is linear. A linear source is one whose impedance and emf are both independent of the load, that is, independent of any incoming wave. The catch in the maximum-power theorem is that the primary sources of microwave power—oscillators and amplifiers—are not in general linear. A good laboratory approximation to a linear source is provided by an oscillator that is isolated from its load by 6 or 10 dB of attenuation.

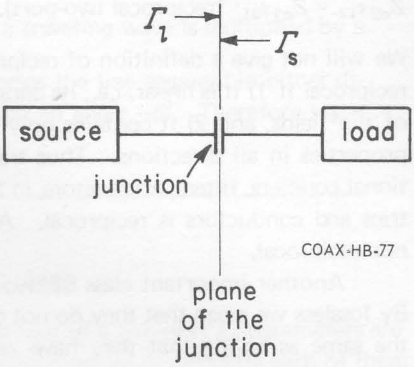


Figure 3.3-1. The amount of power that the source delivers to the load depends on the reflection coefficients Γ_s of the source and Γ_l of the load.

ohms will deliver maximum power to a load impedance of 50 – *j*25 ohms. If the source and load impedances are complex conjugates of each other, so are the source and load reflection coefficients. Thus we may state the maximum power transfer theorem by saying that maximum power is extracted from a source by a load whose reflection coefficient is the complex conjugate of the source reflection coefficient.

$$\left.\begin{aligned} Z_l &= \text{comp. conj. } Z_s \\ \Gamma_l &= \text{comp. conj. } \Gamma_s \end{aligned}\right\} \quad \text{condition for maximum power transfer} \qquad (3.3\text{-}1)$$

The maximum amount of power that the source can deliver is called the source's **available power.**

Example: A source and load are connected at a junction whose characteristic impedance is 50 ohms. When the load is reflectionless ($\Gamma_l = 0$), the power delivered to the load is 0.02 watt. A measurement of the reflection coefficient Γ_s of the source yields a value of $0.5 \measuredangle + 30$ deg. How much power will the source deliver to a conjugate load?

When the source is terminated in a reflectionless load there is only one wave crossing the source-load junction. This is the primary wave emitted by the source. It is totally absorbed by the load, and we calculate that it must have an amplitude of

$$\sqrt{0.02 \text{ watt} \times 50 \text{ ohms}} = 1.0 \text{ volt rms}$$

Let us see what happens when the load is a conjugate match to the source: $\Gamma_l = \text{comp. conj. } \Gamma_s = 0.5 \measuredangle - 30$ deg. Now the 1-volt primary wave is partially reflected by the load. The amplitude of this reflection is 0.5 volt and its phase is −30 degrees with respect to the incident primary wave. The reflected wave in turn experiences a partial reflection from the source, so there is a second wave incident on the load whose amplitude is 0.25 volt $\measuredangle$ 0 deg. There is a second reflection from the load of amplitude 0.125 volt $\measuredangle$ −30 deg, and so on. The total voltage incident on the load is

$$(1 + 0.25 + 0.0625 + \cdot\ \cdot\ \cdot) \text{ volts} \measuredangle 0 \text{ deg.} \ = \ 1.33 \text{ volts} \measuredangle 0 \text{ deg.}$$

The power incident on the load is $(1.33 \text{ volts})^2/50 \text{ ohms} = 0.035$ watt. The total voltage reflected from the load is

$$(0.5 + 0.125 + \cdot\ \cdot\ \cdot) \text{ volt } \measuredangle -30 \text{ deg} = 0.67 \text{ volt} \measuredangle -30 \text{ deg}$$

and the reflected power is

$$(0.67 \text{ volt})^2/50 \text{ ohms} \ = \ 0.0090 \text{ watt}$$

The power absorbed by the load is thus 0.035 − 0.0090 = 0.026 watt. This is the available power of the source.

This example explains the apparent paradox that a reflecting load can absorb more power than a nonreflecting one if the source is not reflectionless.

One kind of mismatch loss is the **conjugate mismatch loss.** This is the loss that occurs because a nonconjugate load does not extract the available power from the source.

$$\text{conjugate-mismatch-loss ratio} \ = \ \frac{\text{source's available power}}{\text{power delivered to load}} \qquad (3.3\text{-}2)$$

The formula for the conjugate-mismatch-loss ratio as a function of the source and load reflection coefficients is

$$\text{conjugate-mismatch-loss ratio} \ = \ \frac{|1 - \Gamma_s \Gamma_l|^2}{(1 - |\Gamma_s|^2)\,(1 - |\Gamma_l|^2)} \qquad (3.3\text{-}3)$$

One rarely knows the angles of Γ_s and Γ_l, and without them one cannot calculate the numerator of (3.3-3). However, even if just the magnitudes of Γ_s and Γ_l are known, (3.3-3) can still be used to calculate the largest and smallest values that the conjugate mismatch loss can have. For given values of $|\Gamma_s|$ and

$|\Gamma_l|$, the right-hand side of (3.3-3) is maximum when arg Γ_s + arg Γ_l = 180 deg and minimum when arg Γ_s + arg Γ_l = 0 deg. Thus, for given magnitudes of source and load reflections,

$$\left.\begin{array}{l} \begin{array}{l}\text{maximum}\\ \text{conjugate-}\\ \text{mismatch-}\\ \text{loss ratio}\end{array} = \dfrac{(1+|\Gamma_s|\,|\Gamma_l|)^2}{(1-|\Gamma_s|^2)(1-|\Gamma_l|^2)} = 1+\dfrac{(r_s r_l-1)^2}{4r_s r_l} \\ \\ \begin{array}{l}\text{minimum}\\ \text{conjugate-}\\ \text{mismatch-}\\ \text{loss ratio}\end{array} = \dfrac{(1-|\Gamma_s|\,|\Gamma_l|)^2}{(1-|\Gamma_s|^2)(1-|\Gamma_l|^2)} = 1+\dfrac{(r_s-r_l)^2}{4r_s r_l} \end{array}\right\} \tag{3.3-4}$$

Several features of these formulas are worth noting. 1) It does not matter which reflection is associated with the source and which with the load, since the formulas are symmetrical with respect to source and load. 2) If one side of the junction is reflectionless, the maximum and minimum losses are equal. 3) If the magnitudes of the two reflections are equal, the minimum mismatch loss is zero (ratio is unity). This is no surprise, since if $|\Gamma_s| = |\Gamma_l|$ there could be a conjugate match if arg $\Gamma_s = -$arg Γ_l.

Example: What are the maximum and minimum conjugate-mismatch-loss ratios when the source SWR is 1.6 and the load SWR is 1.2?

$$\begin{array}{l}\text{maximum}\\ \text{conjugate-}\\ \text{mismatch-}\\ \text{loss ratio}\end{array} = 1+\frac{(1.6\times 1.2-1)^2}{4\times 1.6\times 1.2} = 1.11\ (0.45\,\text{dB})$$

$$\begin{array}{l}\text{minimum}\\ \text{conjugate-}\\ \text{mismatch-}\\ \text{loss ratio}\end{array} = 1+\frac{(1.6-1.2)^2}{4\times 1.6\times 1.2} = 1.021\quad (0.090\,\text{dB})$$

For many reasons a microwave system is usually made as reflectionless as possible, and consequently electrical specifications of microwave components and equipment are usually stated with respect to reflectionless rather than conjugate terminations.

Example: The specified output power of a signal generator is conventionally assumed to be the power that the generator would deliver to a reflectionless load. How large a mismatch error do we make if we use a power meter whose SWR is 1.3 to measure the output power of a signal generator whose SWR is 1.2?

From (3.3-2) we have

$$\frac{\begin{array}{c}\text{power that source would}\\ \text{deliver to a reflectionless load}\end{array}}{\begin{array}{c}\text{power delivered to}\\ \text{actual load}\end{array}} = \frac{\begin{array}{c}\text{conjugate-mismatch-loss ratio}\\ \text{of source and actual load}\end{array}}{\begin{array}{c}\text{conjugate-mismatch-loss ratio}\\ \text{of source and reflectionless load}\end{array}} \tag{3.3-5}$$

As we saw above, we can calculate an upper and a lower limit to the numerator on the right of (3.3-5). Using the two relations in (3.3-4), we have

$$\begin{array}{c}\text{maximum}\\ \text{conjugate-mismatch-}\\ \text{loss ratio of source}\\ \text{and actual load}\end{array} = 1 + \frac{(1.2 \times 1.3 - 1)^2}{4 \times 1.2 \times 1.3} = 1.05$$

and

$$\begin{array}{c}\text{minimum}\\ \text{conjugate-mismatch-}\\ \text{loss ratio of source}\\ \text{and actual load}\end{array} = 1 + \frac{(1.2 - 1.3)^2}{4 \times 1.2 \times 1.3} = 1.0016$$

The denominator on the right of (3.3-5) has a definite value that is independent of the angle of the source's reflection coefficient. When $r_l = 1$, either of the formulas in (3.3-4) gives

$$\begin{array}{c}\text{conjugate-mismatch-}\\ \text{loss ratio of source}\\ \text{and reflectionless load}\end{array} = 1 + \frac{(1.2 - 1)^2}{4 \times 1.2} = 1.0083$$

Thus the power that the generator would deliver to a reflectionless load is somewhere between 1.05/1.0083 = 1.041 and 1.0016/1.0083 = (1 − 0.0067) times the measured power, that is, between 4.1 percent above and 0.67 percent below the measured power.

Before we leave the subject of the losses that are associated with a single junction, let us briefly call attention to the special case in which the source is reflectionless. When there is no reflection from the source, the wave incident on the load is always just the primary wave emitted by the source, and the maximum power transfer to the load occurs when the load also is nonreflecting. Any power that is reflected from the load represents a loss in load power with respect to the source's available power, so that in this special case the conjugate mismatch loss is equal to the **reflection loss**, which we defined in Section 1.7, Chapter 1.

$$\text{reflection-loss ratio} = \frac{1}{1-|\Gamma_l|^2} = 1 + \frac{(r_l-1)^2}{4r_l} = \begin{array}{l}\text{conjugate-mismatch-loss ratio}\\ \text{when source is nonreflecting}\end{array} \tag{3.3-6}$$

It is important to realize that the reflection loss and conjugate mismatch loss are the same only when the source is reflectionless. When this is not the case, reflection loss has very little meaning, since in general maximum power transfer occurs when there is a reflection from the load.

3.4 INSERTION LOSS AND ATTENUATION

Now we turn to the rather confused subject of losses associated with two-ports. The confusion arises because there is not general agreement on the meanings of the terms "insertion loss" and "attenuation."

Insertion loss has its origins in low-frequency filter theory. It is usually[†] defined by

$$\text{insertion-loss ratio} = \frac{\begin{array}{c}\text{power delivered to load}\\ \text{connected directly to source}\end{array}}{\begin{array}{c}\text{power delivered to load}\\ \text{when two-port is inserted}\end{array}} \tag{3.4-1}$$

A problem with insertion loss at microwave frequencies is that, unless the connecting hardware at port 1 is the same as that at port 2, the concept of insertion loss is in principle meaningless because the source and load cannot be connected directly. Perhaps we should also point out that the insertion loss is negative (ratio less than unity) when the two-port improves the power transfer between source and load.

[†] **But not always. "Insertion loss" is sometimes used with the meaning of "transducer loss" and sometimes with the meaning of "characteristic insertion loss," both of which are defined below.**

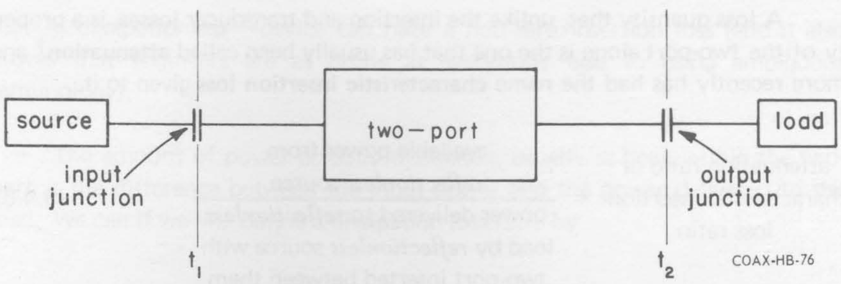


Figure 3.4-1.

The difficulties with insertion loss are obviated in another two-port loss concept. **Transducer loss**, whose definition is universally agreed upon, compares the power delivered to the load when the two-port is inserted with the available power of the source.

$$\text{transducer-loss ratio} = \frac{\text{source's available power}}{\text{power delivered to load when two-port is inserted}} \tag{3.4-2}$$

The transducer loss clearly cannot be negative (loss ratio less than unity), and the definition is valid regardless of the hardware at the ports.

In general the insertion loss and the transducer loss depend in a complicated way on both the source and load as well as on the two-port itself. The formulas for these two losses, derivations of which the interested reader will find in Chapter 4, are

$$\text{insertion-loss ratio} = \left(\frac{Z_{c2}}{Z_{c1}}\ \frac{1}{|s_{21}|^2}\right)\left(\frac{|(1-s_{11}\Gamma_s)\,(1-s_{22}\Gamma_l)-s_{12}s_{21}\Gamma_s\Gamma_l|^2}{|1-\Gamma_s\Gamma_l|^2}\right) \tag{3.4-3}$$

$$\text{transducer-loss ratio} = \left(\frac{Z_{c2}}{Z_{c1}}\ \frac{1}{|s_{21}|^2}\right)\left(\frac{|(1-s_{11}\Gamma_s)\,(1-s_{22}\Gamma_l)-s_{12}s_{21}\Gamma_s\Gamma_l|^2}{(1-|\Gamma_s|^2)(1-|\Gamma_l|^2)}\right) \tag{3.4-4}$$

We call the reader's attention to the term $s_{12}s_{21}\Gamma_s\Gamma_l$. This interaction term arises because of multiple reflections back and forth within the two-port. If *either* of the terminations is nonreflecting, or if the backward transmission s_{12} is zero, the interaction term is zero. Equations 3.4-3 and 3.4-4 differ only in the denominators within the second sets of parentheses. Comparison of these donominators reveals that

$$\text{transducer-loss ratio} = \text{insertion-loss ratio} \times \text{conjugate-mismatch-loss ratio of source and load} \tag{3.4-5}$$

A loss quantity that, unlike the insertion and transducer losses, is a property of the two-port alone is the one that has usually been called **attenuation**[†] and more recently has had the name **characteristic insertion loss** given to it.

$$\begin{array}{c}\text{attenuation ratio or}\\ \text{characteristic-insertion-}\\ \text{loss ratio}\end{array} = \frac{\begin{array}{c}\text{available power from}\\ \textit{reflectionless}\text{ source}\end{array}}{\begin{array}{c}\text{power delivered to }\textit{reflectionless}\\ \text{load by }\textit{reflectionless}\text{ source with}\\ \text{two-port inserted between them}\end{array}} \tag{3.4-6}$$

Comparison with (3.4-2) reveals that attenuation is the same as transducer loss when the source and load are reflectionless. If the source and load are directly mate-able, and if the source and load ports have the same characteristic impedance, the attenuation is also equal to the insertion loss when the source and load are reflectionless.

$$\begin{array}{c}\text{attenuation or}\\ \text{characteristic}\\ \text{insertion loss}\end{array} = \begin{array}{c}\text{transducer loss}\\ \text{when } \Gamma_s = \Gamma_l = 0\end{array} = \begin{array}{c}\text{insertion loss}\\ \text{when } \Gamma_s = \Gamma_l = 0\end{array} \tag{3.4-7}$$

In contrast to the complicated formulas (3.4-3) and (3.4-4), the characteristic insertion loss depends in a very simple way upon quantities that are properties of the two-port alone. If we put $\Gamma_s = \Gamma_l = 0$ in either (3.4-3) or (3.4-4) we have

$$\begin{array}{c}\text{attenuation ratio or}\\ \text{characteristic-insertion-}\\ \text{loss ratio}\end{array} = \frac{Z_{c2}}{Z_{c1}} \cdot \frac{1}{|s_{21}|^2} \tag{3.4-8}$$

The reader might like to verify that (3.4-8) also follows from the definition (3.4-6) of the characteristic insertion loss and the definition (3.1-5) of the forward transmission coefficient s_{21}.

Notice that none of the losses we have defined thus far is a loss in the sense of dissipation. Each is a loss only in the sense of a comparison with some hypothetical coupling of source and load. To say that the insertion loss of a component of a microwave system is 3 dB does not mean that it dissipates half the power that is delivered to it. It means that when the component is put into the system the power arriving at its load side is cut in half. This ambiguity in the meaning of the word "loss" gives rise to the apparent paradox that a lossless—

[†] The trouble with the term "attenuation" is that it is used to designate almost any comparison of power levels.

that is, dissipationless—device can have a non-zero insertion loss (and it also proves that engineers are as prone as everybody else to using ambiguous terminology).

The amount of power actually dissipated, usually as heat, within the two-port is the difference between the input power and the power delivered to the load. We can if we like define a **dissipation-loss** ratio by

$$\text{dissipation-loss ratio} = \frac{\begin{matrix}\text{power delivered}\\ \text{to two-port}\end{matrix}}{\begin{matrix}\text{power delivered}\\ \text{to load}\end{matrix}} \qquad (3.4\text{-}9)$$

and we note that this is just the reciprocal of the familiar quantity **efficiency.** The reflection from the load influences the amount of power that is dissipated, so that the dissipation loss is a function of Γ_l as well as the s-parameters:

$$\begin{matrix}\text{dissipation-loss}\\ \text{ratio}\end{matrix} = \frac{Z_{c2}}{Z_{c1}} \; \frac{1}{|s_{21}|^2} \times \frac{|1-s_{22}\Gamma_l|^2 - |(s_{12}s_{21} - s_{11}s_{22})\,\Gamma_l + s_{11}|^2}{1-|\Gamma_l|^2} \qquad (3.4\text{-}10)$$

It is possible to separate the transducer loss into a part that is due to mismatch at the input junction and a part that is due to dissipation:

$$\begin{matrix}\text{transducer-}\\ \text{loss ratio}\end{matrix} = \begin{matrix}\text{conjugate-mismatch-loss}\\ \text{ratio at input junction}\end{matrix} \times \begin{matrix}\text{dissipation-}\\ \text{loss ratio}\end{matrix} \qquad (3.4\text{-}11)$$

This is merely a fairly obvious identity, and not a particularly useful one at that, because both the mismatch loss and the dissipation loss are in general involved functions. There is one case, however, in which it yields a simple and instructive relation. When $\Gamma_s = \Gamma_l = 0$, (3.4-11) becomes

$$\begin{matrix}\text{characteristic-}\\ \text{insertion-}\\ \text{loss (attenuation)}\\ \text{ratio}\end{matrix} = \underbrace{\frac{1}{1-|s_{11}|^2}}_{\substack{\text{reflection-}\\ \text{loss ratio}\\ \text{at input}}} \cdot \underbrace{\frac{Z_{c2}}{Z_{c1}} \cdot \frac{1-|s_{11}|^2}{|s_{21}|^2}}_{\substack{\text{dissipation-loss}\\ \text{ratio}}} \qquad (\Gamma_s = \Gamma_l = 0) \qquad (3.4\text{-}12)$$

This equation shows that the characteristic insertion loss is due to reflection or dissipation or both. It is important for the reader to bear in mind that the expressions given in (3.4-12) for the reflection- and dissipation-loss ratios are valid only when $\Gamma_s = \Gamma_l = 0$.

Formulas, such as (3.4-3) and (3.4-4), that express the various loss quantities in terms of Γ_s, Γ_l, and the s-parameters of the two-port are usually of no use for computation. Apart from the fact that these formulas are extremely complicated, we usually do not know the angles of the complex numbers that they depend upon. As an example of the sort of computational use to which this sort of equation can be put, we shall discuss the mismatch error that arises in the insertion measurement of attenuation (characteristic insertion loss).

A practical way to measure attenuation is to insert the unknown component into a nominally matched system and record the decrease in transmitted power. Of course what this method actually measures is insertion loss, and when the result is taken as the attenuation there is an error that is due to the mismatch that inevitably exists. Comparison of equations 3.4-8 and 3.4-3 shows that

$$\frac{\begin{array}{c}\text{characteristic-}\\ \text{insertion-loss}\\ \text{(attenuation) ratio}\end{array}}{\begin{array}{c}\text{insertion-loss}\\ \text{ratio}\end{array}} = \frac{|(1 - s_{11}\Gamma_s)\,(1 - s_{22}\Gamma_l) - s_{12}s_{21}\Gamma_s\Gamma_l|^2}{|1 - \Gamma_s\Gamma_l|^2} \qquad (3.4\text{-}13)$$

We can also express this ratio in terms of the input reflection coefficient Γ_1 that the device presents at port 1 or the output reflection coefficient Γ_2 that it presents at port 2:

$$\frac{\begin{array}{c}\text{characteristic-}\\ \text{insertion-loss}\\ \text{(attenuation) ratio}\end{array}}{\begin{array}{c}\text{insertion-loss}\\ \text{ratio}\end{array}} = \frac{|1 - \Gamma_1\Gamma_s|^2\;|1 - s_{22}\Gamma_l|^2}{|1 - \Gamma_s\Gamma_l|^2} = \frac{|1 - s_{11}\Gamma_s|^2\;|1 - \Gamma_2\Gamma_l|^2}{|1 - \Gamma_s\Gamma_l|^2} \qquad (3.4\text{-}14)$$

The reflection coefficients Γ_1 and Γ_2 are measured with the respective opposite terminations in place (we shall go into the subject of input and output reflection coefficients in the next section). If we know the magnitudes but not the angles of the various complex quantities in (3.4-13) and (3.4-14) we can still calculate the limits of the mismatch error.[†]

[†] **Let us warn the reader that this problem is sometimes treated incorrectly in the literature. One can find discussions that ignore the fact that dissipation loss depends on the load, or assume erroneously that insertion loss is the sum of either dissipation loss or attenuation and a mismatch loss at each port. These mistakes have led to wrong answers in publications that have an obligation to be more reliable.**

Example: What can we say about the attenuation (characteristic insertion loss) of an attenuator if the measured insertion loss is 20 dB, the measuring system SWR is 1.05 in either direction, and the insertion SWR of the attenuator is 1.15 at each end?

To begin with, let us note that considerable simplification of this problem results from the fact that 20 dB of attenuation between the ports are enough to swamp the interaction between the small mismatches at each end of the attenuator. Thus, as the reader can easily verify, the interaction term $s_{12}s_{21}\Gamma_s\Gamma_l$ in (3.4-13) is negligible compared with the rest of the numerator. We then have the approximation

$$\frac{\begin{array}{c}\text{characteristic-}\\ \text{insertion-loss}\\ \text{(attenuation) ratio}\end{array}}{\begin{array}{c}\text{insertion-loss}\\ \text{ratio}\end{array}} \doteq \frac{|1 - s_{11}\Gamma_s|^2\ |1 - s_{22}\Gamma_l|^2}{|1 - \Gamma_s\Gamma_l|^2} \qquad \left(\begin{array}{c}\text{neglecting}\\ \text{interaction}\end{array}\right) \tag{3.4-15}$$

The maximum value of (3.4-15) would occur if the quantities $s_{11}\Gamma_s$ and $s_{22}\Gamma_l$ both had angles of 180 degrees and the quantity $\Gamma_s\Gamma_l$ had an angle of 0 degrees:

$$\begin{array}{c}\text{maximum}\\ \text{value}\\ \text{of}\end{array}\left(\frac{\begin{array}{c}\text{characteristic-}\\ \text{insertion-loss}\\ \text{(attenuation) ratio}\end{array}}{\begin{array}{c}\text{insertion-loss}\\ \text{ratio}\end{array}}\right) \doteq \frac{(1 + |s_{11}||\Gamma_s|)^2\ (1 + |s_{22}||\Gamma_l|)^2}{(1 - |\Gamma_s||\Gamma_l|)^2}$$

The minimum value would occur if $s_{11}\Gamma_s$ and $s_{22}\Gamma_l$ had angles of 0 degrees and $\Gamma_s\Gamma_l$ had an angle of 180 degrees:

$$\begin{array}{c}\text{minimum}\\ \text{value}\\ \text{of}\end{array}\left(\frac{\begin{array}{c}\text{characteristic-}\\ \text{insertion-loss}\\ \text{(attenuation) ratio}\end{array}}{\begin{array}{c}\text{insertion-loss}\\ \text{ratio}\end{array}}\right) \doteq \frac{(1 - |s_{11}||\Gamma_s|)^2\ (1 - |s_{22}||\Gamma_l|)^2}{(1 + |\Gamma_s||\Gamma_l|)^2}$$

In the present case the numbers we need are

$$|s_{11}| = |s_{22}| = \frac{1.15 - 1}{1.15 + 1} = 0.070$$

and

$$|\Gamma_s| = |\Gamma_l| = \frac{1.05 - 1}{1.05 + 1} = 0.024$$

and the maximum and minimum values of (3.4-15) are

$$\begin{matrix}\text{maximum} \\ \text{value}\end{matrix} = \frac{(1 + 0.070 \times 0.024)^2 \, (1 + 0.070 \times 0.024)^2}{(1 - 0.024 \times 0.024)^2}$$

$$= \frac{(1 + 0.00168)^4}{(1 - 0.000576)^2}$$

(we now make labor-saving use of the binomial theorem)

$$\doteq 1 + 4 \times 0.00168 + 2 \times 0.000576 = 1 + 0.00787$$

and

$$\begin{matrix}\text{minimum} \\ \text{value}\end{matrix} = \frac{(1 - 0.070 \times 0.024)^2 \, (1 - 0.070 \times 0.024)^2}{(1 + 0.024 \times 0.024)^2}$$

$$\doteq 1 - 4 \times 0.00168 - 2 \times 0.000576 = 1 - 0.00787$$

Thus the true attenuation is between 0.79 percent above and 0.79 percent below the measured insertion loss, or, in other words, mismatch causes an error of ± 0.79 percent or ± 0.034 dB.

3.5 INPUT AND OUTPUT REFLECTION COEFFICIENTS

We have seen that when port 2 has a nonreflecting termination the reflection coefficient that one sees when one looks into port 1 is s_{11}. The standing-wave ratio corresponding to s_{11} is called the **insertion SWR.**

$$\text{insertion SWR at port 1} = \frac{1 + |s_{11}|}{1 - |s_{11}|} = \begin{matrix}\text{SWR at port 1 with} \\ \text{nonreflecting load on port 2}\end{matrix} \qquad (3.5\text{-}1)$$

Likewise

$$\text{insertion SWR at port 2} = \frac{1 + |s_{22}|}{1 - |s_{22}|} = \begin{matrix}\text{SWR at port 2 with} \\ \text{nonreflecting load on port 1}\end{matrix} \qquad (3.5\text{-}2)$$

Example: A connector pair is found to have an insertion SWR of 1.1. What is its attenuation (characteristic insertion loss)?

Connectors can normally be considered lossless (dissipationless), and two-ports in this category satisfy

$$|s_{11}| = |s_{22}|$$

and

$$|s_{12}| = |s_{21}| = \sqrt{1 - |s_{11}|^2}$$

if we assume that $Z_{c1} = Z_{c2}$. Therefore, for our connector pair,

$$|s_{11}| = |s_{22}| = \frac{1.1 - 1}{1.1 + 1} = 0.0476$$

$$|s_{12}|^2 = |s_{21}|^2 = 1 - (0.0476)^2 = 1 - 0.00226$$

and

$$\text{attenuation (dB)} = 10 \log_{10} \frac{1}{|s_{21}|^2} = 10 \log_{10} \frac{1}{1 - 0.00226}$$

$$= 10 \log_{10} (1 + 0.00226) = 0.0098 \text{ dB}$$

In the general case in which port 2 is not terminated in a reflectionless load but in a reflection coefficient $\Gamma_l \neq 0$, the reflection coefficient that one sees when one looks into port 1 is given by the very important formula

$$\Gamma_1 = s_{11} + \frac{s_{12} s_{21} \Gamma_l}{1 - s_{22} \Gamma_l} \tag{3.5-3}$$

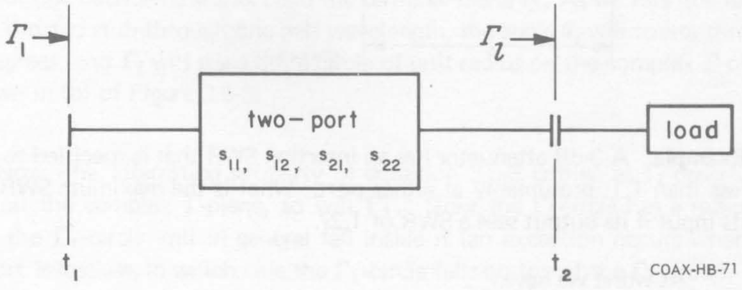


Figure 3.5-1. The reflection coefficient Γ_l of the load is transformed by the two-port into a reflection coefficient Γ_1 that is presented by port 1.

Most of the two-ports one is likely to have dealings with are reciprocal, and usually their ports have the same characteristic impedance. When these conditions are met, $s_{12} = s_{21}$ and (3.5-3) becomes

$$\Gamma_1 = s_{11} + \frac{s_{21}^2 \Gamma_l}{1 - s_{22} \Gamma_l} \quad \text{(reciprocal and } Z_{c1} = Z_{c2}\text{)} \qquad (3.5\text{-}4)$$

Equations 3.5-3 and 3.5-4 show that Γ_1 consists of an intrinsic part s_{11} and a part that is a transformation of Γ_l. When $\Gamma_l = 0$, $\Gamma_1 = s_{11}$.

Example: We saw in Section 3.2 that a transmission line segment of length l can be regarded as a two-port whose s_{11} and s_{22} are zero and whose $s_{21} = s_{12} = 10^{-\frac{1}{20}\alpha(\text{dB/m})l} \measuredangle -\beta l$. Thus, according to (3.5-4), the reflection coefficient presented by the input end of the segment is

$$\Gamma_1 = (10^{-\frac{1}{20}\alpha\,(\text{dB/m})\,l} \measuredangle -\beta l)^2 \Gamma_l = (10^{-\frac{1}{10}\alpha\,(\text{dB/m})\,l} \measuredangle -2\beta l)\,\Gamma_l$$

where Γ_l is the reflection coefficient that terminates the segment at its load end (Figure 3.5-2). The reader should compare this relation with equation 1.8-10, Chapter 1.

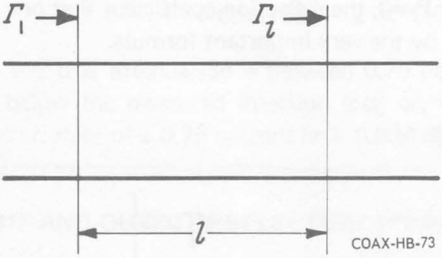


Figure 3.5-2.

Example: A 3-dB attenuator has an insertion SWR that is specified to be less than 1.1, presumably at either port. What is the maximum SWR at its input if its output sees a SWR of 1.2?.

At worst we have

$$|s_{11}| = |s_{22}| = \frac{1.1 - 1}{1.1 + 1} = \frac{0.1}{2.1} = 0.0476$$

Also

$$|\Gamma_l| = \frac{1.2 - 1}{1.2 + 1} = \frac{0.2}{2.2} = 0.091$$

and

$$|s_{21}|^2 = 10^{-\frac{1}{10} \times 3(\text{dB})} = 0.50$$

If the angles of s_{11}, s_{22}, s_{21} and Γ_l are chosen so as to maximize $|\Gamma_1|$, equation 3.5-4 becomes

$$|\Gamma_1|_{\text{max}} = |s_{11}| + \frac{|s_{21}|^2\ |\Gamma_l|}{1 - |s_{22}|\ |\Gamma_l|} = 0.048 + \frac{0.50 \times 0.091}{1 - 0.048 \times 0.091}$$

$$= 0.048 + \frac{0.50 \times 0.091}{1 - 0.0044} = 0.048 + 0.046\ (1 + 0.0044) = 0.094$$

and the maximum input SWR is therefore $(1 + 0.094)/(1 - 0.094) \doteq 1.2$. Notice that the term $|s_{22}||\Gamma_l|$ was too small to influence our result appreciably.

Equation 3.5-3 has a very interesting property, one that is put to good use in many of the procedures for measuring the parameters of two-ports. Suppose we attach to port 2 a variable-length shorted stub, as shown in (a) of Figure 3.5-3. Since the short is totally reflecting, the magnitude $|\Gamma_l|$ of the load reflection coefficient is unity. But the angle θ_l of Γ_l depends upon the adjustable length of line between the short and the terminal plane t_2. As we vary the length of the shorted stub through one half wavelength, the angle θ_l will rotate through 360 degrees, and Γ_l will trace out a circle of unit radius on the complex Γ-plane, as shown in (b) of Figure 3.5-3.

Now, the interesting property of equation 3.5-3 is this: as Γ_l describes a circle on the complex Γ-plane, so will Γ_1. Since the Γ_l-circle has a radius of unity, the Γ_1-circle will in general fall inside it (an exception occurs when the two-port is lossless, in which case the Γ_1-circle falls on top of the Γ_l-circle). The Γ_1-circle will not in general be centered about the origin of the Γ-plane, and points that are uniformly disposed about the Γ_l-circle will not in general correspond to uniformly disposed points on the Γ_1-circle.

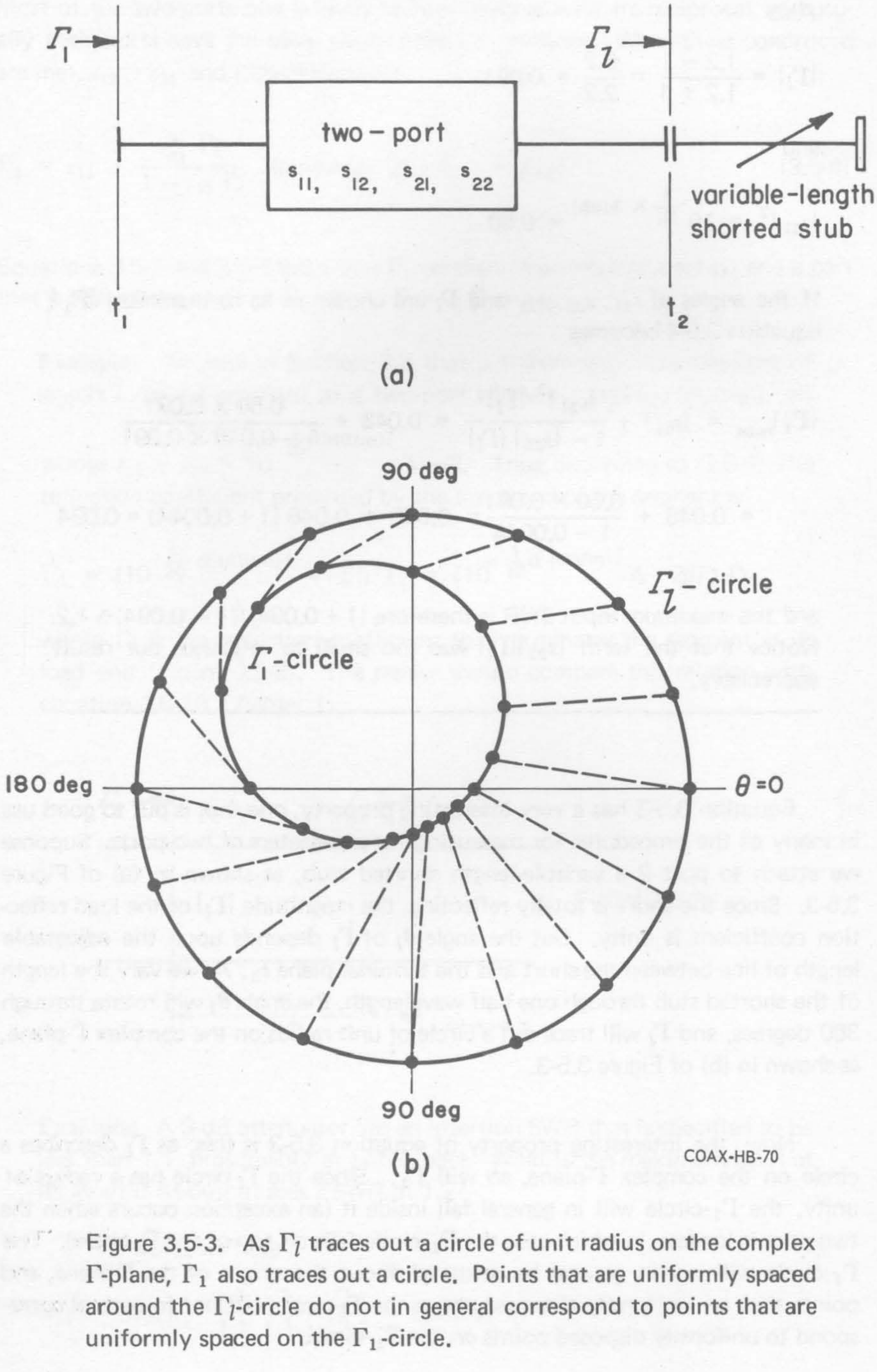


Figure 3.5-3. As Γ_l traces out a circle of unit radius on the complex Γ-plane, Γ_1 also traces out a circle. Points that are uniformly spaced around the Γ_l-circle do not in general correspond to points that are uniformly spaced on the Γ_1-circle.

The radius of the Γ_1-circle and the location of its center are determined by the two-port. The relations between the Γ_1-circle and the s-parameters of the two-port are shown in Figure 3.5-4. The radius of the Γ_1-circle is given by

$$R = \frac{|s_{12}|\,|s_{21}|}{1 - |s_{22}|^2} \tag{3.5-5}$$

or, if reciprocity applies,

$$R = \frac{Z_{c2}}{Z_{c1}} \cdot \frac{|s_{21}|^2}{1 - |s_{22}|^2} \quad \text{(reciprocity)} \tag{3.5-6}$$

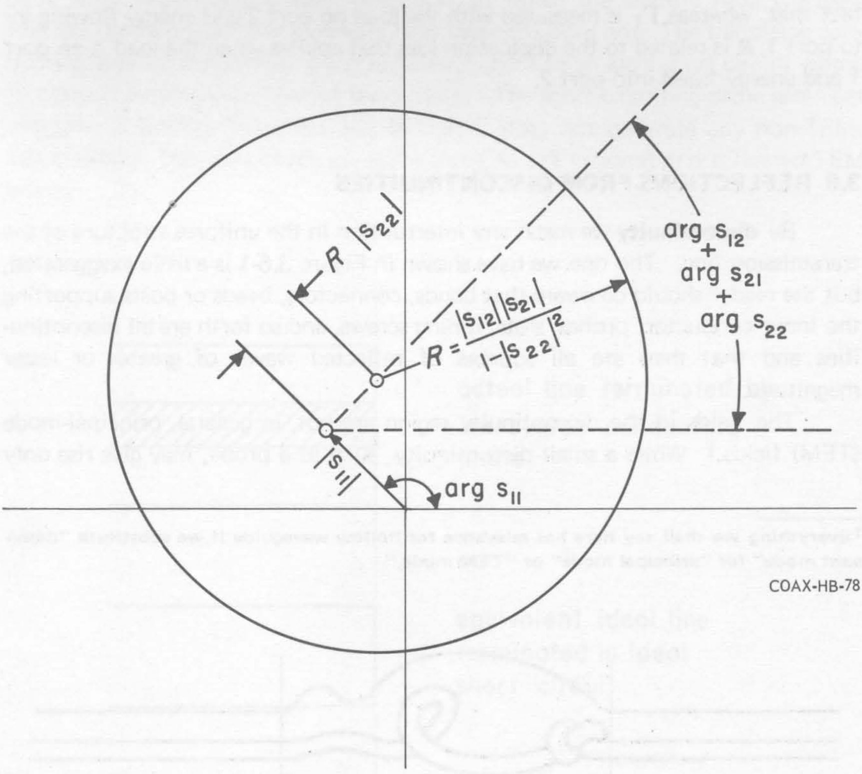


Figure 3.5-4. The Γ_1-circle is a function of the parameters of the two-port. The radius R is the reciprocal of the dissipative component of the backward attenuation ratio, i.e., the attenuation ratio that applies when energy flows into port 2 and out of port 1.

This expression looks like the reciprocal of the dissipative factor in the formula (3.4-12) for the characteristic-insertion-loss (attenuation) ratio, *except that the subscripts 1 and 2 are turned around.* Therefore

$$\frac{1}{R} = \begin{array}{l}\text{dissipation-loss ratio}\\ \text{when } \Gamma_l = \Gamma_s = 0\end{array} = \begin{array}{l}\text{dissipative factor in}\\ \text{attenuation ratio when}\\ \text{load is connected to}\\ \text{port 1 and source is}\\ \text{connected to port 2}\end{array} \tag{3.5-7}$$

When the two-port is not electrically symmetric, so that it matters which port we call #1 and which #2, it is important to keep in mind the seemingly paradoxical fact that, whereas Γ_1 is measured with the load on port 2 and energy flowing into port 1, R is related to the dissipation loss that applies when the load is on port 1 and energy flows into port 2.

3.6 REFLECTIONS FROM DISCONTINUITIES

By **discontinuity** we mean any interruption in the uniform structure of the transmission line. The one we have shown in Figure 3.6-1 is a little exaggerated, but the reader should be aware that bends, connectors, beads or posts supporting the inner conductor, probes, slots, tuning screws, and so forth are all discontinuities and that they are all sources of reflected waves of greater or lesser magnitude.

The fields in the discontinuity region are not, in general, principal-mode (TEM) fields.[†] While a small discontinuity, such as a probe, may give rise only

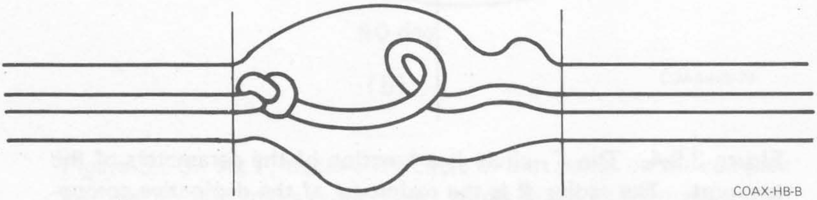


Figure 3.6-1. A rather extreme example of what we call a discontinuity.

[†] Everything we shall say here has relevance for hollow waveguide if we substitute "dominant mode" for "principal mode" or "TEM mode."

to a slight perturbation of the TEM-mode fields, a structure, such as a tee, that totally destroys the uniformity and symmetry of the line can be expected to distort the lines of force to such an extent that they bear little resemblance to the TEM-mode fields in the uniform line. These field perturbations are not necessarily confined to the limits of the discontinuity region. Although non-TEM modes are nonpropagating—we assume that the frequency is below the lowest higher-mode cutoff—the non-TEM-mode fields may penetrate some distance into the uniform line before they are effectively attenuated.

Let us look at a few practical coaxial-line discontinuity structures. The simplest discontinuity one can talk about is a transverse conducting plane that terminates the line, shown in Figure 3.6-2. To the extent that the plane is perfectly conducting, it is an electrical mirror. At the conducting surface the incident wave experiences a total reflection, in which the electric field (voltage) reverses direction but the magnetic field (current) does not. The transverse conducting surface is therefore electrically equivalent to an ideal short circuit whose location coincides with that of the surface. The short-circuiting plane is almost exceptional among discontinuities in that it does not generate any non-TEM-mode fields. The only fields present belong to the incident and reflected TEM waves.

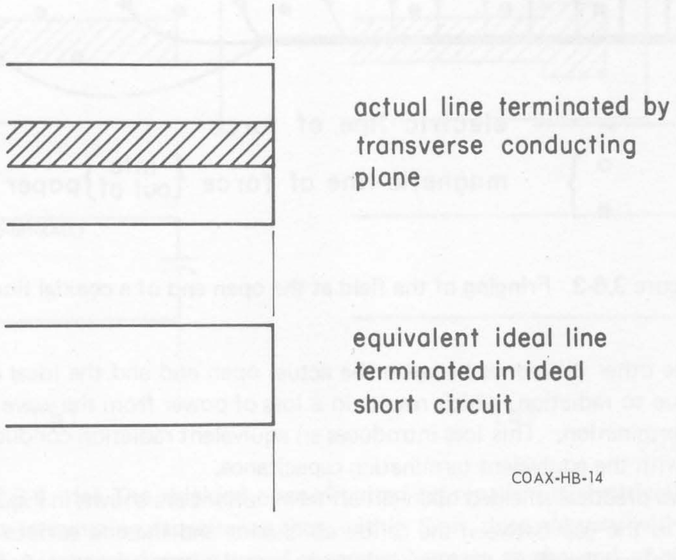


Figure 3.6-2. A perfectly conducting transverse surface that terminates the line looks electrically like an ideal short circuit.

Next let us consider a line whose end is simply left open. We might naively expect by analogy with the previous example that such a termination would look like an ideal open circuit, but it does not. There are two reasons for this. The first is the fringing of the fields shown in Figure 3.6-3, because of which the shunt capacitance c per unit length of line increases toward the end while the series inductance l per unit length decreases (at the very end c is twice and l one half the respective values given in Section 1.5 for an unterminated line). Since the impedance ($Z = V/I$, not Z_c) near an open end is high, it is primarily the increase in capacitance rather than the decrease in inductance that governs the electrical behavior of the open end, and the excess shunt capacitance in the terminal region can be treated as though it were a lumped capacitor connected to the end of the line.

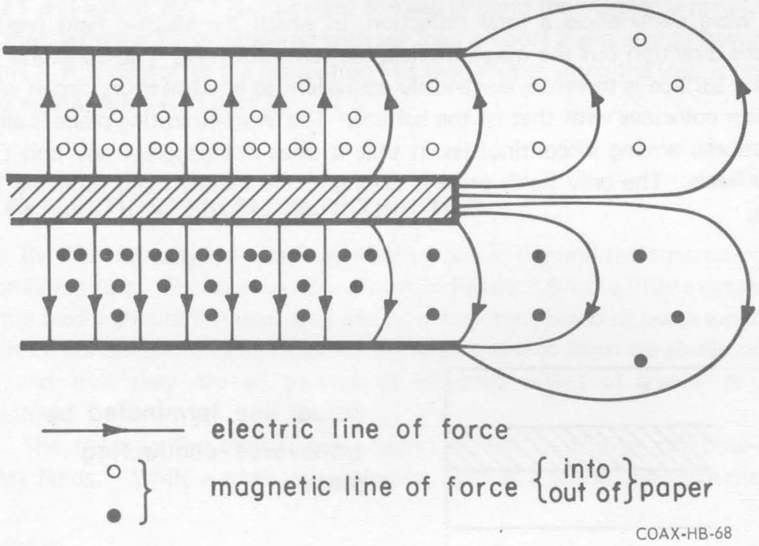


Figure 3.6-3. Fringing of the field at the open end of a coaxial line.

The other difference between the actual open end and the ideal open circuit is due to radiation, which results in a loss of power from the wave incident on the termination. This loss introduces an equivalent radiation conductance in parallel with the equivalent terminating capacitance.

Two practical shielded open-circuit terminations are shown in Figure 3.6-5. As long as the gap between the center conductor and the end surface in (a) is very much smaller than the wavelength, the equivalent lumped capacitance is constant with frequency. Usually what one would like to have is not a constant capacitance but an equivalent ideal open stub of constant electrical length. By

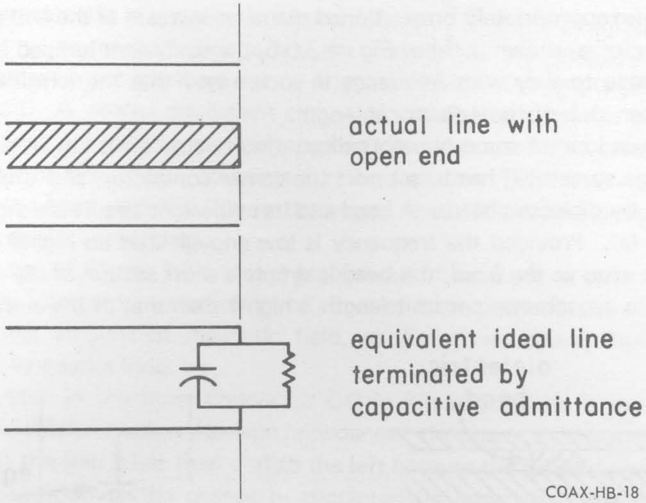


Figure 3.6-4. The effect of fringing fields and radiation at the open end can be accounted for by assuming that the line is ideal but that it is terminated in a capacitive admittance.

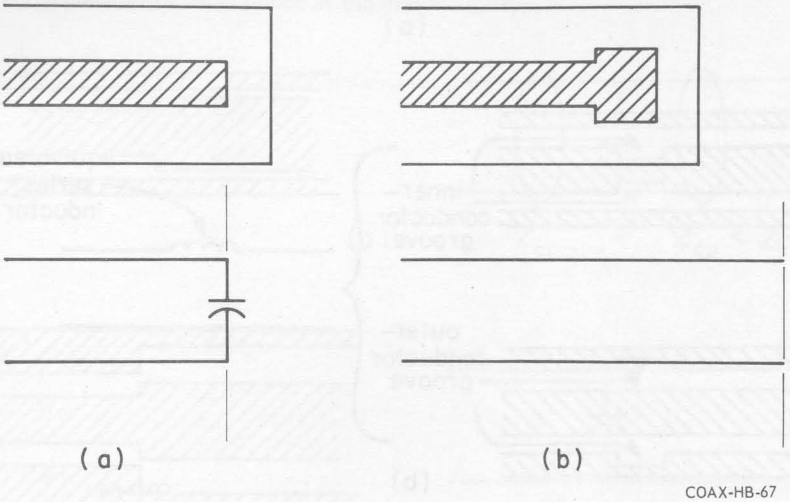


Figure 3.6-5. (a) The shielded open-circuited termination is electrically equivalent to a terminating capacitance that, within limit, does not vary with frequency. (b) A properly proportioned diameter increase at the end of the center conductor makes the termination closely equivalent to an ideal open stub of constant electrical length.

means of an appropriately proportioned diameter increase at the end of the center conductor, as shown in (b) of Figure 3.6-5, the equivalent lumped capacitance can be made to vary with frequency in such a way that the termination looks like an open stub of fixed electrical length.

Let us look at some two-port discontinuity structures. In an air-dielectric coaxial line something has to support the center conductor, and this job is usually done by dielectric beads. A bead and its equivalent circuit are shown in Figure 3.6-6 (a). Provided the frequency is low enough that no higher modes can propagate *even in the bead*, the bead is simply a short section of dielectric-filled line, whose capacitance per unit length is higher than that of the empty line and

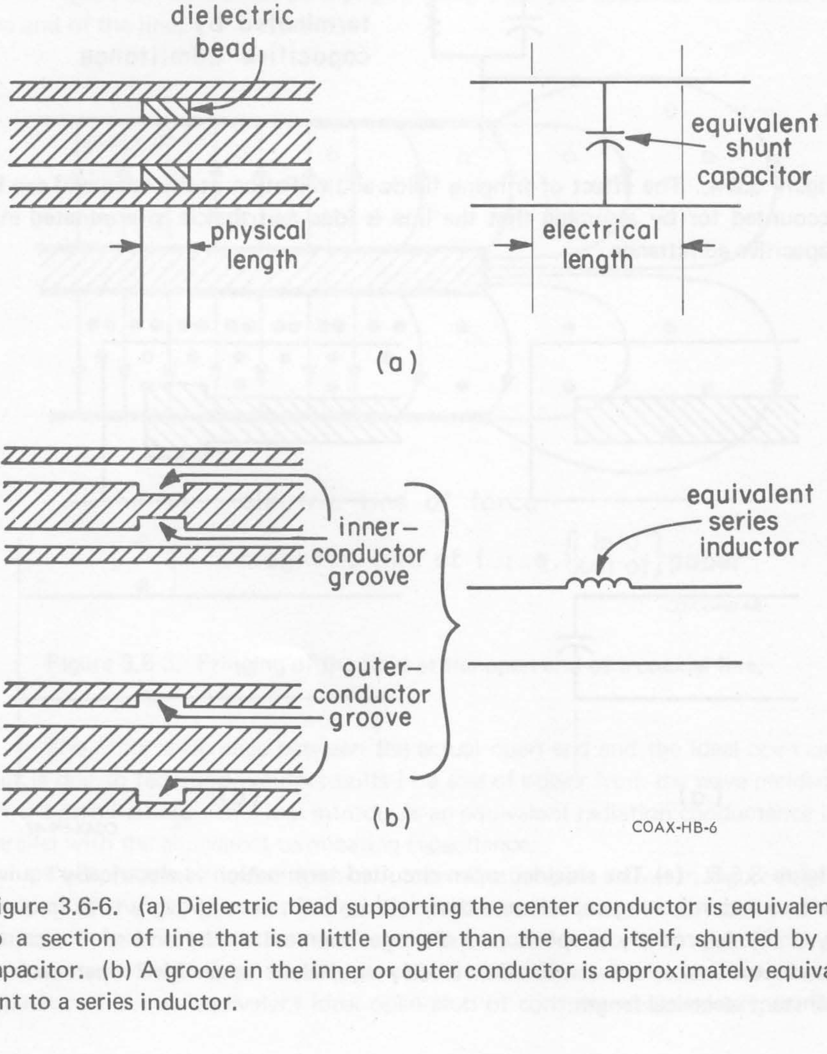


Figure 3.6-6. (a) Dielectric bead supporting the center conductor is equivalent to a section of line that is a little longer than the bead itself, shunted by a capacitor. (b) A groove in the inner or outer conductor is approximately equivalent to a series inductor.

whose characteristic impedance is therefore lower. If the faces of the bead are planes normal to the line's axis, and if the diameters of the inner and outer conductors remain constant, there are no fringing fields in the empty line adjacent to the bead. A precise equivalent circuit of the dielectric bead is a section of ideal line that is a little longer than the bead itself, shunted by a capacitor. The additional electrical length accounts for the fact that TEM waves travel more slowly in the bead than they do in the empty line. The capacitor accounts for the bead's excess capacitance.

Part (b) of Figure 3.6-6 shows grooves cut in the inner or outer conductor of the line. The predominant effect of either of these discontinuities is to create an additional amount of magnetic field, so that either one is approximately equivalent to a series inductor.

The step in the inner conductor OD in Figure 3.6-7 (a) is the boundary between two different characteristic impedances; the characteristic impedance to the right of the step is less than that to the left because the ratio b/a is less on the right. In addition to the change in characteristic impedance, the step causes fringing of the fields, predominantly the electric field, the effect of which is approximated by an equivalent shunt capacitor.

In Figure 3.6-7 (b) the inner conductor OD and the outer conductor ID both jump simultaneously. If the ratio b/a is the same on either side of the step, the characteristic impedance will not change, but the fringing will still introduce an equivalent shunt capacitance at the discontinuity.

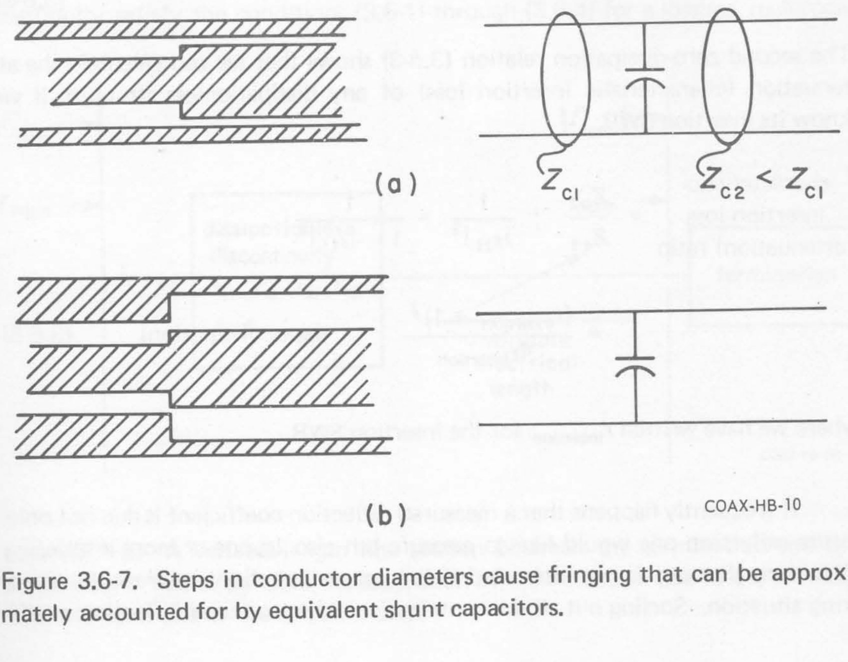


Figure 3.6-7. Steps in conductor diameters cause fringing that can be approximately accounted for by equivalent shunt capacitors.

A discontinuity is a two-port (or a one-port), and everything we have said in the preceding sections of this chapter about two-ports and their s-parameters and about losses applies to discontinuities as to other two-ports. In particular, since all the structures we think of as discontinuities are reciprocal, and virtually all of them may for most purposes be regarded as lossless (dissipationless), their s-parameters practically always satisfy equations 3.2-5 through 3.2-8, which we recapitulate here:

$$Z_{c2}s_{12} = Z_{c1}s_{21} \qquad \text{(reciprocity)} \qquad (3.6\text{-}1)$$

$$\left.\begin{array}{l} |s_{11}| = |s_{22}| \\[1ex] Z_{c2}\ |s_{12}| = Z_{c1}\ |s_{21}| = \sqrt{Z_{c1}Z_{c2}\ (1 - |s_{11}|^2)} \\[1ex] \arg s_{12} + \arg s_{21} = \arg s_{11} + \arg s_{22} \pm\ 1, 3, 5, \ldots \times 180\ \text{deg} \end{array}\right\}\text{(zero-dissipation)} \quad \begin{array}{l}(3.6\text{-}2)\\[1ex](3.6\text{-}3)\\[1ex](3.6\text{-}4)\end{array}$$

The first of the zero-dissipation relations (3.6-2) shows that a dissipationless discontinuity has the same insertion SWR in either direction whether it is symmetric or not.

$$\begin{array}{l}\text{insertion SWR}\\ \text{at port 1}\end{array} = \frac{1 + |s_{11}|}{1 - |s_{11}|} = \frac{1 + |s_{22}|}{1 - |s_{22}|} = \begin{array}{l}\text{insertion SWR}\\ \text{at port 2}\end{array} \quad \text{(zero dissipation)} \qquad (3.6\text{-}5)$$

The second zero-dissipation relation (3.6-3) shows that we can calculate the attenuation (characteristic insertion loss) of any dissipationless structure if we know its insertion SWR.

$$\begin{array}{c}\text{characteristic-}\\ \text{insertion-loss}\\ \text{(attenuation) ratio}\end{array} = \frac{Z_{c2}}{Z_{c1}} \cdot \frac{1}{|s_{21}|^2} = \frac{1}{1 - |s_{11}|^2}$$

$$= \frac{(r_{\text{insertion}} + 1)^2}{4r_{\text{insertion}}} \qquad \text{(zero dissipation)} \qquad (3.6\text{-}6)$$

where we have written $r_{\text{insertion}}$ for the insertion SWR.

It frequently happens that a measured reflection coefficient is due not only to the reflection one would like to measure but also to one or more interfering reflections that one is not interested in but yet cannot eliminate from the measuring situation. Sorting out all the contributions to a measured reflection coefficient would at the very least confront us with a complicated calculation, and, more often than not, such a calculation would not be possible at all because we would not have enough information about the phases of the component reflections.

Figure 3.6-8 shows schematically a measuring arrangement that can sometimes be used to separate the reflection due to a termination from that due to an intervening discontinuity. The reflection coefficient that we see when we look through the discontinuity toward the load is Γ_{input} and the corresponding standing-wave ratio is r_{input}. The reflection coefficient and SWR of the termination are Γ_{term} and r_{term}, and the insertion SWR of the discontinuity is

$$r_{\text{insertion}} = \frac{1 + s_{11}}{1 - s_{11}} \tag{3.6-7}$$

As the length of the variable-length line section changes, Γ_{input} moves in a circular locus on the reflection-coefficient plane, making one complete circle for a half-wavelength change in line length. The circle will not in general be centered at the origin of the Γ-plane. Whether or not the circle encloses the origin depends on the relative size of $|s_{11}|$ and $|\Gamma_{\text{term}}|$. When $|\Gamma_{\text{term}}| > |s_{11}|$, the Γ_{term}-circle encloses the origin of the Γ-plane; when $|\Gamma_{\text{term}}| < |s_{11}|$, it does not.

At some point on the Γ_{input}-circle, $|\Gamma_{\text{input}}|$ is maximum; and at the diametrically opposite point, $|\Gamma_{\text{input}}|$ is minimum. If the s-parameters of the discontinuity satisfy the conditions (3.6-1) through (3.6-4) for a lossless, reciprocal

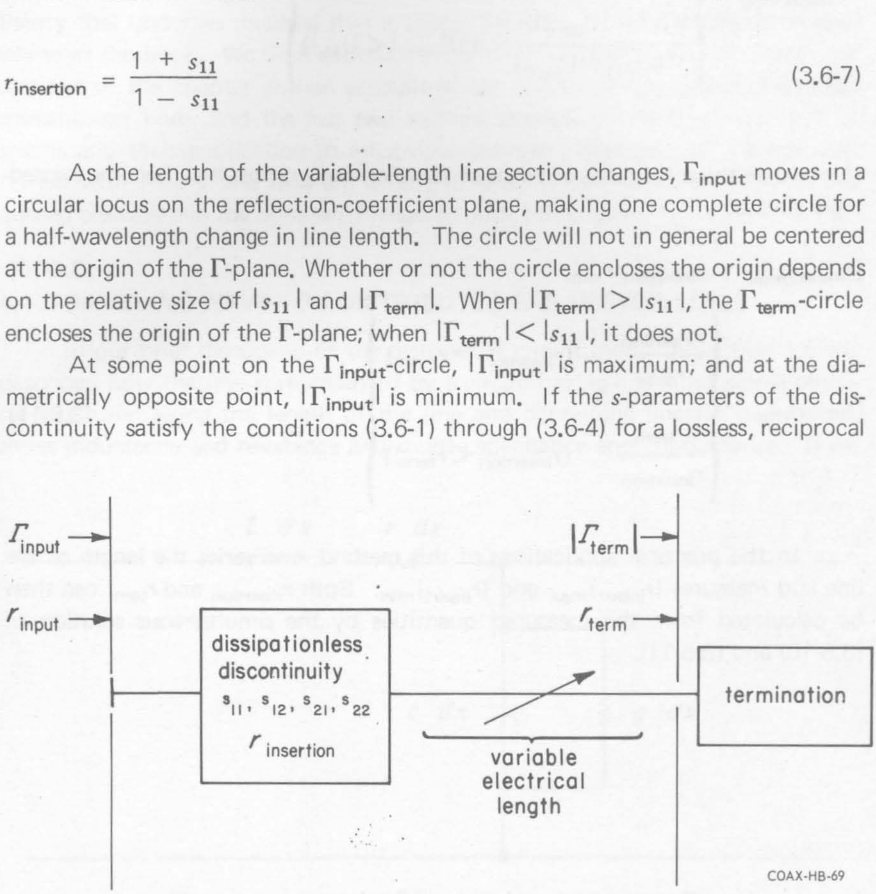


Figure 3.6-8. A variable-length line section between the termination and the discontinuity makes possible the separation of the reflection due to the termination from that due to the discontinuity.

two-port, it can be shown that equation 3.5-3 gives the following formulas for the maximum and minimum values of the input reflection coefficient.

$$|\Gamma_{\text{input}}|_{\max} = \frac{|s_{11}| + |\Gamma_{\text{term}}|}{1 + |s_{11}|\,|\Gamma_{\text{term}}|} \tag{3.6-8}$$

$$|\Gamma_{\text{input}}|_{\min} = \left\{ \begin{array}{ll} \dfrac{|\Gamma_{\text{term}}| - |s_{11}|}{1 - |s_{11}||\Gamma_{\text{term}}|} & (|\Gamma_{\text{term}}| > |s_{11}|) \\ \dfrac{|s_{11}| - |\Gamma_{\text{term}}|}{1 - |s_{11}|\,|\Gamma_{\text{term}}|} & (|\Gamma_{\text{term}}| < |s_{11}|) \end{array} \right\} \tag{3.6-9}$$

If we write these formulas in terms of standing-wave ratios, they become exceedingly simple:

$$(r_{\text{input}})_{\max} = r_{\text{insertion}}\, r_{\text{term}} \tag{3.6-10}$$

$$(r_{\text{input}})_{\min} = \left\{ \begin{array}{ll} \dfrac{r_{\text{insertion}}}{r_{\text{term}}} & (r_{\text{insertion}} > r_{\text{term}}) \\ \dfrac{r_{\text{term}}}{r_{\text{insertion}}} & (r_{\text{insertion}} < r_{\text{term}}) \end{array} \right\} \tag{3.6-11}$$

In the practical applications of this method, one varies the length of the line and measures $(r_{\text{input}})_{\max}$ and $(r_{\text{input}})_{\min}$. Both $r_{\text{insertion}}$ and r_{term} can then be calculated from the measured quantities by the simultaneous solution of (3.6-10) and (3.6-11).

CHAPTER 4

Some Theoretical Background

In this chapter we provide, for those readers who wish it, a little of the theory that underlies material that is presented without much justification elsewhere in the book. We shall confine ourselves to just two topics. The first four sections of the chapter give an account of the theory of distributed parameter transmission lines, and the last two sections provide an introduction to flow graphs and their application to microwave systems. Readers who are not concerned with theory and who are willing to take on faith some of the formulas quoted elsewhere in the book are invited to omit this chapter.

4.1 TRAVELING WAVES ON DISTRIBUTED PARAMETER LINES

In our brief discussion of the distributed circuit model in Section 1.5 we described how the line is represented by a circuit having one-dimensional physical extension along the length of the line and containing linearly distributed series inductance and resistance and shunt capacitance and conductance. If we

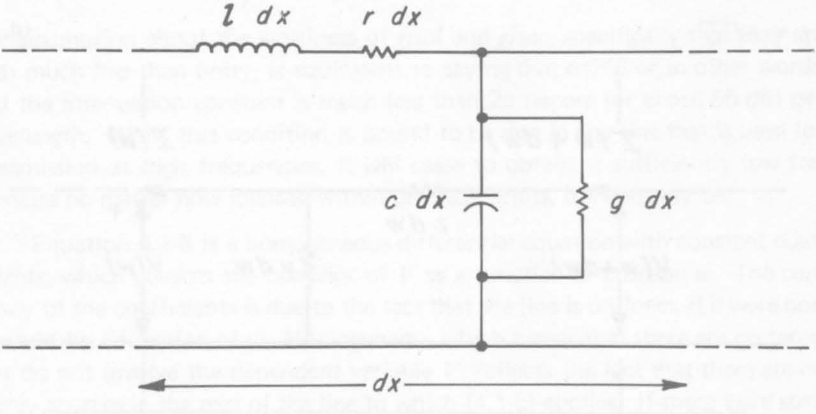


Figure 4.1-1. An infinitessimal length dx of line in the distributed parameter model.

write z for the series impedance per unit length and y for the shunt admittance per unit length, then

$$z = r + j\omega l \tag{4.1-1}$$

and

$$y = g + j\omega c \tag{4.1-2}$$

where r, l, g, and c are respectively the series resistance, series inductance, shunt conductance, and shunt capacitance, all per unit length of line.

Let w be the variable of position along the line's axis, increasing from right to left as shown in Figure 4.1-2. If $V(w)$ and $I(w)$ are the voltage and current in the line at w, defined as in Figure 4.1-2, then their rates of change with position will be given by

$$\frac{dV(w)}{dw} = zI(w) \tag{4.1-3}$$

and

$$\frac{dI(w)}{dw} = yV(w) \tag{4.1-4}$$

Elimination of $I(w)$ between (4.1-3) and (4.1-4) leads to a second-order differential equation in $V(w)$,

$$\frac{d^2 V(w)}{dw^2} - \gamma^2 V(w) = 0 \tag{4.1-5}$$

where

$$\gamma = +\sqrt{zy} \tag{4.1-6}$$

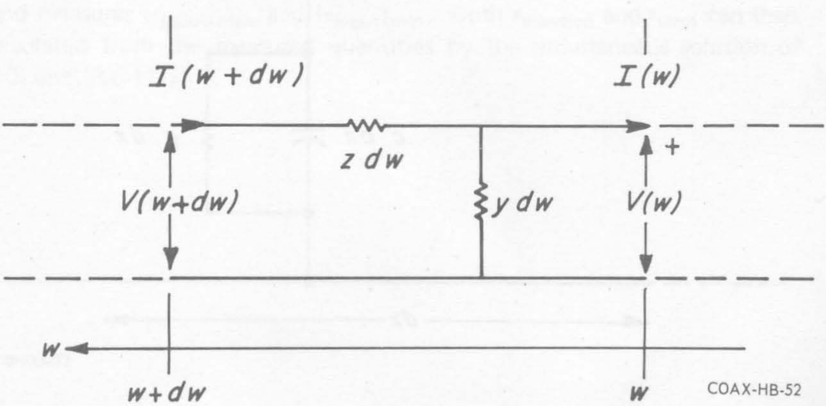


Figure 4.1-2.

is called the **propagation constant.** The real part of γ is the **attenuation constant** α and the imaginary part is the **phase constant** β.

$$\gamma = \alpha + j\beta \tag{4.1-7}$$

When we substitute (4.1-1) for z and (4.1-2) for y in (4.1-6) we get

$$\gamma = \sqrt{(r + j\omega l)\,(g + j\omega c)} = \omega\sqrt{lc} \cdot \sqrt{(j + \frac{r}{\omega l})\,(j + \frac{g}{\omega c})} \tag{4.1-8}$$

The lines we are interested in are not too lossy, so we shall regard $r/\omega l$ and $g/\omega c$ as small quantities. Let us expand (4.1-8) in powers of $r/\omega l$ and $g/\omega c$:

$$\gamma = \omega\sqrt{lc}\left[j + \frac{1}{2}\left(\frac{r}{\omega l} + \frac{g}{\omega c}\right) + \ldots\right] \tag{4.1-9}$$

The largest imaginary terms that are neglected in (4.1-9) are of the second order in small quantities and the largest real terms are of the third order.

The imaginary part of (4.1-9) is the phase constant and the real part is the attenuation constant. Retaining only the leading terms, we have

$$\beta \doteq \omega\sqrt{lc} \tag{4.1-10}$$

and

$$\alpha \doteq \frac{1}{2}\left(r\sqrt{\frac{c}{l}} + g\sqrt{\frac{l}{c}}\right) \tag{4.1-11}$$

Our assumption about the smallness of $r/\omega l$ and $g/\omega c$, specifically that they are both much less than unity, is equivalent to saying that $\alpha \ll \beta$ or, in other words, that the attenuation constant is much less than 2π nepers (or about 55 dB) per wavelength. While this condition is bound to be met in any line that is used for transmission at high frequencies, it will cease to obtain at sufficiently low frequencies no matter how lossless, within practical limits, the line may be.

Equation 4.1-5 is a homogeneous differential equation with constant coefficients, which governs the behavior of V as a function of position w. The constancy of the coefficients is due to the fact that the line is uniform; if it were not, γ would be a function of w. Homogeneity, which means that there are no terms that do not involve the dependent variable V, reflects the fact that there are no energy sources in the part of the line to which (4.1-5) applies. If there were such sources, for example if we were talking about a traveling-wave device with coupling between an electron beam and the transmission line, there would be a driving function on the right-hand side of (4.1-5).

The two independent solutions that we know (4.1-5) must have are $e^{+\gamma w}$ and $e^{-\gamma w}$. The complete solution is a linear combination of these functions, and accordingly we write

$$V(w) = V^+(0)\ e^{+\gamma w} + V^-(0)\ e^{-\gamma w} \tag{4.1-12}$$

where $V^+(0)$ and $V^-(0)$ are arbitrary constants to be determined from the boundary conditions.

The two parts of V, $V^+(0)\ e^{+\gamma w}$ and $V^-(0)\ e^{-\gamma w}$, represent respectively a rightgoing (toward $-w$) and a leftgoing (toward $+w$) wave. We can understand why this is so as follows. Our V's and I's are phasors—time-independent complex quantities that represent sinusoidally time-varying voltages and currents. It is conventional with electrical engineers, although not with physicists, to write the time dependence of sinusoidal functions with a positive imaginary exponent thus:

$$v(t) = \text{real part } (Ve^{j\omega t}) = |V| \cos(\omega t + \arg V) \tag{4.1-13}$$

The angle of the complex number V is the phase of the sinusoidal voltage $v(t)$. Thus the distribution of instantaneous voltage $v(w,t)$ on the line corresponding to the phasor function $V^+(0)\ e^{+\gamma w}$ is

$$v(w,t) = |V^+(0)|e^{\alpha w} \cos(\beta w + \omega t + \arg V^+(0)) \tag{4.1-14}$$

At any instant of time this is a sinusoidal function of w that is attenuated exponentially toward the right (toward $-w$). As time progresses, the exponential envelope remains fixed while the oscillating cosine moves to the right inside it with a **phase velocity** given by

$$\text{phase velocity} = \left(\frac{d(-w)}{dt}\right)_{\substack{\textbf{argument of}\\ \textbf{cosine remains}\\ \textbf{constant}}} = \frac{\omega}{\beta} \tag{4.1-15}$$

Let us write $V^+(w)$ and $V^-(w)$ for the rightgoing and leftgoing parts of the expression (4.1-12) for $V(w)$. Thus

$$\left.\begin{aligned} &V(w) = V^+(w) + V^-(w) \\ &\text{where} \\ &V^+(w) = V^+(0)e^{\gamma w} \quad \text{and} \quad V^-(w) = V^-(0)e^{-\gamma w} \end{aligned}\right\} \tag{4.1-16}$$

We can calculate the current in the line with the help of (4.1-3) and (4.1-6).

$$I(w) = \frac{1}{z}\frac{dV(w)}{dw} = \sqrt{\frac{y}{z}}\;\left(V^+(0)e^{\gamma w} - V^-(0)e^{-\gamma w}\right) \tag{4.1-17}$$

The factor $\sqrt{y/z}$ is the **characteristic admittance** Y_c of the line and its reciprocal is the **characteristic impedance.**

$$Y_c = \frac{1}{Z_c} = \sqrt{\frac{y}{z}} \tag{4.1-18}$$

We recognize the two parts of the right-hand side of (4.1-17) as rightgoing and leftgoing waves, and hence we write, in conformity with the notation we have already adopted for the two parts of $V(w)$,

$$\left.\begin{aligned} &I(w) = I^+(w) + I^-(w) \\ &\text{where} \\ &I^+(w) = Y_c\,V^+(w) \quad \text{and} \quad I^-(w) = -Y_c\,V^-(w) \end{aligned}\right\} \tag{4.1-19}$$

4.2 LOSSLESS COAXIAL LINE

If the radii of the inner and outer conducting surfaces of the coaxial line are a and b respectively, and if there is an amount of charge χ per unit length of line uniformly distributed over the inner conductor, we can write down from Gauss' law that the magnitude $E(r)$ of the electric field at any radius r between the conductors is

$$E(r) = \frac{\chi}{2\pi\epsilon} \cdot \frac{1}{r} \tag{4.2-1}$$

The energy W_E per unit length of line associated with the electric field is

$$W_E = \frac{\epsilon}{2}\int_0^{2\pi}\int_a^b E^2\,(r)\;r\,dr\,d\theta = \frac{\chi^2}{4\pi\epsilon}\;\log_e\frac{b}{a} \tag{4.2-2}$$

and the **capacitance per unit length** is

$$c = \frac{\chi^2}{2W_E} = \frac{2\pi\epsilon}{\log_e \dfrac{b}{a}} \tag{4.2-3}$$

If the current in the inner conductor is i, Ampere's line-integral formula gives for the magnitude $H(r)$ of the magnetic field at any radius r between the conductors

$$H(r) = \frac{i}{2\pi} \cdot \frac{1}{r} \tag{4.2-4}$$

The energy per unit length of line associated with the magnetic field is

$$W_H = \frac{\mu}{2} \int_0^{2\pi} \int_a^b H^2\ (r)\ r\ dr\ d\theta = \frac{\mu i^2}{4\pi} \log_e \frac{b}{a} \tag{4.2-5}$$

and the **external inductance per unit length** l_e is

$$l_e = \frac{2W_H}{i^2} = \frac{\mu}{2\pi} \log_e \frac{b}{a} \tag{4.2-6}$$

The external inductance l_e is that associated with magnetic flux in the dielectric between the conductors as opposed to an internal inductance l_i, which we shall consider in the next section, that arises owing to the flux that penetrates below the surfaces of imperfect conductors. Note that the constant μ in (4.2-6) is the permeability of the dielectric, and hence is equal to μ(vac).

The reader might well ask what c and l_e can possibly have to do with propagation of waves on the line at high frequencies, since they are dc values of capacitance and inductance, and surely the fields at 9 GHz do not look like dc fields. But in fact the fields at 9 GHz do look like dc fields, and this is because we are talking about TEM waves.

It is possible to show in a very elegant manner[†] that, in any axially uniform metallic guiding system, Maxwell's equations separate so as to yield ordinary differential equations in the axial direction that are identical with the distributed-parameter transmission-line equations (4.1-3) and (4.1-4). It is this remarkable fact that justifies the use of the distributed parameter model to represent wave propagation in *any transmission line or waveguide.* The field-theory solution to the problem also shows that in the case of the TEM mode: 1) the fields always have the same form as the dc fields, 2) the quantities that appear as "voltage" and "current" in the field equations can be identified with the ordinary voltage and current in the line, and 3) the distributed series impedance and shunt admittance that appear in the field equations are just the quantities that we arrive at

---

[†] **The interested reader is referred to N. Marcuvitz (ed), "Waveguide Handbook," Vol. 10, MIT Radiation Laboratory Series, McGraw-Hill Book Co., New York, 1951, p 1 ff.**

by assuming a distributed inductance given by (4.2-6) and capacitance given by (4.2-3). Thus the distributed parameter model of the line, with values of inductance and capacitance that are calculated on the assumption that the fields are static, is an accurate description of TEM wave propagation on any lossless transmission line *at frequencies up to those at which higher modes begin to propagate.*

Having, we hope, assured the reader that our unsophisticated treatment of transmission-line propagation is legitimate, we shall conclude this section by calculating the propagation constant and characteristic impedance of a lossless coaxial line.

If we put $z = j\omega l_e$ and $y = j\omega c$ in (4.1-6) and (4.1-18) we get, respectively,

$$\gamma = j\beta = j\omega\sqrt{l_e c} \tag{4.2-7}$$

and

$$Z_c = \sqrt{\frac{l_e}{c}} \tag{4.2-8}$$

Now making use of the forumlas we have just calculated for c and l_e, (4.2-3) and (4.2-6), we have

$$\beta = \omega\sqrt{\mu\epsilon} \tag{4.2-9}$$

or, in view of (4.1-15),

$$\text{phase velocity} = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu\epsilon}} \tag{4.2-10}$$

and

$$Z_c = \frac{1}{2\pi}\sqrt{\frac{\mu}{\epsilon}}\log_e\frac{b}{a} \tag{4.2-11}$$

Bear in mind that μ and ϵ in these formulas are the permeability and permittivity of the dielectric.

4.3 COAXIAL LINES WITH SMALL LOSSES

We pointed out in Chapter 1 that the fields in a lossy line are not strictly TEM. Nevertheless, if the losses are small we are justified in assuming that the fields are virtually the same as they would be in a lossless line and hence that we may continue to use our distributed parameter model to represent the line.

When the line is lossless, the only contribution to the variation of voltage with distance is the changing magnetic flux in the space between the conductors, and, as we saw in the preceding section, this leads to an inductive component $j\omega l_e$ of z. Conductor loss causes the electric field to have a small tangential component at the conducting surfaces. This tangential field component gives rise to an additional contribution to dV/dw which we account for in the model by adding a small impedance in series with l_e.

We shall assume that it is possible to define a **surface impedance** Z_s as the ratio of the tangential component E_{tan} of the electric field at the surface of the metal to the surface current density K (amperes/meter).

$$Z_s = R_s + jX_s = \frac{E_{\text{tan}}}{K} \tag{4.3-1}$$

The surface impedance of an ideal, plane-conducting surface is given by[†]

$$Z_s = (1 + j)\sqrt{\frac{\omega\mu}{2\sigma}} \tag{4.3-2}$$

where μ is equal to μ(vac) for a nonferromagnetic conductor and σ is the conductivity in ohms^{-1}/meter. An interesting aspect of (4.3-2) is that the resistive and reactive components of Z_s are equal.

The case of a real conductor is complicated by the degree of compactness and the surface finish. The irregularities of a rough or porous surface may well extend to depths on the order of, or even much greater than, the skin depth. Then not only is the surface impedance much greater than the ideal value given by (4.3-2), but it is also no longer accurate to assume that the real and imaginary parts of Z_s are equal. It has nevertheless become customary to talk about surface impedance in terms of "effective conductivities" as though (4.3-2) applied rigorously. Accordingly, let us define $\sigma_{eff,\,R}$ and $\sigma_{eff,\,X}$ so that the real and imaginary parts of Z_s are given by

$$R_s = \sqrt{\frac{\omega\mu}{2\sigma_{eff,\,R}}} \tag{4.3-3}$$

and

$$X_s = \sqrt{\frac{\omega\mu}{2\sigma_{eff,\,X}}} \tag{4.3-4}$$

[†] See for example Ramo and Whinnery, "Fields and Waves in Modern Radio," (second edition), John Wiley and Sons, New York, 1953, p 239.

At frequencies high enough that the skin depth is very much smaller than the radius of the inner conductor, the series resistance per unit length r and the additional series reactance per unit length ωl_i due to the surface impedance of the conductors are given by

$$r = \sqrt{\frac{\omega\mu}{2}} \left(\frac{1}{2\pi a \sqrt{\sigma_{eff,\, R \text{ (inner cond)}}}} + \frac{1}{2\pi b \sqrt{\sigma_{eff,\, R \text{ (outer cond)}}}} \right) \tag{4.3-5}$$

and

$$\omega l_i = \sqrt{\frac{\omega\mu}{2}} \left(\frac{1}{2\pi a \sqrt{\sigma_{eff,\, X \text{ (inner cond)}}}} + \frac{1}{2\pi b \sqrt{\sigma_{eff,\, X \text{ (outer cond)}}}} \right) \tag{4.3-6}$$

The component l_i of the inductance per unit length is called **internal inductance** because it is due to magnetic flux within the interior of the conductors.

The conductor-loss component of the line's attenuation is given by the first term on the right of (4.1-11):

$$\alpha(\text{cond}) = \frac{1}{2} r \sqrt{\frac{c}{l}} \tag{4.3-7}$$

Since α itself is a small quantity we may ignore the contribution of l_i to l and write

$$\alpha(\text{cond}) = \frac{1}{2} r Y_c(\text{lossless}) \tag{4.3-8}$$

The inductance l_i causes a small increase in the phase constant (decrease in the velocity of propagation). If we put $l = l_e + l_i$ into (4.1-10) we have

$$\beta = \omega\sqrt{l_e c} \cdot \sqrt{1 + \frac{l_i}{l_e}} \doteq \omega\sqrt{l_e c}\,(1 + \frac{1}{2}\frac{l_i}{l_e}) \tag{4.3-9}$$

or

$$\frac{\Delta\beta\text{ (due to conductors)}}{\beta\text{ (lossless line)}} = \frac{1}{2}\frac{l_i}{l_e} \tag{4.3-10}$$

We are justified in ignoring the effect of r on β. Referring to equation 4.1-9, we see that the largest imaginary terms involving r are of the second order in small quantities, whereas the correction we have just calculated is of the first order.

Both r and l_i affect the characteristic impedance significantly. If we ignore dielectric loss for the moment we have

$$Z_c = \sqrt{\frac{z}{y}} = \sqrt{\frac{j\omega(l_e + l_i) + r}{j\omega c}} \doteq Z_c(\text{lossless}) \left[1 + \frac{1}{2}\left(\frac{l_i}{l_e} - j\frac{r}{\omega l_e} \right) \right] \tag{4.3-11}$$

or

$$\frac{\Delta Z_c(\text{cond})}{Z_c(\text{lossless})} = \frac{1}{2}\frac{l_i}{l_e} - j\frac{1}{2}\frac{r}{\omega l_e} \tag{4.3-12}$$

The foregoing theory of conductor loss finds practical application in the case of precision air-dielectric lines whose conductors have been fabricated so as to have smooth, compact surfaces. For such conducting surfaces one can assume that $\sigma_{eff,\,R} = \sigma_{eff,\,X}$ and hence that $r = \omega l_i$. In this circumstance

$$\begin{aligned}\frac{\alpha(\text{cond})}{\beta(\text{lossless})} = \frac{\Delta\beta(\text{cond})}{\beta(\text{lossless})} &= \text{real part}\left(\frac{\Delta Z_c(\text{cond})}{Z_c(\text{lossless})}\right)\\ &= -\text{ imaginary part}\left(\frac{\Delta Z_c(\text{cond})}{Z_c(\text{lossless})}\right)\end{aligned} \tag{4.3-13}$$

The shunt admittance per unit length of a lossless line is

$$y(\text{lossless}) = j\omega c = j\omega\,\frac{2\pi\epsilon}{\log_e \dfrac{b}{a}} \tag{4.3-14}$$

The shunt admittance of a line with a lossy dielectric is found by substituting the complex permittivity $\epsilon' - j\epsilon'' = \epsilon(1 - j\tan\delta)$, where δ is the loss angle, in place of ϵ in (4.3-14). Thus

$$y(\text{lossy}) = j\omega\,\frac{2\pi\epsilon\,(1 - j\tan\delta)}{\log_e \dfrac{b}{a}} = j\omega c + \omega c\tan\delta \tag{4.3-15}$$

so that the shunt conductance due to dielectric loss is

$$g = \omega c\tan\delta \tag{4.3-16}$$

The contribution that g makes to the line's attenuation is found by putting (4.3-16) into the second term on the right of (4.1-11):

$$\alpha(\text{diel}) = \frac{1}{2}\,\omega c\tan\delta\cdot\sqrt{\frac{l}{c}} = \frac{1}{2}\,\beta\tan\delta \tag{4.3-17}$$

The influence of dielectric loss on the phase constant can be ignored. Even theoretically, g, like r, causes a change in β that is of the second order in small quantities. But even if g were large, its effect on β would be swamped by the uncertainty in the dielectric constant.

Under some circumstances, though, dielectric loss might have a significant effect on the characteristic impedance, since it gives rise to an imaginary component. Ignoring conductor loss, we have

$$Z_c = \sqrt{\frac{z}{y}} = \sqrt{\frac{j\omega l_e}{j\omega c + g}} \doteq Z_c(\text{lossless}) \cdot \left(1 + \frac{j}{2}\,\frac{g}{\omega c}\right) \qquad (4.3\text{-}18)$$

or

$$\frac{\Delta Z_c(\text{diel})}{Z_c(\text{lossless})} = \frac{j}{2}\tan\delta \qquad (4.3\text{-}19)$$

4.4 THE TERMINATED LINE

Let us return to the general expressions (4.1-16) and (4.1-19) for the voltage and current on the line:

$$\left.\begin{aligned} &V(w) = V^+(w) + V^-(w) \\ &\text{where} \\ &V^+(w) = V^+(0)e^{\gamma w} \quad \text{and} \quad V^-(w) = V^-(0)e^{-\gamma w} \end{aligned}\right\} \qquad (4.4\text{-}1)$$

and

$$\left.\begin{aligned} &I(w) = I^+(w) + I^-(w) \\ &\text{where} \\ &I^+(w) = Y_c V^+(w) \quad \text{and} \quad I^-(w) = -Y_c V^-(w) \end{aligned}\right\} \qquad (4.4\text{-}2)$$

Mathematically, various sets of boundary conditions might be used to determine the two constants $V^+(0)$ and $V^-(0)$ in (4.4-1) and (4.4-2). As a matter of fact, though, we do not wish to make a unique determination of $V^+(0)$ and $V^-(0)$. For the present we are interested in the effect of terminating the line at its right-hand end in a load of known impedance or reflection coefficient, and this condition imposes a constraint on (4.4-1) and (4.4-2) that is sufficient only to fix the ratio of $V^+(0)$ and $V^-(0)$. This partial determination of the solution to the problem will enable us to calculate impedances and reflection coefficients, but not the actual values of voltages and currents. To find these we would have to specify another boundary condition, for example the voltage at the load or at the generator.

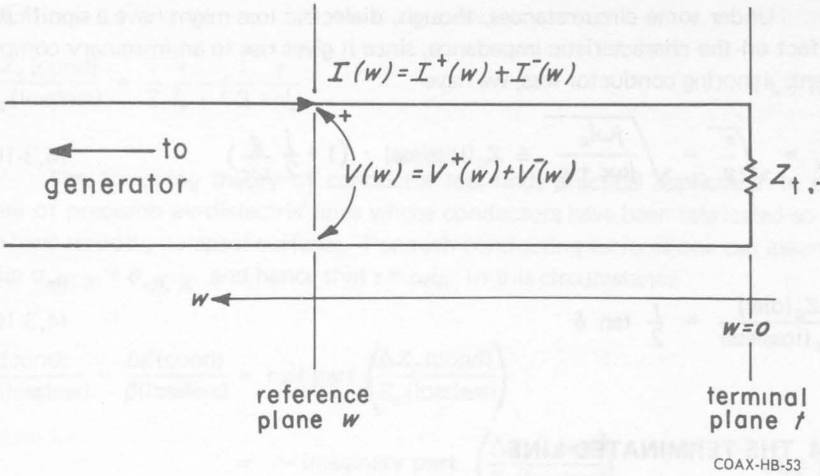


Figure 4.4-1.

Let the load's terminal plane t be at $w = 0$. We can write (4.4-1) and (4.4-2) in the form

$$V(w) = V^+(0)\ [e^{\gamma w} + \Gamma(0)e^{-\gamma w}] \tag{4.4-3}$$

$$I(w) = Y_c V^+(0)\ [e^{\gamma w} - \Gamma(0)e^{-\gamma w}] \tag{4.4-4}$$

where

$$\Gamma(0) = \frac{V^-(0)}{V^+(0)} = \Gamma_t \tag{4.4-5}$$

is the reflection coefficient of the load. Equations (4.4-3) and (4.4-4) emphasize the role of the forward and reflected waves in making up the total voltage and current on the line.

If we solve (4.4-1) and (4.4-2) for $V^+(w)$ and $V^-(w)$ in terms of $V(w)$ and $I(w)$, obtaining

$$V^+(w) = \frac{1}{2}\ [V(w) + Z_c I(w)] \tag{4.4-6}$$

and

$$V^-(w) = \frac{1}{2}\ [V(w) - Z_c I(w)] \tag{4.4-7}$$

and then use these expressions, evaluated at $w = 0$, to eliminate $V^+(0)$ and $V^-(0)$ from (4.4-1) and (4.4-2), we get

$$V(w) = V(0)\cosh\gamma w + Z_c I(0)\sinh\gamma w \tag{4.4-8}$$

$$I(w) = I(0)\cosh\gamma w + Y_c V(0)\sinh\gamma w \tag{4.4-9}$$

Equations (4.4-8) and (4.4-9) put in prominence the 2-port-network aspect of a length of line, since they may be thought of as relations between the voltage and current at an input port and the voltage and current at an output port.

The reflection coefficient $\Gamma(w)$ at the reference plane w is defined by

$$\Gamma(w) = \frac{V^-(w)}{V^+(w)} \tag{4.4-10}$$

and the impedance $Z(w)$ is defined by

$$Z(w) = \frac{V(w)}{I(w)} \tag{4.4-11}$$

Impedance and reflection coefficient are mathematically equivalent. If we combine the definitions (4.4-10) and (4.4-11) with either (4.4-1) and (4.4-2) or (4.4-6) and (4.4-7), we get

$$\Gamma(w) = \frac{\dfrac{Z(w)}{Z_c} - 1}{\dfrac{Z(w)}{Z_c} + 1}, \qquad \frac{Z(w)}{Z_c} = \frac{1 + \Gamma(w)}{1 - \Gamma(w)} \tag{4.4-12}$$

Readers who are acquainted with the theory of conformal mapping will recognize (4.4-12) as a bilinear transformation which maps the right-hand half of the Z-plane into the interior of the unit circle about the origin of the Γ-plane.

From (4.4-1) we have

$$\frac{V^-(w)}{V^+(w)} = \frac{V^-(0)e^{-\gamma w}}{V^+(0)e^{+\gamma w}}$$

so that the relation between the reflection coefficient $\Gamma(w)$ at any plane w and that at the terminal plane $w = 0$ is

$$\Gamma(w) = \Gamma(0)e^{-2\gamma w} \tag{4.4-13}$$

The relation between the impedance $Z(w)$ at w and that at $w = 0$ can be obtained by dividing (4.4-8) by (4.4-9). We get

$$\frac{Z(w)}{Z_c} = \frac{\dfrac{Z(0)}{Z_c} + \tanh \gamma w}{1 + \dfrac{Z(0)}{Z_c} \tanh \gamma w} \qquad (4.4\text{-}14)$$

The voltage and current relations (4.4-8) and (4.4-9) and the impedance relation (4.4-14) are really quite complicated because the arguments of the hyperbolic functions are complex. We can obtain simpler expressions that are valid when the attenuation is small. Making use of well known identities, we have

$$\cosh \gamma w = \cosh (\alpha + j\beta) w = \cosh \alpha w \cos \beta w + j \sinh \alpha w \sin \qquad (4.4\text{-}15)$$

and

$$\sinh \gamma w = \sinh (\alpha + j\beta) w = \sinh \alpha w \cos \beta w + j \cosh \alpha w \sin \beta w \qquad (4.4\text{-}16)$$

If α is small enough that $(\alpha w)^2$ is negligible compared with αw, we can make the approximations $\cosh \alpha w \doteq 1$ and $\sinh \alpha w \doteq \alpha w$. When we do so, (4.4-8) and (4.4-9) become

$$V(w) = (V(0) + Z_c I(0) \alpha w) \cos \beta w + j\,(Z_c I(0) + V(0) \alpha w) \sin \beta w \qquad (4.4\text{-}17)$$

$$I(w) = (I(0) + Y_c V(0) \alpha w) \cos \beta w + j\,(Y_c V(0) + I(0) \alpha w) \sin \beta w \qquad (4.4\text{-}18)$$

and (4.4-14) becomes

$$\frac{Z(w)}{Z_c} = \frac{\left(\dfrac{Z(0)}{Z_c} + \alpha w\right) + j\left(1 + \dfrac{Z(0)}{Z_c}\alpha w\right) \tan \beta w}{\left(1 + \dfrac{Z(0)}{Z_c}\alpha w\right) + j\left(\dfrac{Z(0)}{Z_c} + \alpha w\right) \tan \beta w} \qquad (4.4\text{-}19)$$

The standing-wave pattern on the line is most appropriately expressed in terms of the reflection coefficient $\Gamma(w) = |\Gamma(w)| e^{j\theta(w)}$. Making use of the expressions (4.4-3) and (4.4-4) for the voltage and current, and writing * to denote a complex conjugate, we have

$$\begin{aligned} |V(w)| &= \sqrt{V(w) V^*(w)} \\ &= |V^+(0)| \sqrt{e^{2\alpha w} + |\Gamma(0)|^2 e^{-2\alpha w} + 2|\Gamma(0)| \cos (2\beta w - \theta(0))} \end{aligned} \qquad (4.4\text{-}20)$$

and

$$|I(w)| = \sqrt{I(w)I^*(w)}$$

$$= |Y_c V^+(0)| \sqrt{e^{2\alpha w} + |\Gamma(0)|^2 e^{-2\alpha w} - 2|\Gamma(0)| \cos(2\beta w - \theta(0))} \quad (4.4\text{-}21)$$

Apart from the factor $|Y_c|$ in (4.4-21), the expressions for $|V|$ and $|I|$ differ only in the sign of the cosine term. We leave it as an exercise for the reader to demonstrate the perhaps surprising fact that successive standing-wave maxima and minima are *not* a half wavelength apart when $\alpha \neq 0$.

The power flowing in the line toward the load is

$$P = \frac{1}{2} \text{ real part } (VI^*) \quad (4.4\text{-}22)$$

(* denotes complex conjugate). In terms of forward and reflected waves, (4.4-22) becomes

$$P = \frac{1}{2} \text{ real part } \quad [(V^+ + V^-) \; Y_c^* \quad (V^{+*} - V^{-*})] \quad (4.4\text{-}23)$$

and, if Y_c is real,

$$P = \frac{1}{2} Y_c (|V^+|^2 - |V^-|^2) = \frac{1}{2} Y_c |V^+|^2 (1 - |\Gamma|^2) \quad (4.4\text{-}24)$$

In this expression, $\frac{1}{2} Y_c \; |V^+|^2$ and $\frac{1}{2} Y_c \; |V^-|^2$ are clearly the amounts of power in the forward and reflected waves respectively, and the net forward power is the difference between them.

When the characteristic immittance is not purely real it is not possible to separate the power flow into a forward part that is proportional to $|V^+|^2$ and a reflected part proportional to $|V^-|^2$. We leave it to the reader to show that, with a given forward voltage, the power delivered to the load will be maximum when the load immittance is equal to the complex conjugate of the characteristic immittance. Thus when the characteristic immittance is not real, maximum power is not delivered to a nonreflecting load.

4.5 APPLICATION OF SIGNAL FLOW GRAPHS TO MICROWAVE CIRCUITS[†]

Figure 4.5-1 shows a two-port with a current and a voltage so defined at each port that ½ real part $(V_1 I_1{}^*)$ and ½ real part $(V_2 I_2{}^*)$ are the amounts of power flowing *into* ports 1 and 2 respectively. The characteristic impedances of the lines in which ports 1 and 2 are located are Z_{c1} and Z_{c2}. The ingoing and outgoing voltages at each port are given in terms of the corresponding V and I by

$$V^+_{1,2} = \frac{1}{2}\ (V_{1,2} + Z_{c1,2}\, I_{1,2}) \tag{4.5-1}$$

$$V^-_{1,2} = \frac{1}{2}\ (V_{1,2} - Z_{c1,2}\, I_{1,2}) \tag{4.5-2}$$

and the scattering matrix of the two-port is defined by

$$\begin{pmatrix} V_1^- \\ V_2^- \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} V_1^+ \\ V_2^+ \end{pmatrix} \tag{4.5-3}$$

We gave a simplified introduction to the scattering matrix in Chapter 3, and we shall not elaborate on that here. The reader will find a discussion of *s*-matrix theory in any text on microwave circuits.

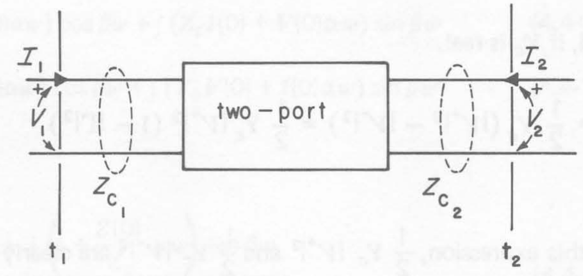


Figure 4.5-1.

Example: The scattering matrix of a length l of transmission line is

$$\mathbf{S}_{\substack{\text{transmission}\\\text{line}}} = \begin{pmatrix} 0 & e^{-\gamma l} \\ e^{-\gamma l} & 0 \end{pmatrix} \tag{4.5-4}$$

[†] J. K. Hunton, "Analysis of Microwave Techniques by Means of Signal Flow Graphs," *IRE Transactions on Microwave Theory and Techniques,* Vol MTT-8, pp 206-212, March 1960.

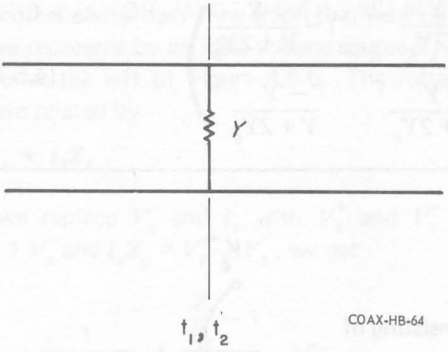


Figure 4.5-2.

Example: Let us calculate the scattering matrix of an admittance Y shunted across a line of characteristic admittance Y_c. We shall take both terminal planes t_1 and t_2 to be coincident with the plane of Y (Figure 4.5-2).

When port 2 is terminated by a reflectionless load, the admittance Y_1 that one sees if one looks into port 1 is $Y + Y_c$, and the reflection coefficient one sees, which is equal to s_{11}, is

$$\left(\Gamma_1\right)_{\text{reflectionless load on port 2}} = \frac{1 - \dfrac{Y_1}{Y_c}}{1 + \dfrac{Y_1}{Y_c}} = \frac{-Y}{Y + 2Y_c} = s_{11} \qquad (4.5\text{-}5)$$

Obviously the total voltages at the two ports have to be equal, so

$$V_1^+ + V_1^- = V_2^+ + V_2^-$$

or, dividing by V_1^+,

$$1 + \frac{V_1^-}{V_1^+} = \frac{V_2^+}{V_1^+} + \frac{V_2^-}{V_1^+}$$

Now, if port 2 is again terminated in a reflectionless load, so that $V_2^+ = 0$, the last equation yields a relation between s_{11} and s_{21}:

$$1 + s_{11} = s_{21} \qquad (4.5\text{-}6)$$

Since the shunt admittance is a symmetric obstacle, the scattering matrix is symmetric and we may combine (4.5-5) and (4.5-6) and write

$$\mathbf{S}_{\text{shunt admittance}} = \begin{pmatrix} \dfrac{-Y}{Y+2Y_c} & 1-\dfrac{Y}{Y+2Y_c} \\ 1-\dfrac{Y}{Y+2Y_c} & \dfrac{-Y}{Y+2Y_c} \end{pmatrix} \tag{4.5-7}$$

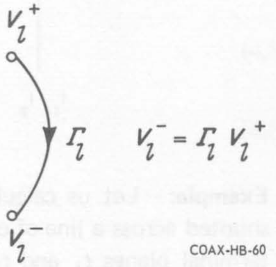


Figure 4.5-3. Flow graph consisting of a single branch.

A flow graph is a connected system of directed line segments that represents a set of algebraic equations. The vertices or **nodes** of the graph are the variables in the equations and the line segments or **branches** are the coefficients. In the flow graphs that are used to describe microwave circuits, the nodes are the ingoing and outgoing voltages, the V^+'s and V^-'s, and the branches are s-parameters and reflection coefficients. Consider for example the equation

$$V_l^- = \Gamma_l V_l^+ \tag{4.5-8}$$

which describes the reflection due to a load. The flow graph of (4.5-8) is shown in Figure 4.5-3.

The flow graph of equation 4.5-3, which describes a two-port in terms of its scattering matrix, is shown in Figure 4.5-4.

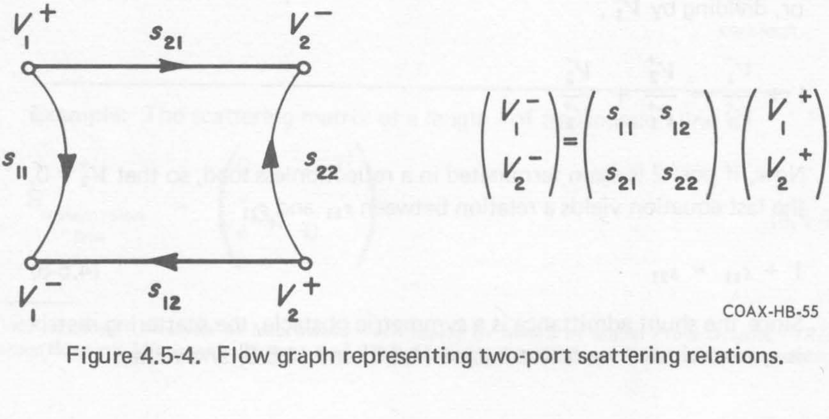


Figure 4.5-4. Flow graph representing two-port scattering relations.

Another elementary flow graph that we shall need to use is that of a source, which we represent by an ideal voltage source E_s in series with an impedance Z_s, as shown at the left of Figure 4.5-5. The voltage and current at the terminal plane t are related by

$$V_s = E_s + I_s Z_s \tag{4.5-9}$$

and if we replace V_s and I_s with V_s^+ and V_s^- by substituting the relations $V_s = V_s^+ + V_s^-$ and $I_s Z_c = V_s^+ - V_s^-$, we get

$$V_s^- = E_s \cdot \frac{1}{\frac{Z_s}{Z_c} + 1} + \frac{\frac{Z_s}{Z_c} - 1}{\frac{Z_s}{Z_c} + 1} V_s^+$$

which we will write

$$V_s^- = S + \Gamma_s V_s^+ \tag{4.5-10}$$

where

$$S = E_s \cdot \frac{1}{\frac{Z_s}{Z_c} + 1} \tag{4.5-11}$$

and

$$\Gamma_s = \frac{\frac{Z_s}{Z_c} - 1}{\frac{Z_s}{Z_c} + 1} \tag{4.5-12}$$

The flow graph of equation 4.5-10 is shown at the right of Figure 4.5-5.

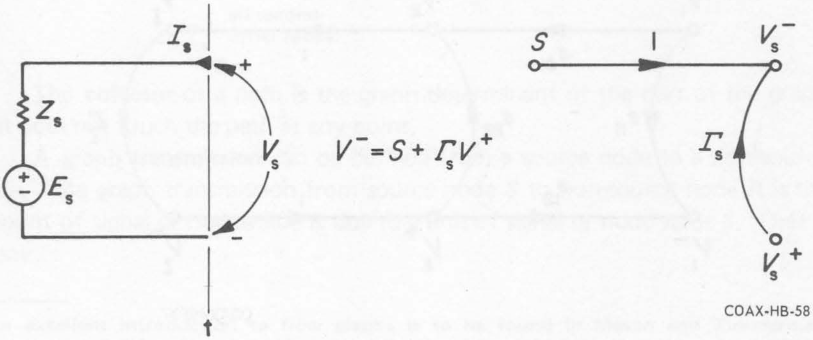


Figure 4.5-5. Equivalent circuit and flow graph of a source.

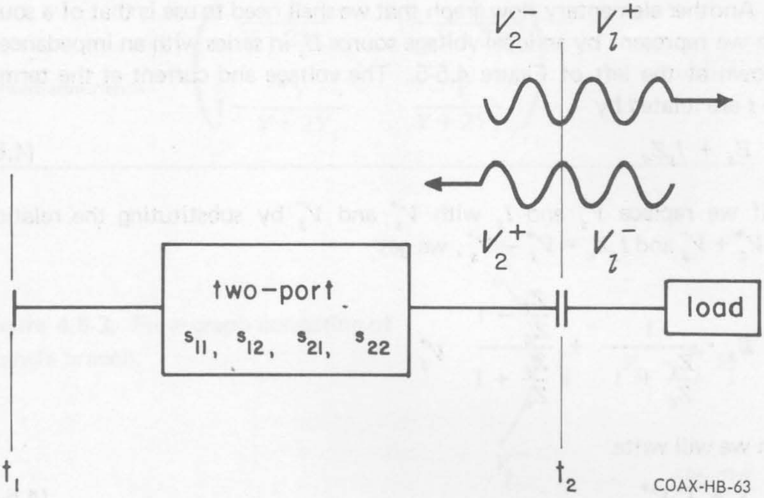


Figure 4.5-6. A two-port terminated in a load.

When two microwave components are connected together, the outgoing wave at one port is the incident wave at the abutting one. Accordingly we may obtain the flow graph of an entire system by plugging together the flow graphs of its individual parts so that each V^- coincides with the abutting V^+. Thus if we connect a load to port 2 of a two-port device, as in Figure 4.5-6, the flow graph of the composite system is obtained by combining the graph of Figure 4.5-4 with that of Figure 4.5-3; it is shown in Figure 4.5-7.

The importance of flow graphs is that they are the basis of a powerful topological method for calculating measurable parameters — immittances, reflection

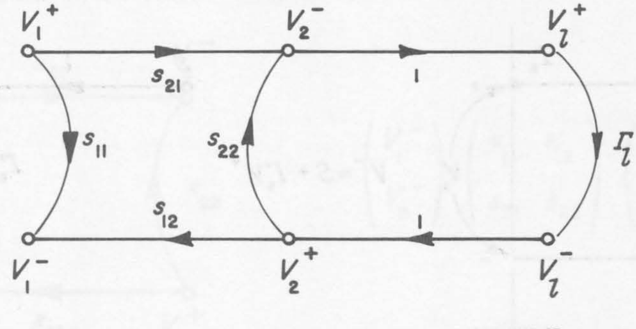


Figure 4.5-7. Flow graph of the two-port and load shown in Figure 4.5-6.

coefficients, losses, phase shifts, etc — in microwave systems. By means of the **nontouching loop rule** one can write down expressions for these quantities more or less from an inspection of the flow graph.[†]

We shall require a few definitions.

The **branches** are the directed line segments out of which the flow graph is made.

The **nodes** of the graph are the points at which the branches begin and end.

A **source node** is a node to which are attached only exiting branches.

A **node value** or **signal** is the value of the variable associated with a node. We practice the economy of using the letter that stands for the value also to label the node. The V^+'s and V^-'s in our graphs are node values.

A **branch transmission** is the value of the coefficient that is associated with a branch. The s's of Figure 4.5-7 are branch transmissions.

A **path** is a set of consecutive, codirectional branches along which no node is encountered more than once.

A **path transmission** is the product of the branch transmissions along a path.

A **first-order loop** is a closed path on which any node is encountered just once per circuit.

The meaning of **first-order loop transmission** will be obvious.

A **second-order loop** is two first-order loops that do not touch.

The meaning of **second-order loop transmission** will be obvious.

A **third-order loop** is three first-order loops that do not touch, etc.

The **graph determinant** is given by

$$\text{graph determinant} = 1 - \sum_{\substack{\text{all first-}\\ \text{order loops}}} \text{first-order loop transmissions} + \sum_{\substack{\text{all second-}\\ \text{order loops}}} \text{second-order loop transmissions} - \ldots \quad (4.5\text{-}13)$$

The **cofactor** of a path is the graph determinant of the part of the graph that does not touch the path at any point.

A **graph transmission** can be defined from a source node to a non-source node. The graph transmission from source node S to non-source node R is the amount of signal or node value R due to a unit of signal or node value S. That is to say,

---

[†] An excellent introduction to flow graphs is to be found in Mason and Zimmerman, "Electronic Circuits, Signals, and Systems," John Wiley and Sons, Inc., New York, 1960, Chapter 4.

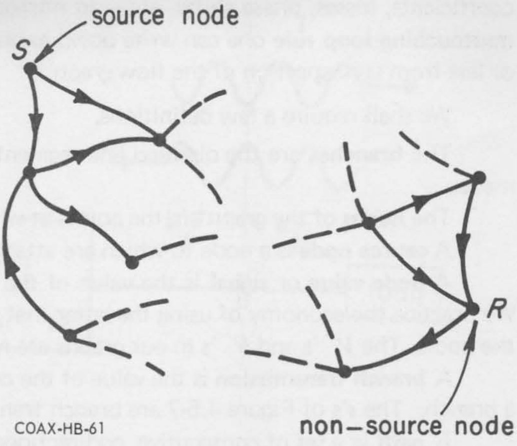


Figure 4.5-8. The graph transmission from source node S to non-source node R is R/S when all other source-node values are zero.

$$\begin{array}{l}\text{graph transmission from}\\ \text{source node } S \text{ to non-source node } R\end{array} = \left(\frac{R}{S}\right)_{\substack{\textbf{all other source-node}\\ \textbf{values equal to zero}}} \tag{4.5-14}$$

Finally, we can state the **nontouching loop rule** for calculating graph transmissions:

$$\begin{array}{l}\text{graph transmission}\\ \text{from } S \text{ to } R\end{array} = \frac{\displaystyle\sum_{\substack{\textbf{all paths}\\ \textbf{from } S \textbf{ to } R}} \begin{array}{c}\text{path}\\ \text{transmission}\end{array} \times \begin{array}{c}\text{path}\\ \text{cofactor}\end{array}}{\text{graph determinant}} \tag{4.5-15}$$

Example: We will illustrate the topological method by deriving the useful formula for the reflection coefficient $\Gamma_1 = V_1^-/V_1^+$ that one sees if one looks into port 1 of the loaded two-port shown in Figure 4.5-6. We observe from the flow graph of Figure 4.5-7 that V_1^+ is a source node and V_1^- is a non-source node, so that V_1^-/V_1^+ is the graph transmission from V_1^+ to V_1^-.

The graph has just a single first-order loop, $s_{22}\Gamma_l$, so that the graph determinant is $1-s_{22}\Gamma_l$. There are two paths from V_1^+ to V_1^-: s_{11} and $s_{21}\Gamma_l s_{12}$. The cofactor of s_{11} is $1-s_{22}\Gamma_l$, and the cofactor of $s_{21}\Gamma_l s_{12}$ is 1. Thus, according to (4.5-15),

$$\frac{V_1^-}{V_1^+} = \frac{s_{11}\ (1 - s_{22}\Gamma_l)\ +\ s_{21}\Gamma_l s_{12}}{1 - s_{22}\Gamma_l}$$

or

$$\Gamma_1 = s_{11} + \frac{s_{21}s_{12}\Gamma_l}{1 - s_{22}\Gamma_l} \qquad (4.5\text{-}16)$$

Example: As a further example let us consider the system shown in Figure 4.5-9, in which a two-port is terminated at port 2 in a load and at port 1 in a source. The flow graph is shown in Figure 4.5-10. We shall calculate the ratio V_2^-/S, the graph transmission from source node S to non-source node V_2^-.

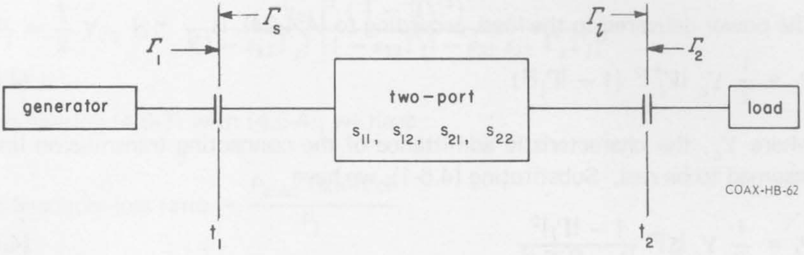


Figure 4.5-9. A two-port with source and load.

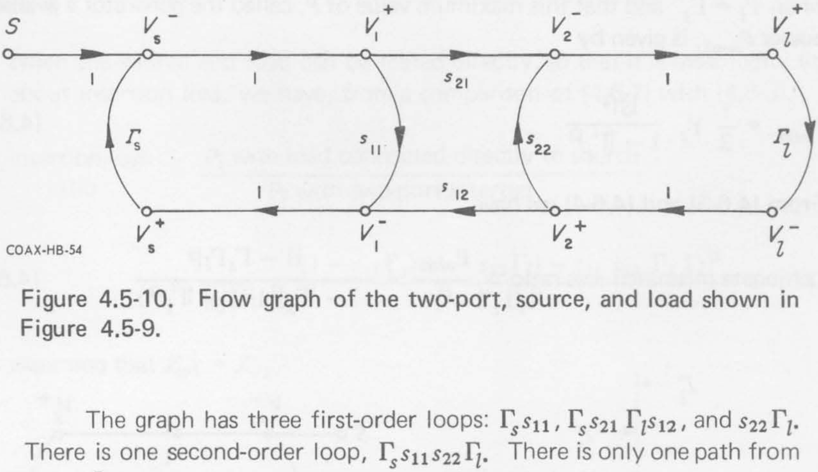


Figure 4.5-10. Flow graph of the two-port, source, and load shown in Figure 4.5-9.

The graph has three first-order loops: $\Gamma_s s_{11}$, $\Gamma_s s_{21}\Gamma_l s_{12}$, and $s_{22}\Gamma_l$. There is one second-order loop, $\Gamma_s s_{11} s_{22}\Gamma_l$. There is only one path from S to V_2^-; its transmission is s_{21} and its cofactor is 1. Thus

$$\frac{V_2^-}{S} = \frac{s_{21}}{1 - (\Gamma_s s_{11} + \Gamma_s s_{21}\Gamma_l s_{12} + s_{22}\Gamma_l) + \Gamma_s s_{11}s_{22}\Gamma_l}$$

$$= \frac{s_{21}}{(1 - s_{11}\Gamma_s)\,(1 - s_{22}\Gamma_l) - s_{21}s_{12}\Gamma_s\Gamma_l} \qquad (4.5\text{-}17)$$

4.6 LOSS FORMULAS

In this section we shall derive the basic formulas for the different kinds of loss that we discussed in Sections 3.3 and 3.4.

Let us first consider the transfer of power from a source to a load that is directly connected to it (Figure 4.6-1). Using the method of the last section, one finds from inspection of the flow graph that the incident wave at the load V_l^+ is

$$V_l^+ = \frac{S}{1 - \Gamma_s \Gamma_l} \tag{4.6-1}$$

The power delivered to the load, according to (4.4-24), is

$$P_l = \frac{1}{2} Y_c \, |V_l^+|^2 \, (1 - |\Gamma_l|^2) \tag{4.6-2}$$

where Y_c, the characteristic admittance of the connecting transmission line, is assumed to be real. Substituting (4.6-1), we have

$$P_l = \frac{1}{2} Y_c \, |S|^2 \, \frac{1 - |\Gamma_l|^2}{|1 - \Gamma_s \Gamma_l|^2} \tag{4.6-3}$$

We leave it to the reader to show that, when S and Γ_s are fixed, P_l is maximum when $\Gamma_l = \Gamma_s^*$ and that this maximum value of P, called the generator's **available power** P_{avail}, is given by

$$P_{\text{avail}} = \frac{1}{2} Y_c \, \frac{|S|^2}{1 - |\Gamma_s|^2} \tag{4.6-4}$$

From (4.6-3) and (4.6-4) we have

$$\text{conjugate mismatch loss ratio} = \frac{P_{\text{avail}}}{P_l} = \frac{|1 - \Gamma_s \Gamma_l|^2}{(1 - |\Gamma_s|^2)\,(1 - |\Gamma_l|^2)} \tag{4.6-5}$$

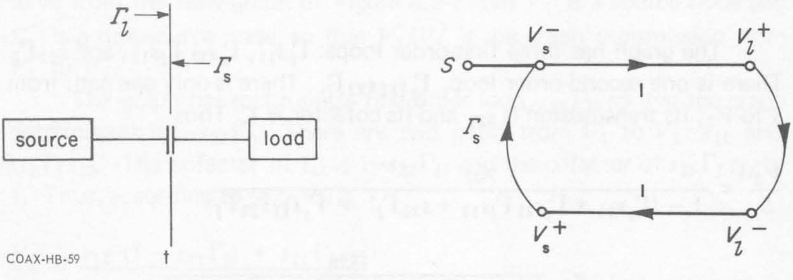


Figure 4.6-1. Schematic diagram and flow graph of source and directly connected load.

If a two-port is inserted between the source and the load, the load power will be

$$P_l = \frac{1}{2} Y_{c2} \; |V_2^-|^2 \; (1 - |\Gamma_l|^2) \tag{4.6-6}$$

where Y_{c2} is the characteristic admittance of the port-2-to-load junction and $V_2^- = V_l^+$ is the voltage of the wave leaving port 2 and entering the load. If we substitute (4.5-17), which gives the ratio V_2^-/S, into (4.6-2) we get

$$P_l = \frac{1}{2} Y_{c2} \; |S|^2 \; \frac{|s_{21}|^2 \; (1 - |\Gamma_l|^2)}{|(1 - s_{11}\Gamma_s)\,(1 - s_{22}\Gamma_l) - s_{21}\, s_{12}\; \Gamma_s \Gamma_l|^2} \tag{4.6-7}$$

comparing (4.6-7) with (4.6-4), we have

$$\text{transducer-loss ratio} = \frac{P_{\text{avail}}\ (\text{source})}{P_l}$$

$$= \frac{Y_{c1}}{Y_{c2}} \cdot \frac{1}{|s_{21}|^2} \cdot \frac{|(1 - s_{11}\Gamma_s)\,(1 - s_{22}\Gamma_l) - s_{21}\, s_{12}\; \Gamma_s\, \Gamma_l|^2}{(1 - |\Gamma_s|^2)\,(1 - |\Gamma_l|^2)} \tag{4.6-8}$$

When the source and load can be mated directly, so that it is meaningful to talk about insertion loss, we have, from a comparison of (4.6-7) with (4.6-3),

$$\begin{array}{c}\text{insertion-loss}\\ \text{ratio}\end{array} = \frac{P_l \text{ with load connected directly to source}}{P_l \text{ with two-port inserted}}$$

$$= \frac{1}{|s_{21}|^2} \cdot \frac{|(1 - s_{11}\Gamma_s)\,(1 - s_{22}\Gamma_l) - s_{21}\; s_{12}\; \Gamma_s\; \Gamma_l|^2}{|1 - \Gamma_s \Gamma_l|^2} \tag{4.6-9}$$

assuming that $Z_{c1} = Z_{c2}$.

CHAPTER 5

Basic Measurement Methods and Procedures

5.1 REFLECTOMETERS, BRIDGES, AND SLOTTED LINES

These are the three basic kinds of instruments that measure immittances or reflection coefficients at microwave frequencies. Although reflection coefficient and immittance are fundamentally one-port parameters, their measurement is the basis of many two-port methods, as well as methods for the measurement of such quantities as the propagation parameters of lines and the electric and magnetic properties of materials.

There are two different kinds of reflectometers. The frequency-domain reflectometer (FDR) makes cw measurements whereas the time-domain reflectometer (TDR) is a pulse instrument.

The principle of the frequency-domain reflectometer is shown in Figure 5.1-1. The directional coupler, whose schematic diagram is shown separately in Figure 5.1-2, is a four-port device with the following property: ports 1 and 2 are each coupled to ports 3 and 4 but not to each other, and likewise ports 3 and 4 are each coupled to ports 1 and 2 but not to each other. Thus a sample of the

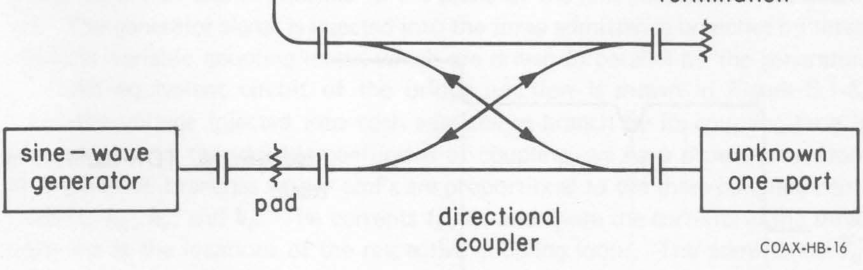


Figure 5.1-1. Basic frequency-domain reflectometer.

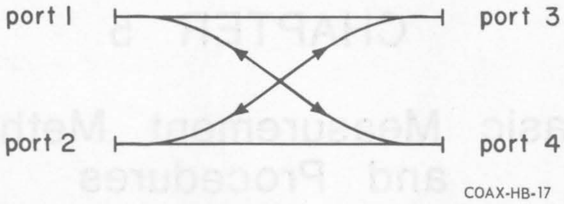


Figure 5.1-2. Schematic diagram of a directional coupler.

wave that is reflected from the unknown load in Figure 5.1-1 is diverted from the main transmission line and measured by the meter.

The basic time-domain reflectometer is shown in Figure 5.1-3. The principle of its operation is obvious; the echo from each discontinuity in the system under examination is displayed on the oscilloscope. A drawing of the sort of trace one might see on the oscilloscope is shown in Figure 5.1-4. The step at t_0 is the front of the generator's pulse. The small dip at t_1 shows that there is a small shunt capacitive discontinuity in the system located at a distance $\frac{1}{2}(t_1 - t_0)/v$ beyond the sampling point. The step at t_2 indicates a resistive load (whose resistance is larger than the characteristic impedance of the line) at a distance $\frac{1}{2}(t_2 - t_0)/v$ beyond the sampling point.

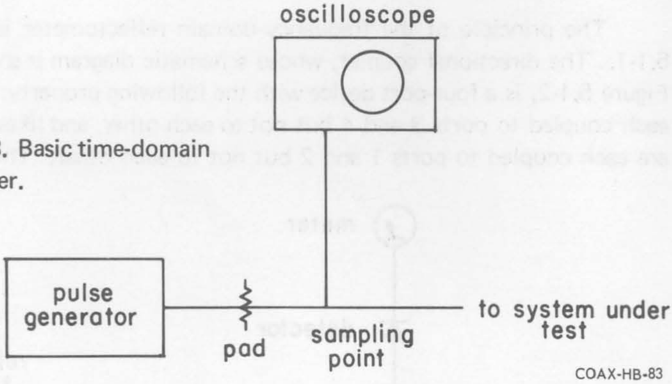


Figure 5.1-3. Basic time-domain reflectometer.

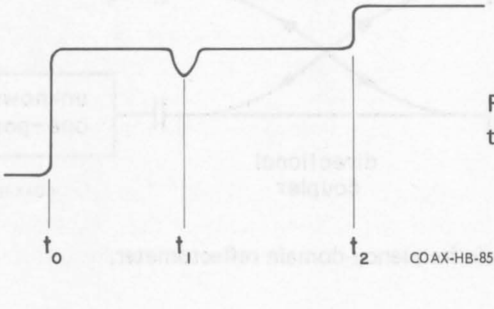


Figure 5.1-4. TDR oscilloscope trace.

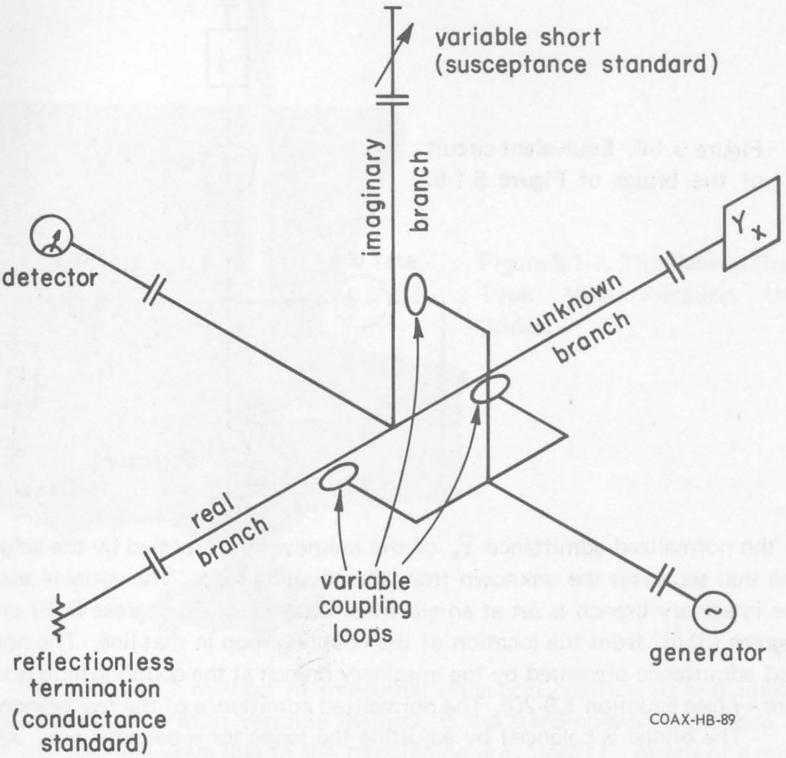
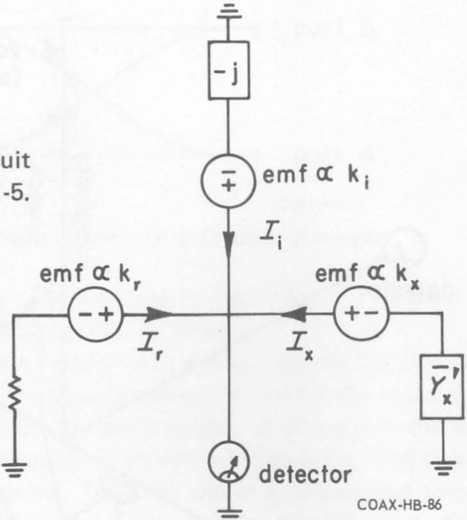


Figure 5.1-5. Schematic diagram of the Thurston bridge.

As an example from the large variety of microwave bridge arrangements, we shall describe the Thurston bridge, shown schematically in Figure 5.1-5. Three coaxial lines, the admittance branches, meet in a T-junction. These lines, the unknown branch, the real branch, and the imaginary branch, are terminated respectively in the unknown admittance, a reflectionless load, and a variable short. A fourth branch, normal to the plane of the junction, goes to the detector. The generator signal is injected into the three admittance branches by three identical variable coupling loops, which are driven in parallel by the generator.

An equivalent circuit of the bridge junction is shown in Figure 5.1-6. Since the voltage injected into each admittance branch by its coupling loop is proportional to the variable coefficient of coupling, we have shown generators in these three branches whose emf's are proportional to the three coupling coefficients, k_x, k_r, and k_i. The currents I_x, I_r, and I_i are the currents in the three branches at the locations of the respective coupling loops. The admittance $\bar{Y}_x'$ is the normalized admittance that we see at the location of the coupling loop in the unknown branch when we look toward the unknown admittance. It it equal

Figure 5.1-6. Equivalent circuit of the bridge of Figure 5.1-5.



to the normalized admittance $\overline{Y}_x$ of the unknown transformed by the length of line that separates the unknown from the coupling loop. The variable short in the imaginary branch is set at an electrical distance of 45 degrees ($\lambda/8$) or 135 degrees ($3\lambda/8$) from the location of the coupling loop in that line. The normalized admittance presented by the imaginary branch at the coupling loop is therefore $-j$ (see equation 1.8-20). The normalized admittance of the real branch is 1.

The bridge is balanced by adjusting the loops for a detector null. At balance, therefore, the currents in the three admittance branches add up to zero *at the junction.* If we assume that the distance from the loops to the junction is negligible, we have $I_x + I_r + I_i = 0$ at balance, and since $I_x \propto \overline{Y}'_x k_x$, $I_r \propto 1 \times k_r$, and $I_i \propto -jk_i$, this null condition leads to

$$\overline{Y}'_x = \frac{1}{k_x}(-k_r + jk_i) \quad \text{at balance}$$

We see from this equation that the real and imaginary components of $\overline{Y}'_x$ are proportional to the coupling coefficients k_r and k_i respectively, and both of them are multiplied by $1/k_x$. An indicator is attached to each coupling loop so as to show its angular position and hence the degree of coupling. The scales corresponding to k_r, k_i, and k_x are calibrated to read respectively the real and imaginary parts of $\overline{Y}'_x$ and a multiplying factor.

The assumption that the distance between the loops and the junction is zero is of course valid only at low frequencies. The bridge can be compensated to correct this error up to about 1500 MHz, which is therefore the upper limit to the useful frequency range of this sort of instrument.

Figure 5.1-7 is a photograph of the General Radio Type 1609 Precision UHF Bridge, which operates on the principle we have just described.

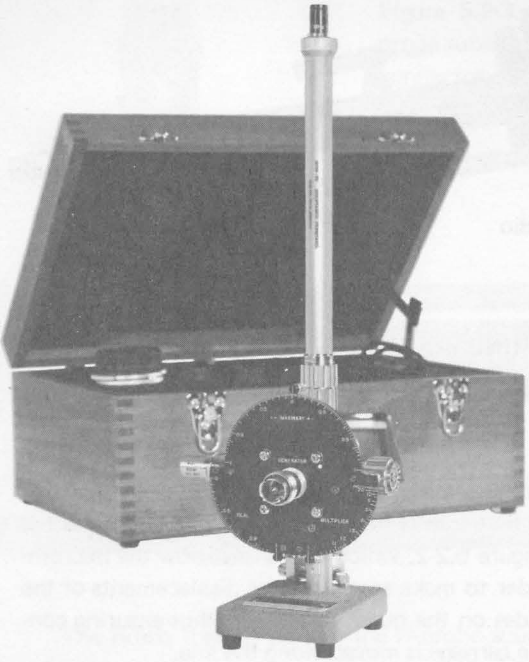


Figure 5.1-7. The General Radio Type 1609 Precision UHF Bridge.

The "classical" method of measuring reflection coefficients and immittances, and still the most versatile method, is the standing-wave technique, in which the standing wave due to the termination is explored by means of a movable probe inserted into a slotted line. The basic slotted-line arrangement for the measurement of an unknown immittance or reflection coefficient is shown in Figure 5.1-8. As we saw in Section 1.9, both the magnitude and phase of the reflection coefficient at the terminal plane can be determined from the standing-wave ratio on the slotted line and the position of the minima.

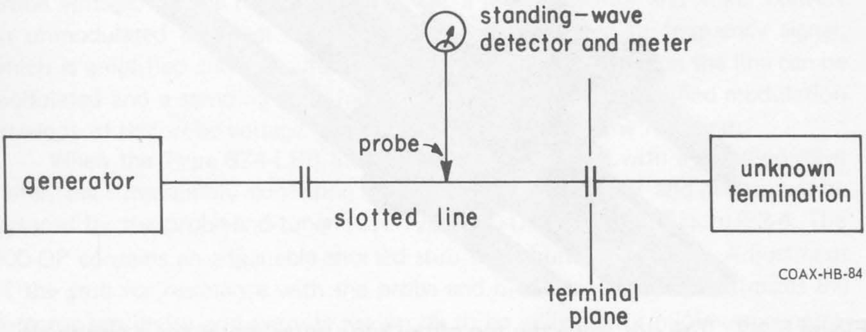


Figure 5.1-8. Basic arrangement for determination of an unknown immittance or reflection coefficient with a slotted line.

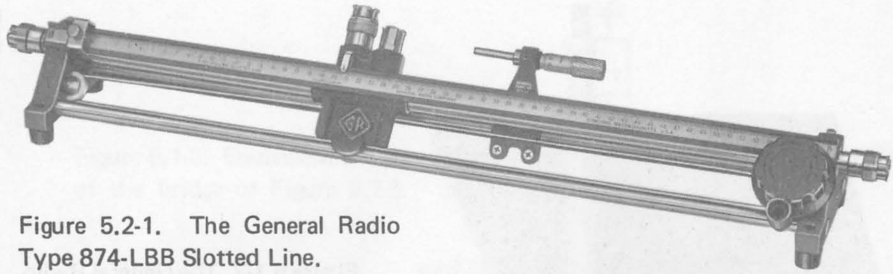


Figure 5.2-1. The General Radio Type 874-LBB Slotted Line.

5.2 STANDING-WAVE MEASURING EQUIPMENT

The General Radio 874-LBB coaxial slotted line is shown in Figure 5.2-1. It is a 50-centimeter section of rigid, 50-ohm air-dielectric line with a narrow axial slot in the outer conductor. A probe, which protrudes through the slot into the region between the conductors, samples the electric field in the line. The probe is mounted on a carriage that travels the length of the line. The slot is clearly seen in the close-up of Figure 5.2-2, which also shows how the micrometer is swung into position in order to make small, precise displacements of the carriage. The probe carriage slides on the outer conductor, thus ensuring constancy of probe penetration as the carriage is moved along the line.

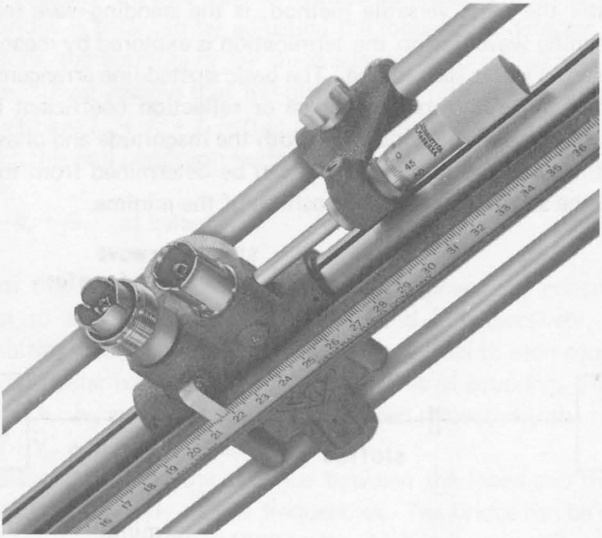


Figure 5.2-2. The slot along the top of the line's outer conductor is clearly visible in this close-up of the carriage assembly, which also shows the micrometer in position.

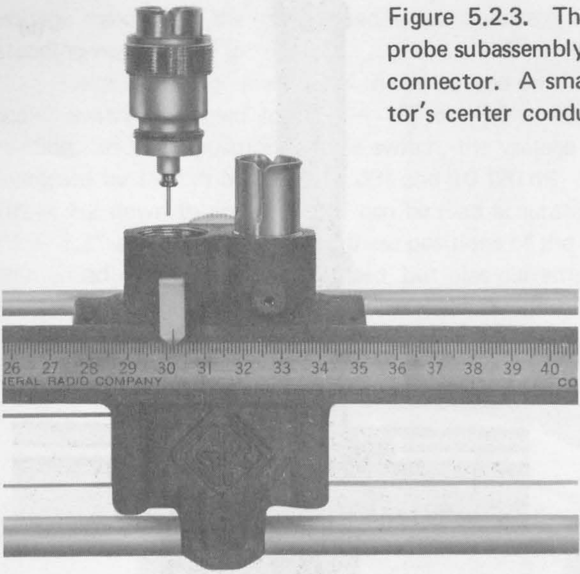


Figure 5.2-3. The probe is part of the rf-probe subassembly that includes the left-hand connector. A small screw inside the connector's center conductor adjusts probe depth.

The probe itself is part of the rf-probe subassembly that includes the left-hand connector on the carriage (Figure 5.2-3). This connector's center conductor is electrically connected to the probe, and a small screw inside the center conductor adjusts the probe depth. A grounded sleeve, part of the main carriage assembly, shields the probe as it passes through the slot, thus preventing changes in probe voltage due to variations in capacitance to the slot walls as the carriage is moved. The probe carriage also houses an envelope detector consisting of a diode rectifier and by-pass capacitor. The rectified probe voltage is brought to the right-hand connector on the carriage.

There are two essentially different methods that can be used to detect the probe voltage. In the heterodyne method, a local oscillator and mixer convert an unmodulated rf signal on the probe to an intermediate-frequency signal, which is amplified and measured. Alternatively, the rf signal on the line can be modulated and a standing-wave meter used to measure the rectified modulation envelope of the probe voltage. We shall discuss this latter method first.

When the Type 874-LBB Slotted Line is to be used with a standing-wave meter, the subassembly consisting of the left-hand-connector and probe may be replaced by the probe-and-tuner (GR Type 900-DP) shown in Figure 5.2-4. The 900-DP contains an adjustable shorted stub that shunts the probe. Adjustment of the stub for resonance with the probe and diode capacitance maximizes the detector sensitivity and permits the probe to be adjusted for minimum penetration into the line. The right-hand connector, at which the rectified modulation envelope is available, is connected to the standing-wave meter. When ease of

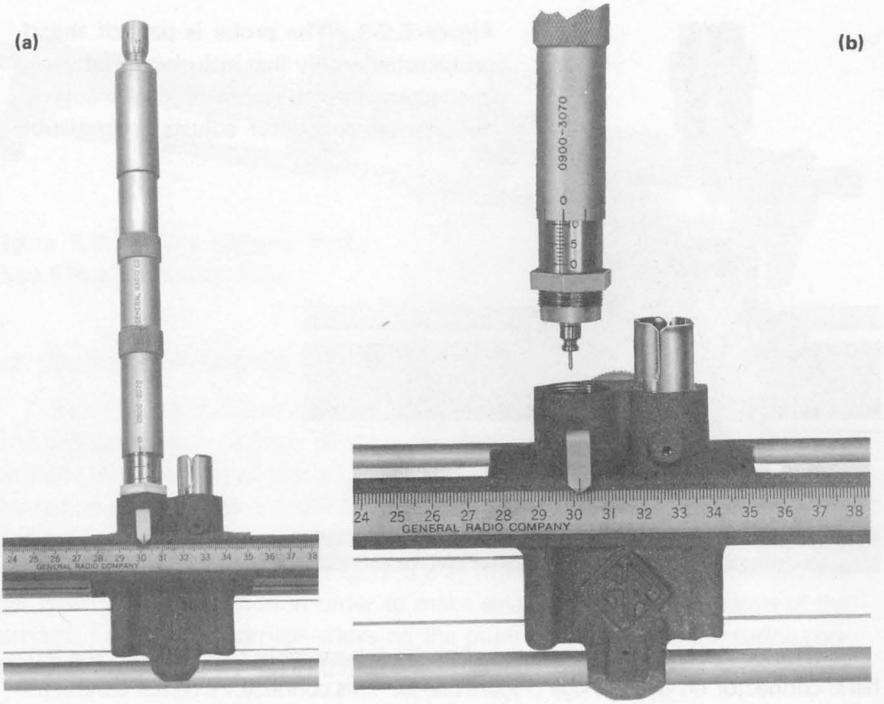


Figure 5.2-4. (a) When the Type 874-LBB Slotted Line is used with a standing-wave meter, the rf-probe subassembly (Figure 5.2-3) is replaced by the probe-and-tuner (Type 900-DP). The micrometer at the top of the 900-DP adjusts probe depth. (b) Close up of 900-DP removed from its seat in the probe carriage.

probe-depth adjustment is not important, an inexpensive alternative to the 900-DP is the rf-probe subassembly with a GR Type 874-D20L Adjustable Stub attached to its connector.

Modulation of the rf generator usually consists in keying the signal on and off with a fifty-percent duty cycle at a rate of 1 kHz. Sine-wave amplitude modulation of an oscillator is usually accompanied by an objectional amount of frequency modulation.

A standing-wave meter is a 1-kHz tuned amplifier preceded by a calibrated adjustable attenuator and followed by a rectifier and meter. The GR 1234 Standing-Wave Meter is shown in Figure 5.2-5. The numbers on the meter's SWR scales are proportional to the reciprocal of the *square* of the 1-kHz voltage at the input; if the detector has a *square-law response*—a point we shall take up presently—the meter readings are therefore proportional to the reciprocal of the probe voltage. Thus, if the meter reading is 1.0 (0 dB) when the probe is at a voltage maximum, the meter reading at a voltage minimum is equal to the standing-wave ratio r (or r(dB)).

Large SWR's (greater than 4.0) can be read on the Type 1234 if the "meter scale" switch is turned to "3.2 – 10" or "10 – 40" to obtain the minimum reading. In these positions of the switch, the voltage gain ahead of the meter is increased by factors of 3.16 (10 dB) and 10 (20 dB) respectively. Small SWR's (from 1.2 down to about 1.001) can be read accurately on the expanded scales "1 – 1.2" and "1 – 1.05." In these positions of the "meter scale" switch, the gain ahead of the meter is increased, but bias currents are applied to the meter

Figure 5.2-5. The GR Type 1234 Standing-Wave Meter.

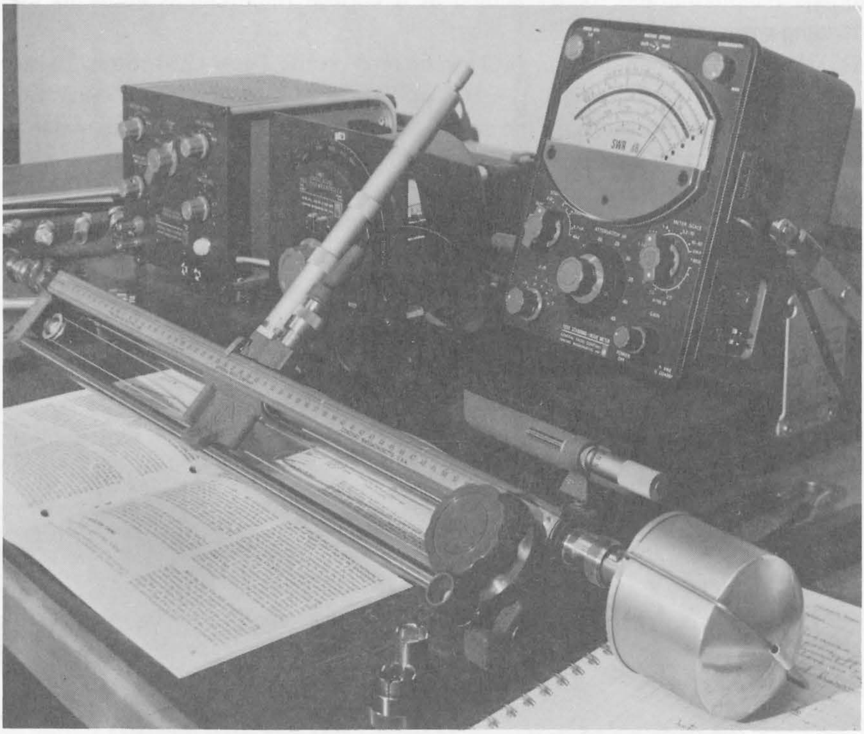


Figure 5.2-6. The General Radio Type 874-LBB Slotted Line with Type 900-DP probe-and-tuner and Type 1234 SWR Meter in a typical laboratory setup. Behind the slotted line on the left is the modulating power supply and next to it is the oscillator. The device under test is mounted in the shielded component mount attached to the right-hand end of the slotted line.

that offset its reading downscale. The combined effect of the increased gain and the offset is that a full-scale reading of 1.0 (0 dB) on the expanded scales occurs for the same detector voltage as on the "1 – 4" scales but, because of the higher gain, decreases in the probe voltage by factors of only 1.2 and 1.05 move the needle all the way downscale in these ranges.

If gross errors are to be avoided, two additional components must be included in the slotted-line setup. These are an attenuator or isolator and a low-pass harmonic filter, both inserted between the generator and the slotted line. The attenuator, 6 or 10 dB, serves to pad the oscillator from changes in its load impedance as various loads, including shorts and opens, are attached to the slotted line. Without the pad the oscillator frequency would be likely to change with such wide variations in load impedance. A ferrite isolator answers for this

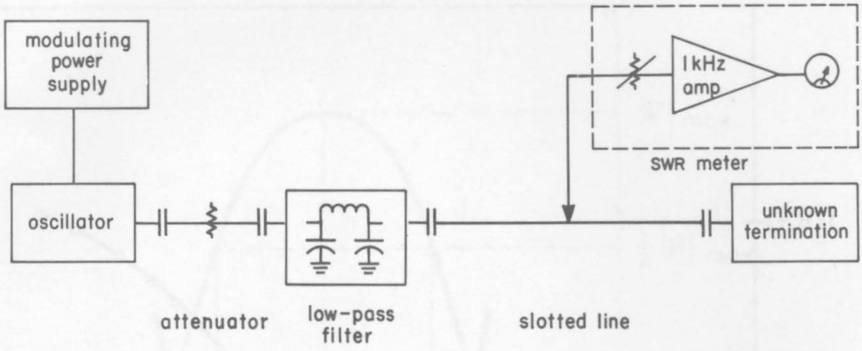


Figure 5.2-7. Schematic diagram of a slotted-line setup using a modulated oscillator and SWR meter.

purpose just as well. The low-pass filter keeps harmonics of the oscillator frequency out of the slotted line. This precaution is unnecessary when a heterodyne detector is used because of the high frequency selectivity of the heterodyne method, but the tuned probe and diode used with the SWR meter may have very little harmonic rejection because of higher-order resonances. When the SWR is high, even a small harmonic signal accompanying the oscillator output could lead to a totally erroneous measurement of the voltage at a standing-wave minimum.

When the applied signal is sufficiently small, a diode detector is a square-law device, that is, the rectified output voltage is proportional to the square of the rf input voltage. We have already said that the calibration of the SWR meter is based on the assumption that the diode is operating in its square-law range. Unfortunately one cannot state what the square-law range of a diode is, because the rectification characteristic depends not only on the individual diode but also on the source and load impedances that the diode sees, but the curve of Figure 5.2-8 is representative.

It is important to be able to determine whether, under given circumstances (frequency, probe depth, stub length, oscillator power, diode), the detector is within its square-law range. The slotted line itself can be made to function as an accurately calibrated attenuator in a precise determination of the detector response characteristic. We show in Appendix A to this chapter that if the slotted line has a totally reflecting termination, the standing wave (whose SWR is of course infinite) has the form of a rectified sinusoid, as shown in Figure 5.2-9. In this special case, the electrical distance along the line between a voltage maximum and the position on either side of the maximum where $|V| = \frac{1}{2} |V|_{max}$ is 60 degrees or 1/6 wavelength. Thus the following procedure may be used as a check on the proper response of the detector. (1) Terminate the slotted line in

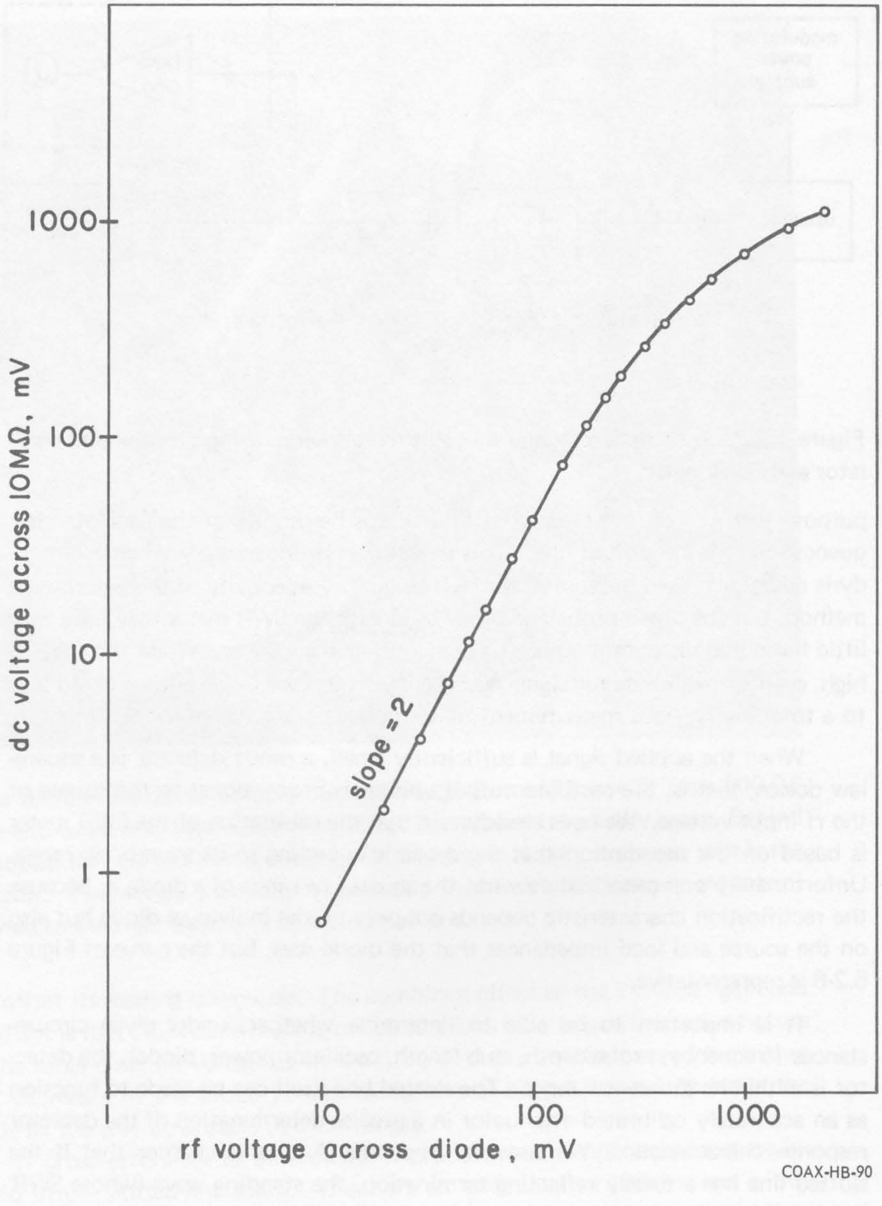


Figure 5.2-8. Typical rectification characteristic of a silicon point-contact microwave diode. In this particular case the diode exhibited a square-law response up to about 100 mV of rf input.

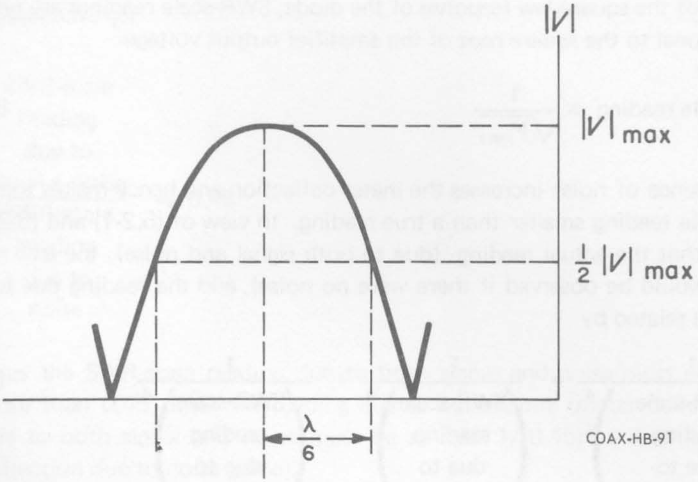


Figure 5.2-9. When the termination on the slotted line is totally reflecting, the standing wave has the form of a rectified sinusoid and the half-voltage points are 60 degrees or λ/6 from the maxima.

an open- or short-circuit. (2) Locate two consecutive minima. The point midway between them is a maximum (we *measure* minimum positions and *compute* maximum positions because minima are sharper). (3) Take a SWR-meter reading at the maximum. (4) Take a second reading at a point λ/6 away from the maximum (2/3 of the way from the maximum to an adjacent minimum). In this latter position the probe voltage is one half its value at the maximum. (5) If the diode is operating in its square-law region, the second SWR-scale reading will be 6.02 dB higher than the first (SWR increases downscale). If the diode is being driven beyond its square-law region, the readings will differ by less than 6.02 dB. In the latter case, the power level at the detector must be reduced – by withdrawing the probe, detuning it, decreasing the oscillator output, or inserting additional attenuation between the oscillator and the slotted line.

The minimum amount of rf signal that must reach the diode is determined by the requirement that the noise generated in the first stage of the 1-kHz amplifier must not cause an appreciable error in the meter reading. One can make an estimate of the minimum usable signal from the following considerations. The output of the 1-kHz amplifier contains two components, one due to the signal and the other due to noise. If V is the voltage of this output signal, then, since signal and noise are uncorrelated,

$$V_{rms}^2 \text{ (signal + noise)} = V_{rms}^2 \text{ (signal)} + V_{rms}^2 \text{ (noise)}. \qquad (5.2\text{-}1)$$

Because of the square-law response of the diode, SWR-scale readings are *inversely* proportional to the *square root* of the amplifier output voltage:

$$\text{SWR-scale reading} \propto \frac{1}{\sqrt{V_{rms}}} \tag{5.2-2}$$

The presence of noise increases the meter deflection and hence makes the actual SWR-scale reading smaller than a true reading. In view of (5.2-1) and (5.2-2) we can see that the actual reading (due to both signal and noise), the true reading (which would be observed if there were no noise), and the reading due to noise alone are related by

$$\frac{1}{\left(\begin{matrix}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{signal + noise}\end{matrix}\right)^4} = \frac{1}{\left(\begin{matrix}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{signal}\end{matrix}\right)^4} + \frac{1}{\left(\begin{matrix}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{noise}\end{matrix}\right)^4} \tag{5.2-3}$$

Example: Suppose that we will tolerate at most a one-percent error due to noise in the SWR-scale reading. Since the SWR-scale reading in the presence of noise is *smaller* than the true reading (meter deflection is larger), the requirement of a one-percent maximum error means that

$$\frac{\begin{matrix}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{signal + noise}\end{matrix}}{\begin{matrix}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{signal}\end{matrix}} \geqslant 0.99$$

or, in view of (5.2-3),

$$\left[1 - \left(\frac{\begin{matrix}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{signal + noise}\end{matrix}}{\begin{matrix}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{noise}\end{matrix}}\right)^4\right]^{\frac{1}{4}} \geqslant 0.99$$

whence we get

$$\frac{\begin{array}{c}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{signal + noise}\end{array}}{\begin{array}{c}\text{SWR-scale}\\ \text{reading}\\ \text{due to}\\ \text{noise}\end{array}} \leqslant 0.45$$

Thus the SWR-scale reading due to both signal and noise must not be more than 0.45 times the reading due to noise alone (meter deflection due to both signal and noise must be at least $1/(0.45)^2 \doteq 4$ times the deflection due to noise alone).

The noise reading should be made by turning off the rf generator; disconnecting the SWR meter from the detector would radically change the source impedance that the 1-kHz amplifier sees and hence also the amount of noise that it generates.

The dynamic range of a point-contact diode used as a SWR detector, limited at the bottom by noise and at the top by deviation from square-law response, is typically 30 dB or better. A bolometer, while it is on the order of 10 dB less sensitive than a diode, has a dynamic range of about 50 dB. The bolometers that are usually used as SWR detectors are barretters rather than thermistors. The barretter is an ohmic device consisting of a piece of very fine platinum wire installed in the same kind of package that houses a point-contact diode. When the barretter is used as a standing-wave detector, it is supplied with a bias current of a few milliamperes by the SWR meter. The presence of an rf current in the wire causes a temperature rise — in addition to the already-elevated temperature due to the bias current. The increase in temperature causes an increase in the wire's dc resistance, which in turn causes an increase in the dc voltage drop along it. The thermal time constant of the barretter wire is short enough that the resistance changes can follow the 1-kHz modulation of the rf signal, and the 1-kHz fluctuations in the dc voltage drop across the barretter are applied to the input of the 1-kHz amplifier. The barretter's response is very precisely square-law as long as the dc bias current is very much larger than the rf current.

Heterodyne detection can be used as an alternative to the SWR meter when greater sensitivity and more accuracy are wanted, as is the case, for example, in the measurement of very high standing-wave ratios.

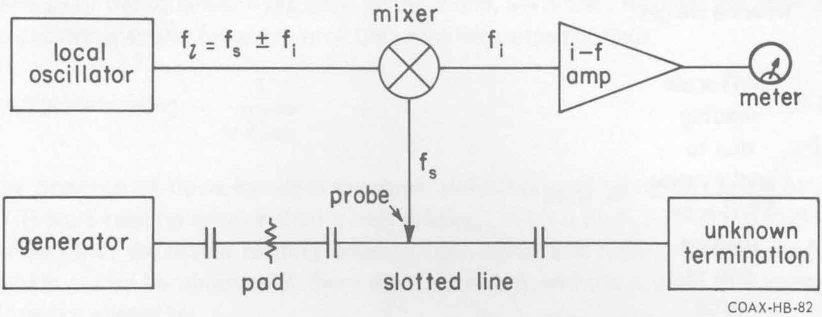


Figure 5.2-10. The heterodyne method of standing-wave detection.

When the Type 874-LBB Slotted-Line is used in a heterodyne arrangement, the 900-DP probe-and-tuner is not used. The rf-probe subassembly is installed in the left side of the probe carriage, the diode is removed from the probe carriage (the right-hand connector is unused), and the rf-input arm of a diode mixer (Type 874-MRAL, Figure 5.2-11) is connected directly to the probe through the left-hand connector. The signal on the slotted line is unmodulated and the local-oscillator frequency is offset from the signal frequency by an amount equal to the intermediate frequency, usually 30 MHz.

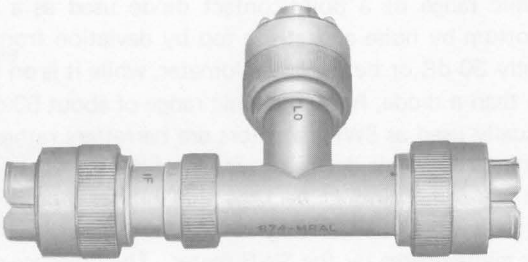
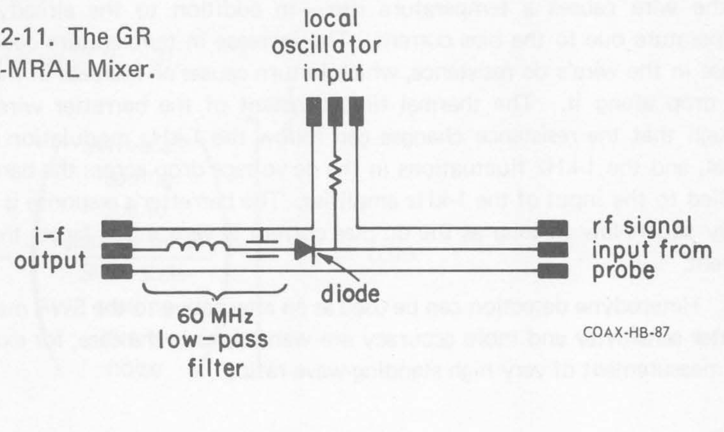


Figure 5.2-11. The GR Type 874-MRAL Mixer.



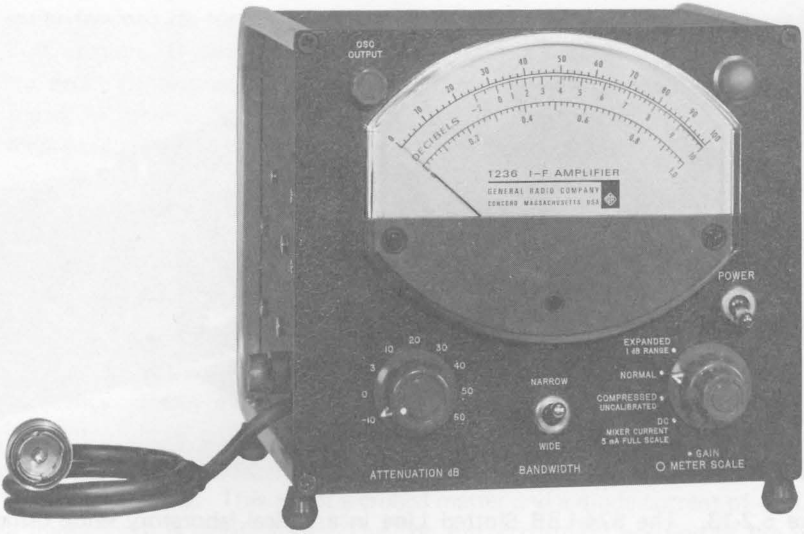


Figure 5.2-12. The GR Type 1236 30-MHz i-f amplifier and meter.

The GR Type 1236, a 30-MHz i-f amplifier and meter combination, is shown in Figure 5.2-12. To measure a standing-wave ratio with this instrument, one adjusts the "attenuation" and "gain" controls for a meter reading of 0 dB when the probe is at a standing-wave minimum. The reading on the dB scale when the probe is moved to a maximum, plus any change in the attenuator setting that is needed to keep the needle on the scale, is equal to the SWR in dB. The "expanded 1-dB range" position on the "meter scale" switch spreads the segment of the main scale between 9 and 10 dB out over the entire movement of the needle. This scale is used when the SWR is less than 1 dB (ratio less than 1.12). The "compressed uncalibrated" position of the "meter scale" switch turns on an automatic-gain-control loop whose threshold corresponds to a meter deflection of about 35 percent of full scale. A very wide range of input-signal levels is then compressed into the upper 65 percent of the meter movement. The "compressed uncalibrated" range facilitates the initial location of maxima and minima when the SWR is high.

Some care is needed in the adjustment of the local-oscillator frequency for heterodyne detection to make sure that the i-f signal is not due to harmonics of the local oscillator beating with harmonics of the probe signal. Two correct local oscillator frequencies are given by

$$f_l = f_s \pm f_i \quad \text{(correct local-oscillator frequency)} \qquad (5.2\text{-}4)$$

where the subscripts l, s, and i refer to the local oscillator, the probe signal, and the i-f signal respectively. Thus, with the usual intermediate frequency of

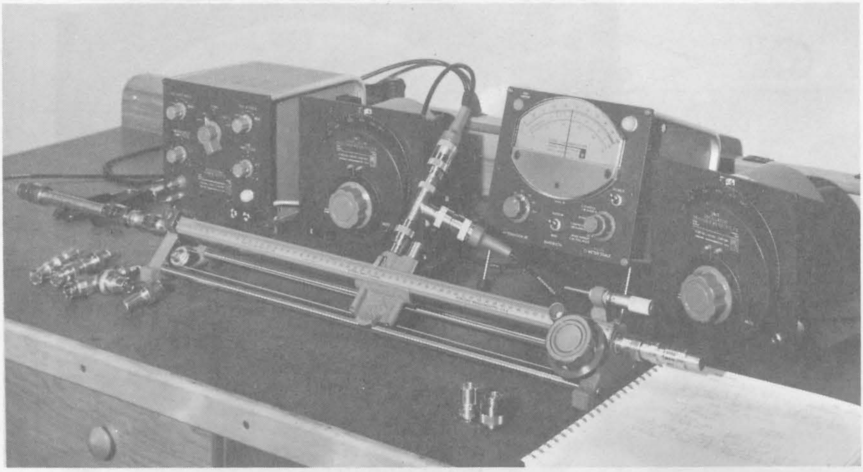


Figure 5.2-13. The 874-LBB Slotted Line in a typical laboratory setup using heterodyne detection. The Type 874-MRAL Mixer is attached to the left-hand connector on the probe carriage. The local oscillator (extreme right) is supplied with power from an auxiliary power take-off plug on the Type 1236 I-F Amplifier. To the left of the Type 1236 is the oscillator that drives the slotted line, and its power supply (extreme left) is switched to the cw mode since an unmodulated signal is detected by the heterodyne method.

30 MHz, correct i-f signals will be observed at two local-oscillator frequencies, 60 MHz apart, one on either side of the slotted-line frequency. If the i-f signal is due to the n^{th} harmonic of the slotted-line signal beating with the n^{th} harmonic of the local oscillator, the local-oscillator frequency will be

$$f_l = f_s \pm \frac{1}{n} f_i \quad \text{(spurious i-f signal due to } n^{\text{th}} \text{ harmonic)} \qquad (5.2\text{-}5)$$

These pairs of spurious images occur between the pair of correct local-oscillator frequencies. Any doubt about the local-oscillator setting can be resolved quickly by a check of the slotted-line wavelength to see that it corresponds to the generator's fundamental frequency.

We have seen that when a microwave diode is used as a demodulator it is operated at low enough levels—a few millivolts—that its detection characteristic is square-law. But as a mixer the diode functions as a switch that is turned on and off by the local oscillator. The conversion characteristic of the mixer is linear; that is, the level of the i-f signal is proportional to the level of the rf signal from the probe.

The functioning of the diode as an efficient mixer requires a relatively large local-oscillator signal, since the local-oscillator voltage has to push the diode quite far into its forward conduction region and quite far into its back-biased "off" region. If conversion is to be linear, the change in diode resistance due to the probe signal must be negligible compared with that due to the local-oscillator signal; in other words, the probe voltage must be very much smaller than the local-oscillator voltage.

The rectified direct current that flows through the diode is a measure of the mixer's conversion efficiency. Of course the diode current depends primarily on the local-oscillator signal level, but it is also somewhat frequency dependent—more power is needed to produce a given diode current at higher frequencies. Furthermore, as one can see from the schematic diagram of the mixer in Figure 5.2-11, the probe arm of the mixer shunts the path over which the local-oscillator signal gets to the diode, and the impedance of this shunt at the local-oscillator frequency affects the coupling of local-oscillator power to the diode. The amount of diode current that is optimum is a question of signal-to-noise ratio: larger currents generate more diode noise and smaller currents yield lower conversion efficiency. This is not a crucial matter and a diode current of about half a milliampere is satisfactory. The "dc mixer current" position of the "meter scale" switch on the Type 1236 I-F Amplifier puts the front-panel meter in series with the center conductor of the i-f input so that the diode current in the Type 874-MRAL Mixer can be checked. Since the probe is an open circuit for direct current, the local-oscillator output must provide a dc path to ground for the diode current.

5.3 MEASUREMENT OF ONE-PORT REFLECTION COEFFICIENTS AND IMMITTANCES BY THE STANDING-WAVE METHOD

In Section 1.9, Chapter 1, we discussed the way in which the terminal reflection coefficient on a lossless line determines the relative amount of standing-wave voltage variation and also the position of the standing wave. We saw there that the magnitude of the terminal reflection coefficient determines the standing-wave ratio, while its angle determines the position of the standing wave.

When the standing-wave ratio is neither too large nor too small it can be read directly from the scale of a standing-wave indicator in a completely straightforward manner. Very large or small SWR's may require some special techniques that we shall take up presently.

The angle of the terminal reflection coefficient is a little more complicated to find. The procedure involves the following steps. 1) Determination of the position of a standing-wave minimum when the unknown load terminates the slotted line. (One always measures the positions of minima rather than maxima. This is partly because minima are sharper and partly because they are perturbed less than maxima by the presence of the probe.) 2) Determination of the position of a standing-wave minimum when the slotted line is terminated in a short.

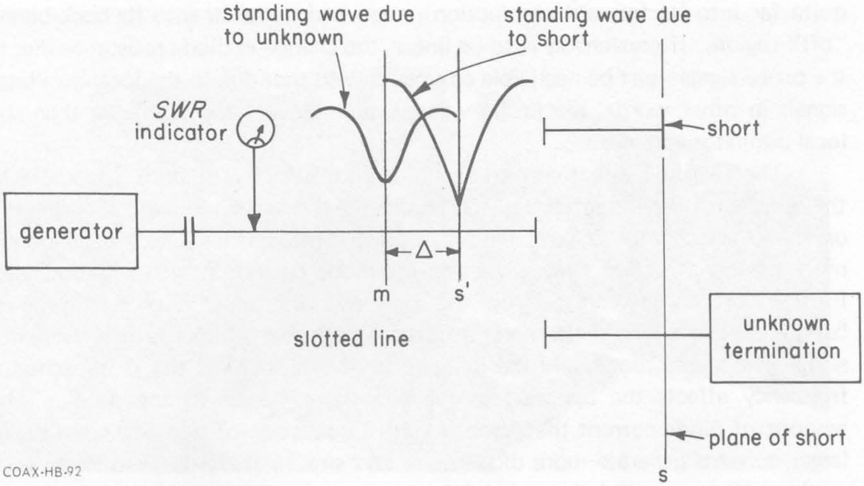


Figure 5.3-1. When a short replaces the unknown termination, the shift in the position of a standing-wave minimum determines the angle of the unknown reflection coefficient.

3) Measurement of the wavelength on the slotted line. This is just twice the distance between adjacent minima. (A slotted-line measurement of wavelength is likely to be far more accurate than a reading on the frequency dial of a microwave oscillator.) 4) When the short replaces the unknown termination, the new minima will not in general fall where the old ones did. Let Δ be the physical distance from a minimum due to the unknown to a minimum due to the short.[†] The angle θ_s of the unknown reflection coefficient *at a reference plane s that coincides with the plane of the short* is given by

$$\theta_s = 180 \text{ deg} \pm 2\beta\Delta \tag{5.3-1}$$

The $\left\{\pm\right\}$ sign applies when the minimum due to the short is on the $\left\{\begin{matrix}\text{load}\\ \text{generator}\end{matrix}\right\}$ side of the one due to the unknown.

Let us see why this procedure works. Determination of a minimum position when the line is shorted serves to identify a reference plane, let us call it s', that is exactly an integral number of half wavelengths from the plane s of the short. *At s', no matter how the line is terminated, the impedance and reflection coefficient are the same as those at s.* Now, when the unknown load terminates the line, a minimum will not in general fall at s'. Let us label with an m the posi-

---

[†] Since a standing wave has minima every half wavelength, a minimum due to the short will always fall within a quarter wavelength of a given minimum due to the unknown. Thus we shall assume for the sake of a definite picture that Δ is not greater than a quarter wavelength, although the result we shall give is valid when Δ is the distance from any minimum due to the unknown to any minimum due to the short.

tion of a minimum due to the unknown. The distance between s' and m is Δ. Since the angle θ of the reflection coefficient at a standing-wave minimum is 180 degrees, the angle $\theta(s')$ at s', hence also the angle $\theta(s)$ at s, is equal to 180 deg $\pm 2\beta\Delta$, + when s' is on the load side of m, − when s' is on the generator side of m.

Example: When an unknown load is connected to the slotted line the SWR is 1.6. When a GR Type 874-WN3 Short-Circuit Termination is connected to the line, the minimum position shifts 5.1 cm toward the load and the distance between minima is 12.0 cm. The plane of the short in the Type 874-WN3 is 3.0 cm toward the load from the front (genera-

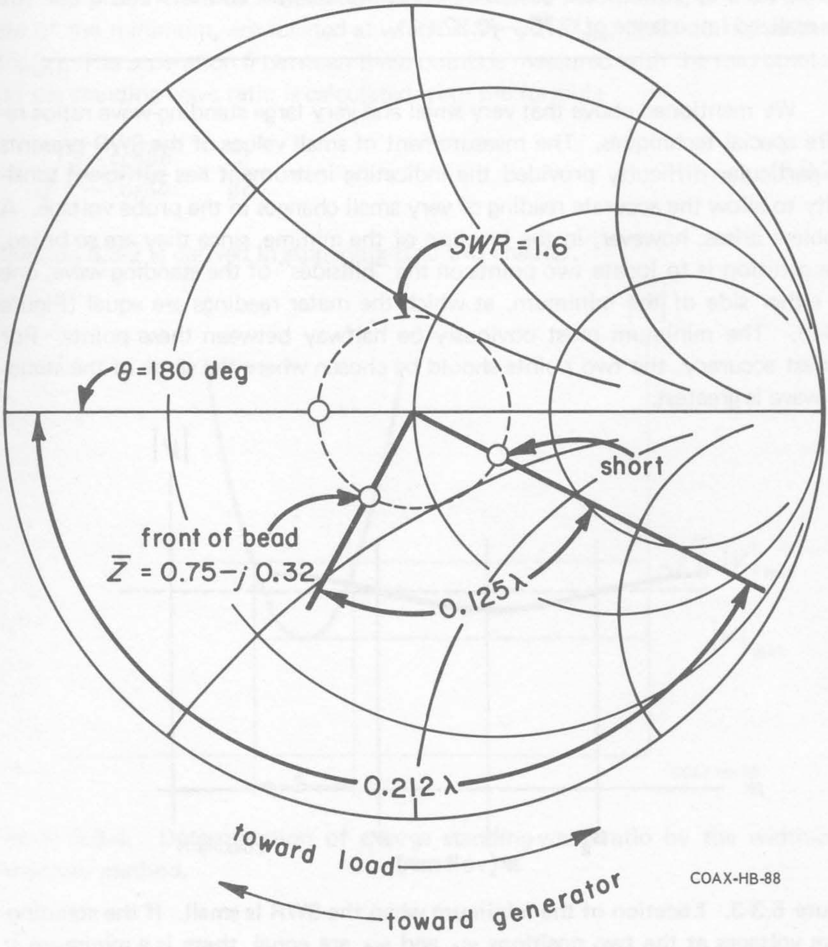


Figure 5.3-2.

tor side) surface of the bead in the 874-WN3's connector. What is the impedance of the unknown load at the front surface of the bead in the load's connector?

The wavelength λ is 2 X 12.0 cm = 24.0 cm, so that the minimum shift Δ of 5.1 cm is equal to 0.212 wavelength. If we enter the Smith chart (Figure 5.3-2) on the θ = 180 deg radial at a radius corresponding to a SWR of 1.6, and then go around the chart toward the load (because the minimum shifted toward the load) 0.212 wavelength, we arrive at the point on the chart that corresponds to the plane s of the short. But we don't want the impedance at this plane; we want it at the front surface of the bead, 3.0 cm = 0.125 wavelength toward the generator from s. So we back up toward the generator 0.125 wavelength and arrive at a normalized impedance of $0.75 - j0.32$.

We mentioned above that very small and very large standing-wave ratios require special techniques. The measurement of small values of the SWR presents no particular difficulty provided the indicating instrument has sufficient sensitivity to allow the accurate reading of very small changes in the probe voltage. A problem arises, however, in the location of the minima, since they are so broad. The solution is to locate two points on the "hillsides" of the standing wave, one on either side of the minimum, at which the meter readings are equal (Figure 5.3-3). The minimum must obviously be halfway between these points. For highest accuracy, the two points should be chosen where the slope of the standing wave is greatest.

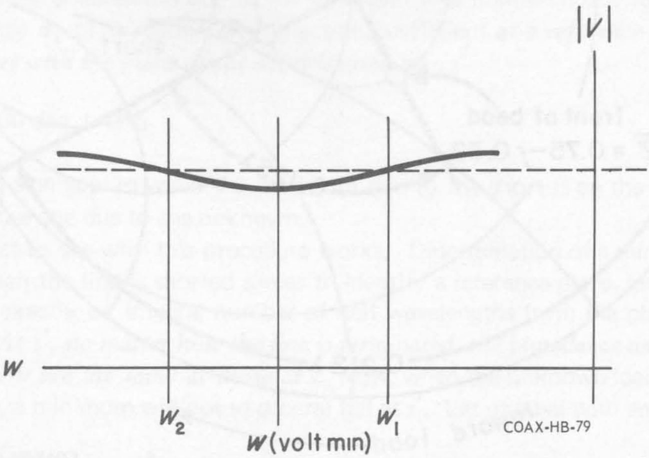


Figure 5.3-3. Location of the minimum when the SWR is small. If the standing-wave voltages at the two positions w_1 and w_2 are equal, there is a minimum at w(volt min) = ½ $(w_2 - w_1)$.

Large standing-wave ratios present the reverse problem. The minimum is sharp and easily located but the measured SWR is likely to be in error owing to perturbation of the standing wave by the probe. The probe introduces a shunt discontinuity into the slotted line. Since the impedance at a standing-wave minimum on a line with a high SWR is small, the effect of the probe is negligible when the probe is at a minimum. But at a standing-wave maximum, where the impedance is very high, the probe admittance may be a significant load across the line. The effect of loading by the probe is that measured SWR's are smaller than they would be if the probe were not there.

Because of probe loading, standing-wave ratios larger than about 10 should be determined by the **width-of-minimum** method, in which all readings are taken with the probe close to a standing-wave minimum. Two points, one on either side of the minimum, are located at which $|V| = \sqrt{2}|V|_{\min}$ (3.01 dB higher than $|V|_{\min}$). The separation δ between these points is measured with the micrometer, and the standing-wave ratio is calculated from the formula

$$r = \sqrt{\frac{3 - \cos\beta\delta}{1 - \cos\beta\delta}} \doteq \frac{2}{\beta\delta} \qquad (5.3\text{-}2)$$

Equation 5.3-2 is derived in Appendix B to this chapter.

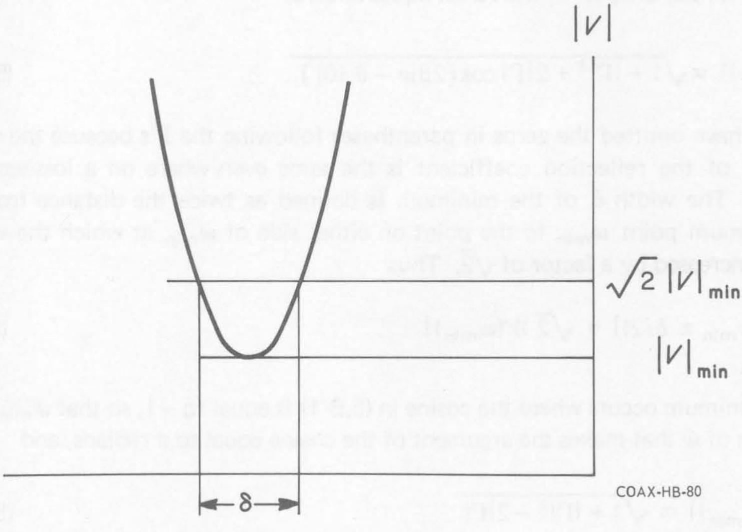


Figure 5.3-4. Determination of a large standing-wave ratio by the width-of-minimum method.

Appendix to Chapter 5

A. Shape of the Standing Wave when the Termination is Totally Reflecting

Equation 4.4-20, Chapter 4, gives the magnitude of the standing-wave voltage distribution. If we apply this formula to the case of an open or short on the end of a lossless line by setting $\alpha = 0$ and $\Gamma(0) = \pm 1$, we have

$$|V(w)| \propto \sqrt{2 \pm 2 \cos 2\beta w} \;=\; 2\sqrt{\frac{1}{2}\,(1 \pm \cos 2\beta w)} \;=\; \left\{\begin{matrix} 2|\cos\beta w| \\ 2|\sin\beta w| \end{matrix}\right\} \qquad (5.\text{A-}1)$$

Since the magnitudes of the cosine and sine fall to ½ at 60 degrees either side of a maximum, the half-voltage points are 60 degrees or λ/6 away from the maxima.

B. The Width-of-Minimum Formula

The standing-wave voltage distribution on a lossless line is given by equation 4.4-20, Chapter 4, with α set equal to zero:

$$|V(w)| \propto \sqrt{1 + |\Gamma|^2 + 2|\Gamma| \cos(2\beta w - \theta\ (0))} \qquad (5.\text{B-}1)$$

(We have omitted the zeros in parentheses following the Γ's because the magnitude of the reflection coefficient is the same everywhere on a lossless line.)

The width δ of the minimum is defined as twice the distance from the minimum point w_{min} to the point on either side of w_{min} at which the voltage has increased by a factor of $\sqrt{2}$. Thus

$$|V(w_{\text{min}} \pm \delta/2)| = \sqrt{2}\,|V(w_{\text{min}})| \qquad (5.\text{B-}2)$$

A minimum occurs where the cosine in (5.B-1) is equal to -1, so that w_{min} is the value of w that makes the argument of the cosine equal to π radians, and

$$|V(w_{\text{min}})| \propto \sqrt{1 + |\Gamma|^2 - 2|\Gamma|} \qquad (5.\text{B-}3)$$

At a distance $\delta/2$ either side of w_{min}, the argument of the cosine differs from π by $\beta\delta$, and since $\cos(\pi \pm \beta\delta) = -\cos\beta\delta$ we have

$$|V(w_{\text{min}} \pm \delta/2)| \propto \sqrt{1 + |\Gamma|^2 - 2|\Gamma|\cos\beta\delta} \qquad (5.\text{B-}4)$$

Substitution of (5.B-3) and (5.B-4) in (5.B-2) yields the equation

$$1 + |\Gamma|^2 - 2|\Gamma|\cos\beta\delta = 2(1 + |\Gamma|^2 - 2|\Gamma|)$$

and the replacement of $|\Gamma|$ by $(r-1)/(r+1)$ leads to

$$r = \sqrt{\frac{3-\cos\beta\delta}{1-\cos\beta\delta}} \tag{5.B-5}$$

If we replace $\cos\beta\delta$ by the first two terms, $1 - \frac{1}{2}\beta^2\delta^2$, of its power series, we get the approximate formula

$$r \doteq \sqrt{\frac{2+\frac{1}{2}\beta^2\delta^2}{\frac{1}{2}\beta^2\delta^2}} = \frac{2}{\beta\delta}\sqrt{1+\frac{\beta^2\delta^2}{4}} \doteq \frac{2}{\beta\delta}$$

A. Shape of the Standing [illegible]

Equation 4.4.20, Chapter 4, gives the [illegible] of the [illegible]. If we apply this formula to the case of [illegible] end of a lossless line by setting α = 0 and [illegible], we have [illegible]

[illegible]

Since the magnitudes of the voltage [illegible] a maximum, the [illegible]

B. The Width of Minimum Formula

The standing-wave voltage distribution on a lossless line is given by equation 4.4.20, Chapter 4, with α set equal to zero:

$$|V(x)| = \sqrt{1 + |\Gamma|^2 + 2|\Gamma|\cos(2\beta x - \theta_\Gamma)} \qquad (5.B.1)$$

We have omitted the [illegible] in parentheses following the Γ's because the magnitude of the reflection coefficient is the same everywhere on a lossless line.

The width Δ of the minimum is defined as twice the distance from the minimum point x_{min} to the point on either side of x_{min} at which the voltage has increased by a factor of $\sqrt{2}$. Thus

$$|V(x_{min} \pm \Delta/2)| = \sqrt{2}\,|V(x_{min})| \qquad (5.B.2)$$

A minimum occurs where the cosine in (5.B.1) is equal to −1, so that x_{min} is the value of x that makes the argument of the cosine equal to π radians, and

$$|V(x_{min})| = \sqrt{(1 - |\Gamma|)^2} \qquad (5.B.3)$$

At a distance Δ/2 either side of x_{min}, the argument of the cosine differs from π by βΔ, and since cos(π ± βΔ) = −cos βΔ we have

$$|V(x_{min} \pm \Delta/2)| = \sqrt{1 + |\Gamma|^2 - 2|\Gamma|\cos\beta\Delta} \qquad (5.B.4)$$

Table of Symbols

A — total attenuation of line section
a — radius of inner conductor of coaxial line
b — radius of outer conductor of coaxial line
c — shunt capacitance/unit length of line
velocity of light
D — dissipation factor of dielectric
e — 2.718 . . .
E — electric field strength
phasor electromotive force
f — frequency
g — shunt conductance/unit length of line
G — conductance
H — magnetic field strength
i — instantaneous current
I — phasor current
I^+ — phasor current of forward (or ingoing) wave
I^- — phasor current of reflected (or outgoing) wave
I_t — phasor current in the termination
$I(w)$ — phasor current at the location w
j — imaginary operator $(j = \sqrt{-1})$
K — surface current density
l — physical length of a section of line
series inductance/unit length of line
n — index of refraction
an integer
turns ratio of transformer
P — power
Q — dielectric Q
r — standing-wave ratio
series resistance/unit length of line
R — return loss
resistance
s — scattering parameter
S — amplitude of a source
t — symbol used to label reference plane of the termination
v — velocity
instantaneous voltage
V — phasor voltage

V^{+} —	phasor voltage of forward (or ingoing) wave
V^{-} —	phasor voltage of reflected (or outgoing) wave
V_t —	phasor voltage across the termination
$V(w)$ —	phasor voltage at the location w
w —	variable of position on a transmission line
W_E —	energy/unit length of line associated with the electric field
W_H —	energy/unit length of line associated with the magnetic field
y —	shunt admittance/unit length of line
Y —	admittance
$\overline{Y}$ —	normalized admittance
Y_c —	characteristic admittance of line
Y_t —	admittance of termination
z —	series impedance/unit length of line
Z —	impedance
$\overline{Z}$ —	normalized impedance
Z_c	characteristic admittance of line
Z_t —	admittance of termination
$Z(w)$ —	ratio of voltage to current at the location w
α —	attenuation constant
β —	phase constant
γ —	propagation constant ($\gamma = \alpha + j\beta$)
Γ —	reflection coefficient
Γ_t —	reflection coefficient of the termination
$\Gamma(w)$ —	ratio of reflected to forward voltages at the location w
δ —	loss angle of dielectric skin depth "width" of standing-wave minimum
Δ —	shift in minimum position
ϵ —	permittivity
ϵ_r —	relative permittivity
$\widetilde{\epsilon}$ —	complex permittivity
$\widetilde{\epsilon}_r$ —	complex relative permittivity
ϵ' —	real part of $\widetilde{\epsilon}$
ϵ'' —	imaginary part of $\widetilde{\epsilon}$
ϵ_r' —	real part of $\widetilde{\epsilon}_r$
ϵ_r'' —	imaginary part of $\widetilde{\epsilon}_r$, loss factor
θ —	angle of reflection coefficient
ϕ —	angle of a phasor
λ —	wavelength
μ —	permeability
σ —	conductivity
χ —	charge/unit length of line
ω —	angular frequency

Index

GENERAL RADIO
WEST CONCORD, MASSACHUSETTS 01781

•NEW YORK •BRIDGEPORT •PHILADELPHIA •WASHINGTON D.C. •SYRACUSE •ROCHESTER
•ATLANTA •COCOA BEACH •HUNTSVILLE •BOSTON •CHICAGO •DETROIT •INDIANAPOLIS
•CLEVELAND •DAYTON •DALLAS •HOUSTON •ALBUQUERQUE •LOS ANGELES •SAN DIEGO
•DENVER •SAN FRANCISCO •SEATTLE •CANADA •TORONTO •MONTREAL •OTTAWA

GENERAL RADIO COMPANY (OVERSEAS) CH 8034 Zurich 34, Switzerland
GENERAL RADIO GMBH D8 Munchen 80, West Germany
GENERAL RADIO COMPANY (U.K.) LIMITED Bourne End Buckinghamshire, England

REPRESENTATIVES IN PRINCIPAL OVERSEAS COUNTRIES