1

Introduction to Microwave Measurements

COPYRIGHTED MATERIAL

"To measure is to know."[1] This is a book about the art and science of measuring microwave components. While this work is based entirely on science, there is some art in the process, and the terms *skilled-in-the-art* and *state-of-the-art* take on particular significance when viewing the task of measuring microwave components. The goal of this work is to provide the latest, state-of-the-art methods and techniques for acquiring the optimum measurements of the myriad of microwave components. This goal naturally leads to the use of the vector network analyzer (VNA) as the principal test equipment, supported by the use of power meters, spectrum analyzers (SAs), signal sources and noise sources, impedance tuners, and other accessories.

Note here the careful use of the word *optimum*; this implies there are trade-offs between the cost and complexity of the measurement system, the time or duration of the measurement, the analytically computed uncertainty and traceability, and some heretofore unknown intangibles that all affect the overall measurement. For the best possible measurement, ignoring any consequence of time or cost, one can often go to national standards laboratories to find these best methods, but they would not suit a practical or commercial application. Thus, here the attempt is to strike an optimum balance between minimal errors in the measurement and practical consequences of the measurement techniques. The true value of this book is in providing insight into the wide range of issues and troubles that one encounters in trying to carefully and correctly ascertain the characteristics of one's microwave component. The details have been gathered from decades of experience in hundreds of direct interactions with actual measurements; some problems are obvious and common, and others are subtle and rare. It is hoped that the reader can use this handbook to avoid many hours of unproductive test time.

For the most part, the mathematical derivations in this book are intended to provide the reader with a straightforward connection between the derived values and the underlying characteristics. In some cases, the derivation will be provided in full if it is not accessible from

[1] Lord Kelvin, "On Measurement."

Handbook of Microwave Component Measurements: With Advanced VNA Techniques, Second Edition. Joel P. Dunsmore.
© 2020 John Wiley & Sons Ltd. Published 2020 by John Wiley & Sons Ltd.

existing literature; in other cases, a reference to the derivation will be provided. There are extensive tables and figures, with key sections providing many of the important formulas. The mathematical level of this handbook is geared to a college senior or working engineer with the intention of providing the most useful formulas in an approachable way. As such, sums will be preferred to integrals; finite differences will be preferred to derivatives; and divs, grads, and curls will be entirely eschewed.

The chapters are intended to self-standing for the most part. In many cases, there will be common material to many measurement types, such as the mathematical derivation of the parameters or the calibration and error-correction methods, and these will be gathered in the introductory chapters, though well referenced in the measurement chapters. In some cases, older methods of historical interest are given (there are many volumes on these older techniques), but by and large only the most modern techniques are presented. The focus here is on the practical microwave engineer facing modern, practical problems.

1.1 Modern Measurement Process

Throughout the discussion of measurements, a six-step procedure will be followed that applies to most measurement problems. When approaching a measurement, these steps are as follows:

Pretest: This important first step is often ignored, resulting in meaningless measurements and wasted time. During the pretest, measurements of the device-under-test (DUT) are performed to coarsely determine some of its attributes. During pretest, it is also determined if the DUT is plugged in, turned on, and operating as expected. Many times the gain, match, or power handling is discovered to be different than expected, and much time and effort can be saved by finding this out early.

Optimize: Once the coarse attributes of the device have been determined, the measurement parameters and measurement system can be optimized to give the best results for that particular device. This might include adding an attenuator to the measurement receivers, adding booster amplifiers to the source, or just changing the number of points in a measurement to capture the true response of the DUT. Depending upon the device's particular characteristic response relative to the system errors, different choices for calibration methods or calibration standards might be required.

Calibrate: Many users will skip to this step, only to find that something in the setup does not provide the needed conditions and they must go back to the first step, retest, and optimize before recalibration. Calibration is the process of characterizing the measurement system so that systematic errors can be removed from the measurement result. This is not the same as obtaining a calibration sticker for an instrument but really is the first step, the *acquisition* step of the error correction process that allows improved measurement results.

Measure: Finally, some stimulus is applied to the DUT, and its response to the stimulus is measured. During the measurement, many aspects of the stimulus must be considered, as well as the order of testing and other testing conditions. These include not only the specific test conditions but also pre-conditions such as previous power states to account for non-linear responses of the DUT.

Analyze: Once the raw data is taken, error correction factors (the *application* step of error correction) are applied to produce a corrected result. Further mathematical manipulations on the measurement result can be performed to create more useful figures of merit, and the data from one set of conditions can be correlated with other conditions to provide useful insight into the DUT.

Save data: The final step is saving the results in a useful form. Sometimes this can be as simple as capturing a screen dump, but often it means saving results in such a way that they can be used in follow-up simulations and analysis.

1.2 A Practical Measurement Focus

The techniques used for component measurements in the microware world change dramatically depending upon the attributes of the components; thus, the first step in describing the optimum measurement methods is understanding the expected behavior of the DUT. In describing the attributes and measurements of microwave components it is tempting to go back to first principles and derive all the underlying mathematics for each component and measurement described, but such an endeavor would require several volumes to complete. One could literally write a book on the all the attributes of almost any *single* component, so for this book the focus will be on only those final results useful for describing practical attributes of the components to be characterized, with quotes and references of many results without the underlying derivation.

There have been examples of books on microwave measurements that focus on the metrology kind of measurements (Collier and Skinner 2007) made in national laboratories such as the National Institute for Standards and Technology (NIST, USA) or the National Physical Laboratory (NPL, UK), but the methods used there don't transfer well or at all to the commercial market. For the most part, the focus of this book will be on practical measurement examples of components found in commercial and aerospace/defense industries. The measurements focus will be commercial characterization rather than the kinds of metrology found in standards labs.

Also, while there has been a great deal written about components in general or ideal terms, as well as much academic analysis of these idealized components, in practice these components contain significant parasitic effects that cause their behavior to differ dramatically from that described in many textbooks. Unfortunately, these effects are often not well understood, or difficult to consider in an analytic sense, and so are revealed only during an actual measurement of a physical devices. In this chapter, the idealized analysis of many components is described, but the descriptions are extended to some of the real-world detriments that cause these components' behavior to vary from the expected analytical response.

1.3 Definition of Microwave Parameters

In this section, many of the relevant parameters used in microwave components are derived from the fundamental measurements of voltage and current on the ports. For simplicity, the derivations will focus on measurements made under the conditions of termination in real valued impedances, with the goal of providing mathematical derivations that are straightforward to follow and readily applicable to practical cases.

In microwave measurements, the fundamental parameter of measurement is power. One of the key goals of microwave circuit design is to optimize the power transfer from one circuit to another such as from an amplifier to an antenna. In the microwave world, power is almost always referred to as either an incident power or a reflected power, in the context of power traveling along a transmission structure. The concept of traveling waves is of fundamental importance to understanding microwave measurements, and to engineers who haven't had a course on transmission lines and traveling waves, and even some who have, the concept of power flow and traveling waves can be confusing.

1.3.1 S-Parameter Primer

S-parameters have been developed in the context of microwave measurements but have a clear relationship to voltages and currents that are the common reference for most electrical engineers. This section will develop the definition of traveling waves and from that the definition of S-parameters, in a way that is both rigorous and ideally intuitive; the development will be incremental, rather than just quoting results, in hopes of engendering an intuitive understanding.

This signal traveling along a transmission line is known as a *traveling wave* (Marks and Williams 1992) and has a forward component and a reverse component. Figure 1.1 shows the schematic of a two-wire transmission structure with a source and a load.

If the voltage from the source is sinusoidal, it is represented by the phasor notation

$$v_s(t) = \mathrm{Re}(|V_s|e^{j(\omega t+\phi)}),\ or\ V_s = |V_s|e^{j(\omega t+\phi)} \tag{1.1}$$

The voltage and current at the load are

$$V_L = |V_L|e^{j\phi_L^V}, I_L = |I_L|e^{j\phi_L^I} \tag{1.2}$$

The voltage along the line is defined as $V(z)$, and the current at each point is $I(z)$. The impedance of the transmission line provides for a relationship between the voltage and the current. At the reference point, the total voltage is $V(0)$ and is equal to V_1; the total current is $I(0)$. The power delivered to the load can be described as

$$P_L = P^F - P^R \tag{1.3}$$

where P^F is called the *forward power*, and P^R is called the *reverse power*. To put this in terms of the voltage and current of Figure 1.1, the total voltage at the port can be defined as the sum

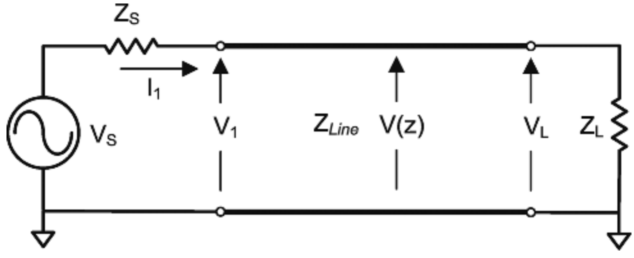

Figure 1.1 Voltage source and two-wire system.

of the forward voltage wave traveling into the port and the reverse voltage wave emerging from the port.

$$V_1 = V_F + V_R \tag{1.4}$$

The forward voltage wave represents a power traveling toward the load, or transferring from the source to the load, and the reflected voltage wave represents power traveling toward the source. To be formal, for a sinusoidal voltage source, the voltage as a function of time is

$$v_1(t) = V_1^p \cos(\omega t + \phi) = \mathrm{Re}(V_1^p\, e^{j(\omega t+\phi)}) \tag{1.5}$$

From this it is clear that V_1^p is the peak voltage and the root-mean-square (rms) voltage is

$$V_1 = \frac{V_1^p}{\sqrt{2}} \tag{1.6}$$

The $\sqrt{2}$ factor shows often in the following discussion of power in a wave, and it is sometimes a point of confusion; but if one remembers that rms voltage is what is used to compute power in a sine wave, and is used to refer to the wave amplitude of a sine wave in the following equations, then it will make perfect sense.

Considering the source impedance Z_S and the line or port impedance Z_0, and simplifying a little by making $Z_S = Z_0$ and considering the case where Z_0 is pure-real, one can relate the forward and reverse voltage to an equivalent power wave. If one looks at the reference point of Figure 1.1 and one had the possibility to insert a current probe as well as had a voltage probe, one could monitor the voltage and current.

The source voltage must equal the sum of the voltage at port 1 and the voltage drop of the current flowing through the source impedance.

$$V_S = V_1 + I_1 Z_0 \tag{1.7}$$

Defining the forward voltage as

$$V_F = \frac{1}{2}(V_1 + I_1 Z_0) \tag{1.8}$$

we see that the forward voltage represents the voltage at port 1 in the case where the termination is Z_0. From this and Eq. (1.4), one finds that the reverse voltage must be

$$V_R = \frac{1}{2}(V_1 - I_1 Z_0) \tag{1.9}$$

If the transmission line in Figure 1.1 is long (such that the load effect is not noticeable) and the line impedance at the reference point is the same as the source, which may be called the port reference impedance, then the instantaneous current going into the transmission line is

$$I_F = V_S\left(\frac{Z_0}{Z_0 + Z_S}\right) = \left.\frac{V_S}{2Z_0}\right|_{Z_0=Z_S} \tag{1.10}$$

The voltage at that point is same as the forward voltage and can be found to be

$$V_F = V_S\left(\frac{Z_0}{Z_0 + Z_S}\right) = \left.\frac{V_S}{2}\right|_{Z_0=Z_S} \tag{1.11}$$

The power delivered to the line (or a Z_0 load) is

$$P_F = V_F I_F = \left(\frac{V_F^2}{Z_0}\right) = \frac{V_S}{4Z_0} \tag{1.12}$$

From these definitions, one can now refer to the incident and reflected power waves using the normalized incident and reflected voltage waves, a and b as (Keysight Technologies 1968).

$$a = \frac{V_F}{\sqrt{Z_0}}, \quad b = \frac{V_R}{\sqrt{Z_0}} \textit{ provided } Z_0 \textit{ is real} \tag{1.13}$$

Or, more formally as a power wave definition

$$a = \frac{1}{2}\left(\frac{V_1 + I_1 Z_0}{\sqrt{|\mathrm{Re}Z_0|}}\right), \ b = \frac{1}{2}\left(\frac{V_1 - I_1 Z_0^*}{\sqrt{|\mathrm{Re}Z_0|}}\right) \tag{1.14}$$

where Eq. (1.14) includes the situation in which Z_0 is not pure real (Kurokawa 1965). However, it would be an unusual case to have a complex reference impedance in any practical measurement.

For real values of Z_0, one can define the forward or incident power as $|a|^2$ and the reverse or scattered power as $|b|^2$ and see that the values a and b are related to the forward and reverse voltage waves, but with the units of square root of power. In practice, the definition of Eq. (1.13) is typically used, as the definition of Z_0 is almost always either 50 or 75 Ω. In the case of waveguide measurements, the impedance is not well defined and changes with frequency and waveguide type. It is recommended to simply use a normalized impedance of 1 for the waveguide impedance. This does not represent 1 Ω but is used to represent the fact that measurements in a waveguide are normalized to the impedance of an ideal waveguide. In Eq. (1.13) incident and reflected waves are defined, and in practice the incident waves are the independent variables, and the reflected waves are the dependent variables. Consider Figure 1.2, a 2-port network.

There are now sets of incident and reflected waves at each port i, where

$$a_i = \frac{V_{Fi}}{\sqrt{Z_{0i}}}, b_i = \frac{V_{Ri}}{\sqrt{Z_{0i}}} \tag{1.15}$$

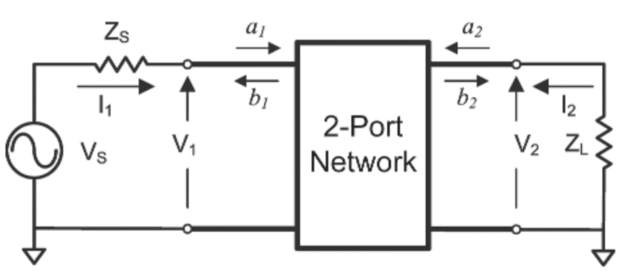

Figure 1.2 2-port network connected to a source and load.

The voltages and currents at each port can now be defined as

$$V_i = \sqrt{Z_{0i}}(a_i + b_i)$$
$$I_i = \frac{1}{\sqrt{Z_{0i}}}(a_i - b_i) \tag{1.16}$$

where Z_{0i} is the reference impedance for the ith port. An important point here that is often misunderstood is that the reference impedance does not have to be the same as the port impedance or the impedance of the network. It is a "nominal" impedance; that is, it is the impedance that we "name" when we are determining the S-parameters, but it need not be associated with any impedance in the circuit. Thus, a 50 Ω test system can easily measure and display S-parameters for a 75 Ω device, referenced to 75 Ω.

The etymology of the term *reflected* derives from optics and refers to light reflecting off a lens or other object with an index of refraction different from air, whereas it appears that the genesis for the scattering or S-matrix was derived in the study of particle physics, from the concept of wavelike particles scattering off crystals. In microwave work, scattering or S-parameters are defined to relate the independent incident waves to the dependent waves; for a 2-port network they become

$$b_1 = S_{11}a_1 + S_{12}a_2$$
$$b_2 = S_{21}a_1 + S_{22}a_2 \tag{1.17}$$

which can be placed in matrix form as

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{1.18}$$

where a's represent the incident power at each port, that is, the power flowing into the port, and b's represent the scattered power, that is, the power reflected or emanating from each port. For more than two ports, the matrix can be generalized to

$$\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} S_{11} & \cdots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1} & \cdots & S_{nn} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad or \quad [b_n] = [S] \cdot [a_n] \tag{1.19}$$

From Eq. (1.17) it is clear that it takes four parameters to relate the incident waves to the reflected waves, but Eq. (1.17) provides only two equations. As a consequence, solving for the S-parameters of a network requires that at least two sets of linearly independent conditions for a_1 and a_2 be applied, and the most common set is one where first a_2 is set to zero, the resulting b waves are measured, and then a_1 is set to zero, and finally a second set of b waves are measured. This yields

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0} \quad S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0}$$
$$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0} \quad S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0} \tag{1.20}$$

which is the most common expression of S-parameter values as a function of a and b waves, and often the only one given for their definition. However, there is nothing in the definition of S-parameters that requires one or the other incident signals to be zero, and it would be just as valid to define them in terms of two sets of incident signals, a_n and a'_n, and reflected signals, b_n and b'_n.

$$S_{11} = \left(\frac{b_1 a'_2 - a_2 b'_1}{a_1 a'_2 - a_2 a'_1}\right) \quad S_{12} = \left(\frac{b_1 a'_1 - a_1 b'_1}{a_2 a'_1 - a_1 a'_2}\right)$$
$$S_{21} = \left(\frac{b_2 a'_2 - a_2 b'_2}{a_1 a'_2 - a_2 a'_1}\right) \quad S_{22} = \left(\frac{b_2 a'_1 - a_1 b'_2}{a_2 a'_1 - a_1 a'_2}\right) \tag{1.21}$$

From Eq. (1.21) one sees that S-parameters are in general defined for a pair of stimulus drives. This will become quite important in more advanced measurements and in the actual realization of the measurement of S-parameters, because in practice it is not possible to make the incident signal go to zero because of mismatches in the measurement system.

These definitions naturally lead to the concept that S_{nn} parameters are reflection coefficients and are directly related to the DUT port input impedance and S_{mn} parameters are transmission coefficients and are directly related to the DUT gain or loss from one port to another.

Now that the S-parameters are defined, they can be related to common terms used in the industry. Consider the circuit of Figure 1.3, where the load impedance Z_L may be arbitrary and the source impedance is the reference impedance.

From inspection one can see that

$$V_1 = V_S\left(\frac{Z_L}{Z_L + Z_0}\right), \quad I_1 = V_S\left(\frac{1}{Z_L + Z_0}\right) \tag{1.22}$$

which is substituted into Eq. (1.8) and Eq. (1.9), and from (1.15) one can directly compute a_1 and b_1 as

$$a_1 = \frac{V_S}{2\sqrt{Z_0}}, \quad b_1 = \frac{V_S}{2\sqrt{Z_0}}\left(\frac{Z_L - Z_0}{Z_L + Z_0}\right) \tag{1.23}$$

From here S_{11} can be derived from inspection as

$$S_{11} = \frac{b_1}{a_1} = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{1.24}$$

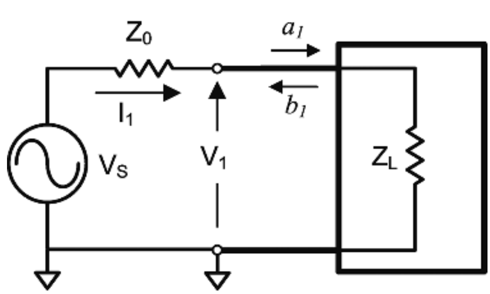

Figure 1.3 1-port network.

It is common to refer to S11 informally as the input impedance of the network, where

$$Z_{In} = \frac{V_1}{I_1} \tag{1.25}$$

This is clearly true for a 1-port network and can be extended to a 2-port or n-port network if all the ports of the network are terminated in the reference impedance; but in general, one cannot say that S_{11} is the input impedance of a network without knowing the termination impedance of the network. This is a common mistake that is made with respect to determining the input impedance or S-parameters of a network. S_{11} is defined for any terminations by Eq. (1.21), but it is the same as the input impedance of the network only under the condition that it is terminated in the reference impedance, thus satisfying the conditions for Eq. (1.20). Consider the network of Figure 1.2 where the load is not the reference impedance; as such, it is noted that a_1 and b_1 exist, but now Γ_1 (also called Γ_{In} for a 2-port network) is defined as

$$\Gamma_1 = \frac{b_1}{a_1} \tag{1.26}$$

with the network terminated in an arbitrary impedance. As such, Γ_1 represents the input impedance of a system comprised of the network and its terminating impedance. The important distinction is that S-parameters of a network are invariant to the input of output terminations, providing they are defined to a consistent reference impedance, whereas the input impedance of a network depends upon the termination impedance at each of the other ports. The value of Γ_1 of a 2-port network can be directly computed from the S-parameters and the terminating impedance, Z_L, as

$$\Gamma_1 = \left(S_{11} + \frac{S_{21}S_{12}\Gamma_L}{1 - S_{22}\Gamma_L} \right) \tag{1.27}$$

where Γ_L computed as in Eq. (1.24) is

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{1.28}$$

or in the case of a 2-port network terminated by an arbitrary load then

$$\Gamma_L = \frac{a_2}{b_2} \tag{1.29}$$

Similarly, the output impedance of a network that is sourced from an arbitrary source impedance is

$$\Gamma_2 = \left(S_{22} + \frac{S_{21}S_{12}\Gamma_S}{1 - S_{11}\Gamma_S} \right) \tag{1.30}$$

Another common term for the input impedance is the voltage standing wave ratio, called VSWR (also simply called SWR), and it represents the ratio of maximum voltage to minimum voltage that one would measure along a Z_0 transmission line terminated in some arbitrary load impedance. It can be shown that this ratio can be defined in terms of the S-parameters of the network as

$$VSWR = \left(\frac{1 + |\Gamma_1|}{1 - |\Gamma_1|} \right) \tag{1.31}$$

If the network is terminated in its reference impedance, then Γ_1 becomes S_{11}. Another common term used to represent the input impedance is the reflection coefficient, ρ_{In}, where

$$\rho_{In} = |\Gamma_{In}| \tag{1.32}$$

It's also common to write

$$VSWR = \left(\frac{1+\rho}{1-\rho}\right) \tag{1.33}$$

Another term related to the input impedance is *return loss*, which is alternatively defined as

$$RL = 20 \cdot \log_{10}(\rho),\ \ or\ RL = -20 \cdot \log_{10}(\rho) \tag{1.34}$$

with the second definition being most properly correct, as loss is defined to be positive in the case where a reflected signal is smaller than the incident signal. But, in many cases, the former definition is more commonly used; the microwave engineer must simply refer to the context of the use to determine the proper meaning of the sign. Thus, an antenna with 14 dB return loss would be understood to have a reflection coefficient of 0.2, and the value displayed on a measurement instrument might read −14 dB.

For transmission measurements, the figure of merit is often gain or insertion loss (sometimes called *isolation* when the loss is very high). Typically this is expressed in dB, and similarly to return loss, it is often referred to as a positive number. Thus

$$Gain = 20\log_{10}(|S_{21}|) \tag{1.35}$$

Insertion loss or isolation is defined as

$$Insertion\ Loss = Isolation = -20\log_{10}(|S_{21}|) \tag{1.36}$$

Again, the microwave engineer will need to use the context of the discussion to understand that a device with 40 dB isolation will show on an instrument display as −40 dB, due to the instrument using the evaluation of Eq. (1.35).

Notice that in the return loss, gain, and insertion loss equations, the dB value is given by the formula $20\log_{10}(|S_{nm}|)$, and this is often a source of confusion because common engineering use of dB has the computation as $X_{dB} = 10\log_{10}(X)$. This apparent inconsistency comes from the desire to have power gain when expressed in dB be equal to voltage gain, also expressed in dB. In a device sourced from a Z_0 source and terminated in a Z_0 load, the power gain is defined as the power delivered to the load relative to the power delivered from the source, and the gain is

$$Power\ gain = 10\log_{10}\left(\frac{P_{To\_Load}}{P_{From\_Source}}\right) \tag{1.37}$$

The power from the source is the incident power $|a_1|^2$, and the power delivered to the load is $|b_2|^2$. The S-parameter gain is S_{21} and in a matched source and load situation is simply

$$S_{21} = \frac{b_2}{a_1},\quad |S_{21}|^2 = \left|\frac{b_2}{a_1}\right|^2 = \frac{|b_2|^2}{|a_2|^2} = Power\ Gain \tag{1.38}$$

So computing power gain as in Eq. (1.37) and converting to dB yields the familiar formula

$$Power\ Gain_{dB} = 10\log_{10}(|S_{21}|^2) = 20\log_{10}(|S_{21}|) \tag{1.39}$$

A few more comments on power are appropriate, as power has several common meanings that can be confused if not used carefully. For any given source, as shown in Figure 1.1, there exists a load for which the maximum power of the source may be delivered to that load. This maximum power occurs when the impedance of the load is equal to the conjugate of the impedance of the source, and the maximum power delivered is

$$P_{\max} = \frac{|V_S|^2}{4 \cdot \mathrm{Re}(Z_S)} \tag{1.40}$$

But it is instructive to note that the maximum power as defined in Eq. (1.40) is the same as $|a_1|^2$ provided the source impedance is real and equals the reference impedance; thus, the incident power from a Z_0 source is always the maximum power that can be delivered to a load. The actual power delivered to the load can be defined in terms of a and b waves as well.

$$P_{del} = |a|^2 - |b|^2 \tag{1.41}$$

If one considers a passive two-port network and conservation of energy, power delivered to the load must be less than or equal to the power incident on the network minus the power reflected, or in terms of S-parameters

$$|S_{21}|^2 \leq 1 - |S_{11}|^2 \tag{1.42}$$

which leads the well-known formula for a lossless network

$$|S_{21}|^2 + |S_{11}|^2 = 1 \tag{1.43}$$

1.3.2 Phase Response of Networks

While most of the discussion thus far about S-parameters refers to powers, including incident, reflected, and delivered to the load, the S-parameters are truly complex numbers and contain both a magnitude and phase component. For reflection measurements, the phase component is critically important and provides insight into the input elements of the network. These will be discussed in great detail as part of Chapter 2, especially when referencing the Smith chart.

For transmission measurements, the magnitude response is often the most cited value of a system, but in many communications systems, the phase response has taken on more importance. The phase response of a network is typically given by

$$\phi_{S21} = \arctan\left[\frac{\mathrm{Im}(S_{21})}{\mathrm{Re}(S_{21})}\right] \tag{1.44}$$

where the region of the arctangent is usually chosen to be $\pm 180°$. However, it is sometimes preferable to display the phase in absolute terms, such that there are no phase discontinuities in the displayed value. This is sometimes called the *unwrapped* phase, in which the particular cycle of the arctangent must be determined from the previous cycle, starting from the DC value. Thus, the unwrapped phase is uniquely defined for an S_{21} response only when it includes all values down to zero frequency (DC).

The linearity of the phase response has consequences when looking at its effect on complex modulated signals. In particular, it is sometimes stated that linear networks cannot cause distortion, but this is true only of single-frequency sinusoidal inputs. Linear networks can cause distortion in the envelope of complex modulated signals, even if the frequency response (the magnitude of S_{21}) is flat. That is because the phase response of a network directly affects the relative time that various frequencies of a complex modulated signal take to pass through the network. Consider the signal in Figure 1.4.

For this network, the phase of S_{21} defines how much shift occurs for each frequency element in the modulated signal. Even though the amplitude response is the same in both Figure 1.4a,b, the phase response is different, and the envelope of the resulting output is changed. In general, there is some delay from the input to the output of a network, and the important definition

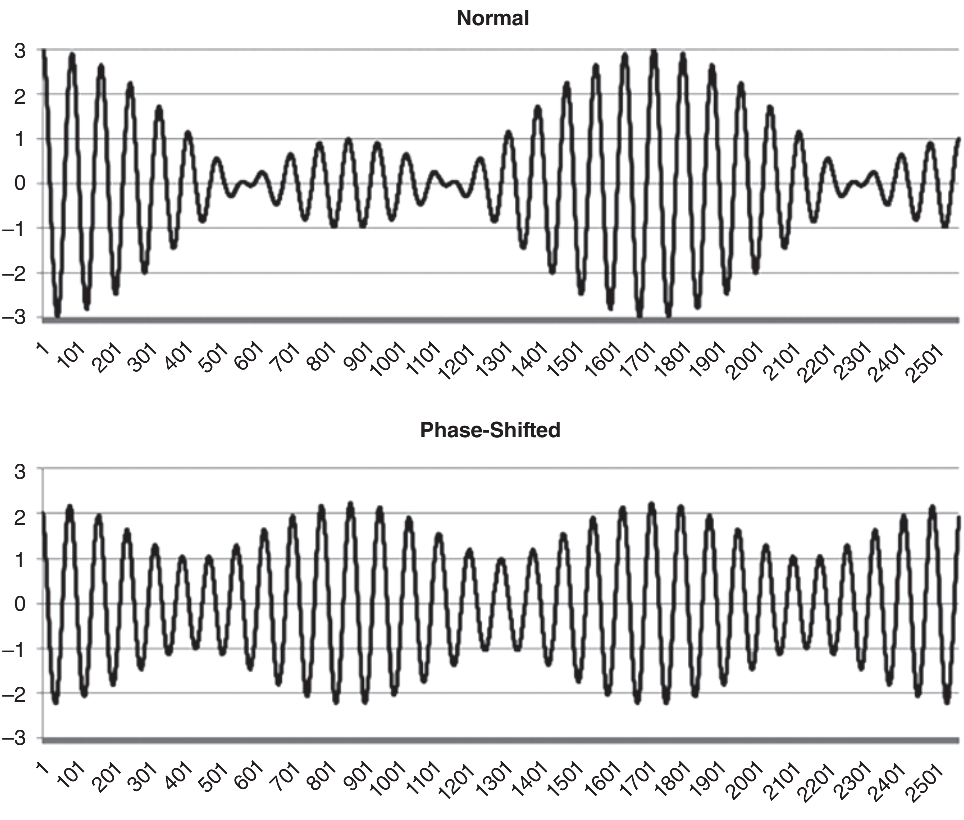

Figure 1.4 Modulated signal through a network showing distortion due to only phase shift: normal (upper), shifted (lower).

that is most commonly used is the group delay of the network, defined as

$$\tau_{GD} = -\frac{d\phi_{S21}^{rad}}{d\omega} = \frac{-d\phi_{S21}^{\circ}}{360 \cdot df} \tag{1.45}$$

While easily defined, the group delay response may be difficult to measure and/or interpret. This is because measurement instruments record discrete values for phase, and the group delay is a derivative of the phase response. Using discrete differentiation can generate numerical difficulties; Chapter 5 shows some of the difficulties encountered in practice when measuring group delay, as well as some solutions to these difficulties.

For most complex signals, the ideal goal for phase response of a network is that of a linear phase response. Deviation from linear phase is a figure of merit for the phase flatness of a network, and this is closely related another figure of merit, group delay flatness. Thus, the ideal network has a flat group delay, meaning a linear phase response. However, many complex communications systems employ equalization to remove some of the phase response effects. Often, this equalization can account for first- or second-order deviations in the phase; thus, another figure of merit is deviation from parabolic phase, which is effectively a measure of the quality of fit of the phase response to a second order polynomial. These measurements are discussed further in Chapter 5.

1.4 Power Parameters

1.4.1 Incident and Reflected Power

Just as there are a variety of S-parameters, which are derived from the fundamental parameters of incident and reflected waves a and b, so too are there many power parameters that can be identified with the same waves. As inferred earlier, the principal power parameters are incident and reflected, or forward and reverse, powers at each port, which for Z_0 real, are defined as

$$P_{Incident} = P_F = |a|^2, \quad P_{Reflected} = P_R = |b|^2 \tag{1.46}$$

The proper interpretation of these parameters is that incident and reflected power is the power that would be delivered to a nonreflecting (Z_0) load. If one were to put an ideal Z_0 directional-coupler in line with the signal, it would sample or couple the incident signal (if the coupler were set to couple the forward power) or the reflected signal (if the coupler were set to couple the reverse power). In simulations, ideal directional-couplers are often used in just such a manner.

1.4.2 Available Power

The maximum power that can delivered from a generator is called the *available power*, or $P_{Available}$, and can be defined as the power delivered from a Z_S

$$P_{Available} = P_{AS} = \frac{|a_S|^2}{(1 - |\Gamma_S|^2)} \tag{1.47}$$

where Γ_S is computed as in Eq. (1.24) as

$$\Gamma_S = \frac{Z_S - Z_0}{Z_S + Z_0} \tag{1.48}$$

This maximum power is delivered to the load when the load impedance is the conjugate of the source impedance, $Z_L = Z_S^*$.

1.4.3 Delivered Power

The power that is absorbed by an arbitrary load is called the *delivered power*, and it is computed directly from the difference between the incident and reflected power.

$$P_{del} = |a|^2 - |b|^2 \tag{1.49}$$

For most cases, this is the power parameter that is of greatest interest. In the case of a transmitter, it represents the power that is delivered to the antenna, for example, which in turn is the power radiated less the resistive loss of the antenna.

1.4.4 Power Available from a Network

A special case of available power is the power available from the output of a network, when the network is connected an arbitrary source. In this case, the available power is only a function of the network and the source impedance and is not a function of the load impedance. It represents the maximum power that could be delivered to a load under the condition that the load impedance was ideally matched and can be found by noting that the available output power is similar to Eq. (1.47) but with the source reflection coefficient replaced by the output reflection coefficient of the network Γ_2 from Eq. (1.30) such that

$$P_{Out\_Available} = P_{OA} = \frac{|b_2|^2}{(1 - |\Gamma_2|^2)} \tag{1.50}$$

When a 2-port network is connected to a generator with arbitrary impedance, the output scattered wave into matched load is

$$b_2 = \frac{a_S S_{21}}{1 - \Gamma_S S_{11}} \tag{1.51}$$

Here the incident wave is represented as a_S rather than a_1 as an indication that the source is not matched, and Γ_1 is defined by Eq. (1.27). The output power incident to the load is

$$|b_2|^2 = \frac{|a_S|^2 |S_{21}|^2}{|1 - \Gamma_S S_{11}|^2} \tag{1.52}$$

Combining Eqs. (1.52) and (1.50), the available power at the output from a network that is driven from a generator with source impedance of Γ_S is

$$P_{OA} = \frac{|b_2|^2}{(1 - |\Gamma_2|^2)} = \frac{|a_S|^2 |S_{21}|^2}{|1 - \Gamma_S S_{11}|^2 (1 - |\Gamma_2|^2)} \tag{1.53}$$

With Γ_2 defined as in Eq. (1.30).

1.4.5 Available Gain

Available gain is the gain that an amplifier can provide to a conjugately matched load from a source or generator of a given impedance and is computed with the formula

$$G_A = \frac{(1 - |\Gamma_S|^2)|S_{21}|^2}{|1 - \Gamma_S S_{11}|^2(1 - |\Gamma_2|^2)} \tag{1.54}$$

$$\text{where } \Gamma_2 = \left(S_{22} + \frac{S_{21}S_{12}\Gamma_S}{1 - S_{11}\Gamma_S}\right)$$

Other derived values such as maximum available gain and maximum stable gain are discussed in detail in Chapter 6.

1.5 Noise Figure and Noise Parameters

For a receiver, the key figure of merit is its sensitivity, or the ability to detect small signals. This is limited by the intrinsic noise of the device itself, and for amplifiers and mixers, this is represented as noise figure. Noise figure is defined as a signal-to-noise ratio at the input divided by signal to noise at the output expressed in dB.

$$NF \equiv N_{Figure} = 10\log_{10}\left(\frac{Signal_{Input}/Noise_{Input}}{Signal_{Output}/Noise_{Output}}\right) = 10\log_{10}\left(\frac{(S/N)_I}{(S/N)_O}\right) \tag{1.55}$$

Its related value, noise factor, which is unitless, is

$$N_F \equiv N_{Factor} = \left(\frac{Signal_{Input}/Noise_{Input}}{Signal_{Output}/Noise_{Output}}\right) = \frac{(S/N)_I}{(S/N)_O} \tag{1.56}$$

Here the signal and noise values are represented as a power; traditionally, this is available power, but incident power can be used as well with a little care. Rearranging Eq. (1.55), one can obtain

$$N_{Factor} = \frac{N_O}{Gain \cdot N_I} = \frac{N_{O\_Avail}}{G_{Avail} \cdot N_{I\_Avail}} \tag{1.57}$$

In most cases, the input noise is known very well, as it consists only of thermal noise associated with the temperature of the source resistance. This is the noise available from the source and can be found from

$$N_{Avail} = N_a = kTB \tag{1.58}$$

where k is Boltzmann's constant (1.38×10^{-23} J K^{-1}), B is the noise bandwidth, and T is the temperature in Kelvin. Note that the available noise power does not depend upon the impedance of the source. From the definition in Eq. (1.57) it is clear that if the temperature of the source impedance changes, then the noise figure of the amplifier using this definition would change as well. Therefore, by convention, a fixed value for the temperature is presumed, and this value, known as T_0, is 290 K.

This is the noise power that would be delivered to a conjugately matched load. Alternatively, the noise power can be represented as a noise wave, much like a signal, and one can define an incident noise (sometimes called the *effective noise power*), which is defined as the noise delivered to a nonreflecting nonradiating load and is found as

$$N_{Incident} = N_E = N_A(1 - |\Gamma_S|^2) \tag{1.59}$$

which is consistent with the definition of Eq. (1.47). Since the available noise at the output of a network doesn't depend upon the load impedance, the available gain from a network similarly doesn't depend upon the load impedance, and the available noise at the input of the network can be computed as Eq. (1.58), the measurement of noise figure defined in this way is not dependent upon the match of the noise receiver. One way to understand this is to note that the available gain is the maximum gain that can be delivered to a load. If the load is not conjugately matched to Γ_2, both the available gain and the available noise power at the output would be reduced by equal amounts, leaving the noise figure unchanged and independent of the noise receiver load impedance. Thus, for the case of noise measurements, the available noise power and available gain have been the important terms of use historically.

Recently more advanced techniques have been developed and made practical based on incident noise power and gain. If the impedance is known, the incident noise power can be computed as in Eq. (1.59); and if the output incident noise power N_{OE} can be measured, then one can compute the output available noise as

$$N_{OA} = \frac{N_{OE}}{(1 - |\Gamma_2|^2)} \tag{1.60}$$

Substituting into Eq. (1.57) to find

$$N_F = \frac{1}{G_A} N_{OA} \frac{1}{N_{IA}} = \frac{|1 - \Gamma_S S_{11}|^2 \cancel{(1 - |\Gamma_2|^2)}}{(1 - |\Gamma_S|^2)|S_{21}|^2} \frac{N_{OE}}{\cancel{(1 - |\Gamma_2|^2)}} \frac{1}{(kTB)} = \frac{|1 - \Gamma_S S_{11}|^2 N_{OE}}{(1 - |\Gamma_S|^2)|S_{21}|^2 (kTB)} \tag{1.61}$$

When the source is a matched source, this simplifies to

$$N_F = \frac{N_{OE}}{|S_{21}|^2 (kTB)} \tag{1.62}$$

Thus, for a simple system of an amplifier sourced with a Z_0 impedance and terminated with a Z_0 load, the noise factor can be computed simply from the noise power measured in the load and the S21 gain. However, Eq. (1.61) defines the noise figure of the amplifier in terms of the source impedance, and this is a key point. In general, although the 50 Ω noise figure is the most commonly quoted, it is measured only when the source impedance provided is exactly 50 Ω. In the case where the source impedance is not 50 Ω, the 50-Ω noise figure cannot be simply determined.

1.5.1 Noise Temperature

Because of the common factor of temperature in many noise figure computations, the noise power is sometimes redefined as available noise temperature.

$$T_A = \frac{N_A}{kB} \tag{1.63}$$

From this definition, the noise factor becomes

$$N_F = \frac{T_A}{G_A 290} = \frac{T_{RNA}}{G_A} \tag{1.64}$$

where T_{RNA} is the relative available noise temperature, expressed in Kelvin above 290 K.

1.5.2 Effective or Excess Input Noise Temperature

For very low noise figure devices, it is often convenient to express their noise factor or noise figure in terms of the excess power that would be at the input due to a higher temperature generator termination, which would result in the same available noise temperature at the output. This can be computed as

$$T_e = 290(N_F - 1) \tag{1.65}$$

Thus, an ideal noiseless network would have a zero input noise temperature, and a 3 dB noise figure amplifier would have a 290 K excess input noise temperature, or 290 K above the reference temperature.

1.5.3 Excess Noise Power and Operating Temperature

For an amplifier under test, the noise power at the output, relative to the kTB noise power, is called the *excess noise power*, P_{NE}, and is computed as

$$P_{NE} = N_F |S_{21}|^2 \frac{(1 - |\Gamma_S|^2)}{|1 - \Gamma_S S_{11}|^2 (1 - |\Gamma_2|^2)} \tag{1.66}$$

For a matched source and load, it is the excess noise, above kTB, that is measured in the terminating resistor and can be computed as

$$P_{NE} = (|S_{21}|^2 N_F) \tag{1.67}$$

which is sometimes called the *incident relative noise* or $\mathrm{RNP_I}$ (as opposed to available, or RNP). Errors in noise figure measurement are often the result of not accounting properly for the fact that the source or load impedances are not exactly Z_0. A related parameter is the operating temperature, which is analogous to the input noise temperature at the amplifier output, and is computed as

$$T_O = \frac{T_{OA}}{(1 - |\Gamma|^2)} \tag{1.68}$$

While the effect of load impedance may be overcome with the use of available gain, which is independent of load impedance, the effect of source impedance mismatch must be dealt with a much more complicated way, as shown next.

1.5.4 Noise Power Density

The excess noise is measured relative to the kTB noise floor and is expressed in dBc relative to the T_0 noise floor. However, the noise power could also be expressed in absolute terms such as dBm. But the measured noise power depends upon the bandwidth of the detector, and so the noise power density provides a reference value with a bandwidth equivalent to 1 Hz. Thus, the noise power density is the related to the excess noise by

$$P_{Noise\ Power\ Density} = \frac{P_{NE}}{B} = k(T_0 + T_e) \tag{1.69}$$

1.5.5 Noise Parameters

The formal definition of noise figure for an amplifier defines the noise figure only for the impedance or reflection coefficient of current source termination, but this noise figure is *not* the 50-Ω noise figure. Rather, it is the noise figure of the amplifier for the impedance of the source. In general, one cannot compute the 50 Ω noise figure from this value without additional information about the amplifier. If one considers the amplifier in Figure 1.5, with internal noise sources, the effect of the noise sources is to produce noise power waves that may be treated similarly to normalized power waves, *a* and *b*.

The source termination produces an incident noise wave a_{NS} and adds to the internal noise created in the amplifier, which can be represented as an input noise source a_{Namp}. There are scattered noise waves represented by the noise emitted from the input of the amplifier, b_{N1}, and the noise incident on the load is b_{N2}. From this figure, one can make a direct comparison to the S-parameters and see that reflected noise power might add or subtract to the incident noise power and affect the total noise power. However, at the input of the amplifier, the noise generated inside the amplifier is in general not correlated with the noise coming from the source termination so that they don't add together in a simple way. Because of this, the noise power at the output of the amplifier, and therefore the noise figure, depends upon the source impedance in a complex way. This complex interaction is defined by two real valued parameters and one complex parameter, known collectively as the *noise parameters*. The noise figure at any source reflection coefficient may be computed as

$$N_F = N_{Fmin} + \frac{4R_n}{Z_0}\frac{|\Gamma_{opt} - \Gamma_S|^2}{|1 + \Gamma_{opt}|^2(1 - |\Gamma_S|^2)} \tag{1.70}$$

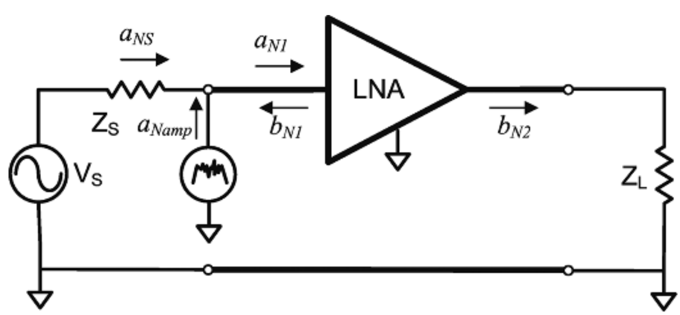

Figure 1.5 An amplifier with internal noise sources.

where N_{Fmin} is the minimum noise figure; Γ_{Opt}, called *gamma-opt*, is the reflection coefficient (magnitude and phase) that gives the minimum noise figure; and R_N, sometimes called the *noise resistance*, describes how the noise figure increases as the source impedances varies from the gamma-opt. The characterization required to determine these values is quite complex and is covered in Chapter 6.

1.6 Distortion Parameters

Up to now, all the parameters described have been under the consideration that the DUT is linear. However, when a DUT, particularly an amplifier, is driven with a large signal, non-linear transfer characteristics become significant, leading to an entirely new set of parameters used to describe these non-linear characteristics.

1.6.1 Harmonics

One of the first noticeable effects of large signal drive is the generation of harmonics at multiples of the input frequency. Harmonics are described by their order and either by their output power or, more commonly, by the power relative to the output power of the fundamental, and almost always in dBc (dB relative to the carrier). Second harmonic is short for second-order harmonic and refers to the harmonic found at two times the fundamental, even though it is in fact the first of the harmonic frequency above the fundamental; third harmonic is found at three times the fundamental, and so on. Surprisingly, there are not well-established symbols for harmonics; for this book, we will use H2, H3 … Hn to represent the dBc values of harmonics or order 2, 3 … respectively. In Chapter 6, the measurements of harmonics are fully developed as part of the description of X-parameters and utilize the notation $b_{2,m}$ to describe the output normalized wave power at port 2 for the *m*th harmonic. A similar notation is used for harmonics incident on the amplifier.

One important attribute of harmonics is that for most devices the level of the harmonics increases in dB value as the power of the input increases and to a rate directly proportional to the harmonic order, as shown in Figure 1.6. In this figure, the x-axis is the drive power, and the y-axis is the measured output power of the fundamental and the harmonics.

1.6.2 Second-Order Intercept

This pattern of increasing power as the input power is increased, but to the slope related to the order of the harmonic, cannot continue indefinitely or the harmonic power would exceed the fundamental power. While theoretically possible, in practice the harmonic power saturates just as the output power does and never crosses the level of the output power. However, if one uses the lower power regions to project a line from the fundamental and each of the harmonics, they will intersect at some power, as shown in Figure 1.6. The level that these lines converge is called the *intercept point*, and the most common value is the second-order intercept (SOI), and intercept points beyond third order are seldom used.

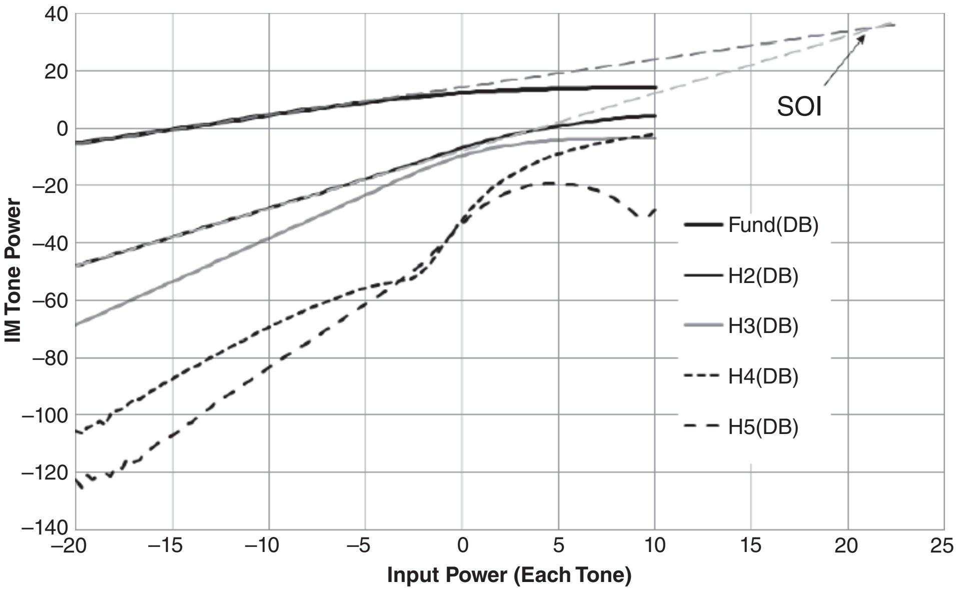

Figure 1.6 Output power of harmonics of an amplifier.

There is sometimes confusion in the use of the term SOI; while it is most commonly used to refer to the second harmonic content, in some case, it has also been used to refer to the two-tone SOI, which is a distortion product that occurs at the sum of the two tones. Most properly, one should always use the term two-tone SOI if one is to distinguish from the more common harmonic SOI.

1.6.3 Two-Tone Intermodulation Distortion

While the harmonic measurement provides a direct characterization of distortion, it suffers from the fact that the harmonic frequencies are far away from the fundamental, and in many circuits, the network response is such that the harmonic content is essentially filtered out. Thus, it is not possible to discern the non-linear response of such a network by measuring only the output signal. Of course, if the gain is measured, compression of the amplifier will show that the value of S21 changes with the input drive level. But it is convenient to have a measure or figure of merit of the distortion of an amplifier that relies only on the output signal. In such a case, two signals of different frequencies can be applied at the amplifier input, at a level sufficiently large to cause a detectible non-linear response of the amplifier. Figure 1.7 shows a measurement of a two-tone signal applied to the input of an amplifier (lower trace) and measured on the output of the amplifier (upper trace).

It is clear that several other tones are present at the output and are the result of higher-order products mixing in the amplifier due to its non-linear response and creating other signals. The principal signals of interest are the higher and lower intermodulation (IM) products, PwrN_Hi and PwrN_Lo, where N is the order of intermodulation distortion (IMD). Normally,

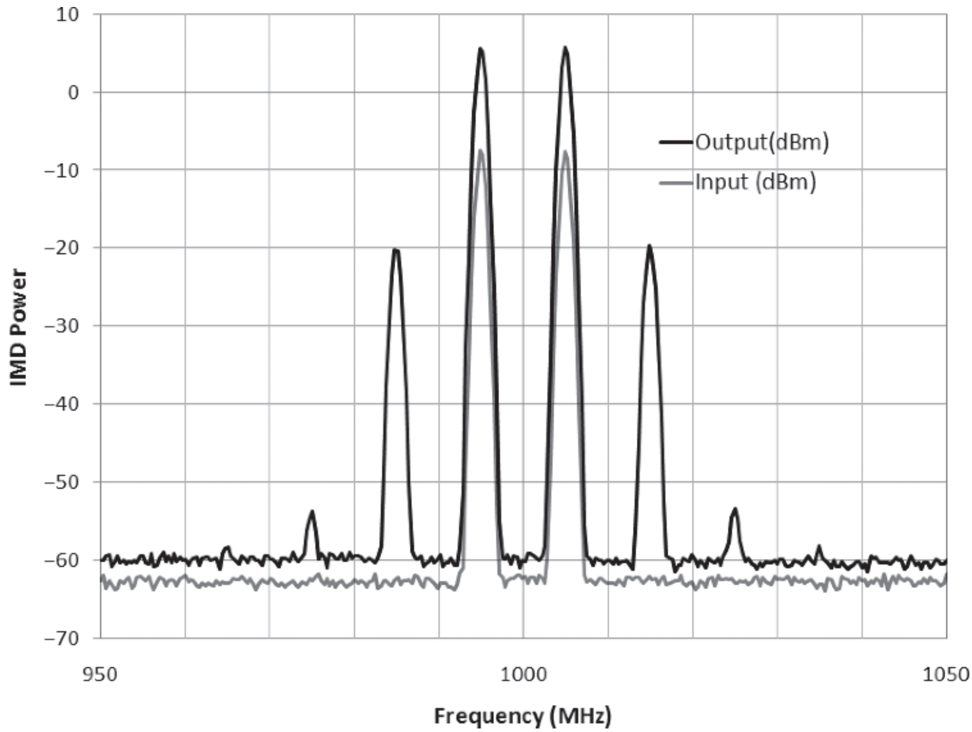

Figure 1.7 Measurement of a two-tone signal at the input and output of an amplifier.

IM products refer to the power of the IM product relative to the carrier, in dBc, and these terms are called IMN_Hi and IMN_Lo. For example, the power in the lower third-order tone is Pwr3_Lo; the level of the upper third-order tone relative to the carrier is called IM3_Hi. The frequencies of the higher and lower tones are found at

$$f_{3Hi} = 2f_{Hi} - f_{Lo},\ f_{3Lo} = 2f_{Lo} - f_{Hi} \tag{1.71}$$

And more generally

$$f_{mHi} = \left(\frac{m+1}{2}\right)f_{Hi} - \left(\frac{m-1}{2}\right)f_{Lo},\ f_{mLo} = \left(\frac{m+1}{2}\right)f_{Lo} - \left(\frac{m-1}{2}\right)f_{Hi}\Bigg|_{m\ odd}$$
$$f_{mHi} = (m-1)f_{Hi} + (m-1)f_{Lo},\ f_{mLo} = (m-1)f_{Ho} - (m-1)f_{Lo}\big|_{m\ even} \tag{1.72}$$

In Figure 1.7, the amplifier is driven such that the fifth-order IM product is just visible above the noise floor in the upper trace.

IM products have the same attribute as harmonics with respect to drive power, and the power in the IM product (sometimes called the *tone power*, or PWR*m* for the *m*th-order IM power) increases in direct proportion to the input power and the order of the IM product. Thus, if the tone power is plotted along with the output power against an x-axis of input power, the plot will look like Figure 1.8, where the extension of the slope of the output power

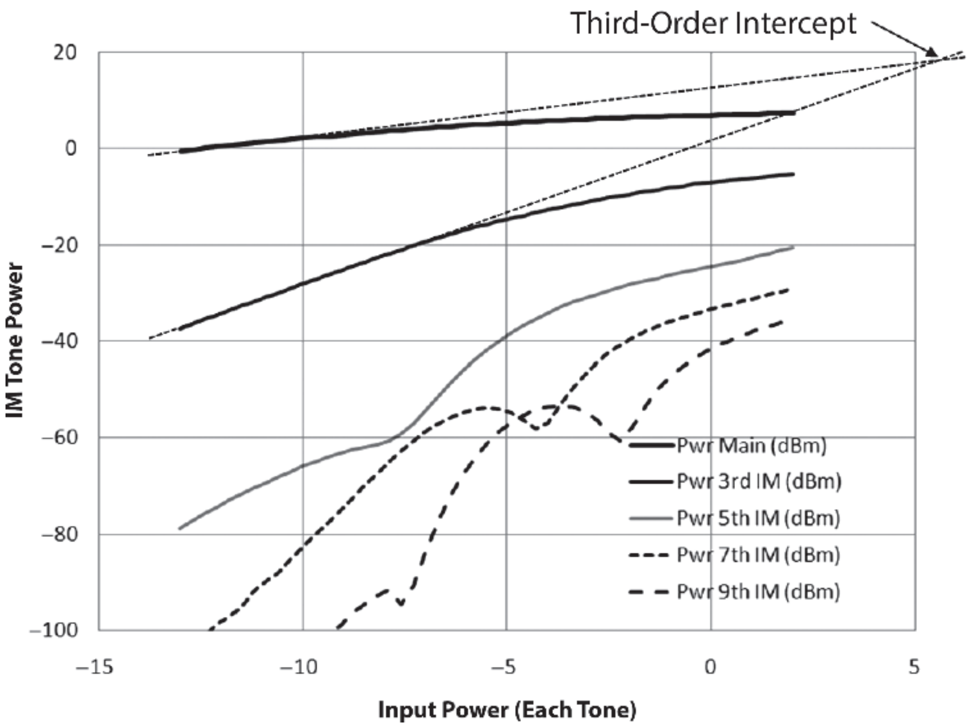

Figure 1.8 Output power and IM tone-power versus input power.

and IM tone-powers at low drives will intersect. This point of intersection for the third-order IM product is known as the third-order intercept point, or IP3. Similarly, IP5 is the fifth-order intercept point, etc.

It is also interesting to note that in general at high powers, the IM tone-powers may not increase but may decrease or have local minima. This is because of the effect of high-order IM products re-mixing and creating significant signals that lie on the lower-order products and can increase or decrease their level, depending upon the phasing of the signals.

There is often some confusion about third-order IM products (IM3) and third-order intercept point (IP3), and both are sometimes referred to as *third-order intermod*. For clarity, in this book, the intercept point will always be referred to as IP.

Finally, for amplifiers used as a low-noise amplifier (LNA) at the input of a receiver chain, it is often desired to refer the IP level to the input power, which would produce an intercept point at the output. This is distinguished as the input intercept point (IIP), and in the case of ambiguity, the normal intercept point referencing to the output power should be most properly referred to as the output-referred intercept point (OIP). The most common intercept points are the third-order ones, OIP3 and IIP3. The input and output intercept points differ by the gain of the amplifier at drive level where the measurements are made.

The details of two-tone IM measurements are discussed at length in Chapter 8.

1.6.4 Adjacent Channel Power and Adjacent Channel Level Ratio

One figure of distortion common with modulated signals is the adjacent channel power (ACP) and adjacent channel level ratio (ACLR). Sometimes a third term, adjacent channel power ratio (ACPR), is used instead of ACLR. All are measures of out-of-channel spectral regrowth caused principally by the third-order intermodulation distortion occurring because of a modulated signal. During testing, a modulated signal waveform is applied to the DUT. Figure 1.9 shows the output spectrum of a signal modulated with 16 quadrature amplitude modulation (16 QAM) over a 40 MHz BW, applied to an amplifier.

It is a repetitive periodic waveform from an arbitrary waveform generator, which must be comprised of a multiple sinewave signals, typically thousands of tones, each of which can intermodulate with each other one. In a typical modulated signal, each tone can have a nearly random amplitude and phase, so it is quite complicated to measure each distortion product directly. In general, this figure of merit measures the intermodulation products, which appear in the adjacent channel to the channel under test, as a total integrated power using band power measurements.

In the figure, the lower and upper ACP region is identified, and the signal here is caused by the third-order distortion in the amplifiers. Also identified is the outline of the distortion profile of the amplifier. ACP is used as a figure of merit as it is easy to discern the distortion level

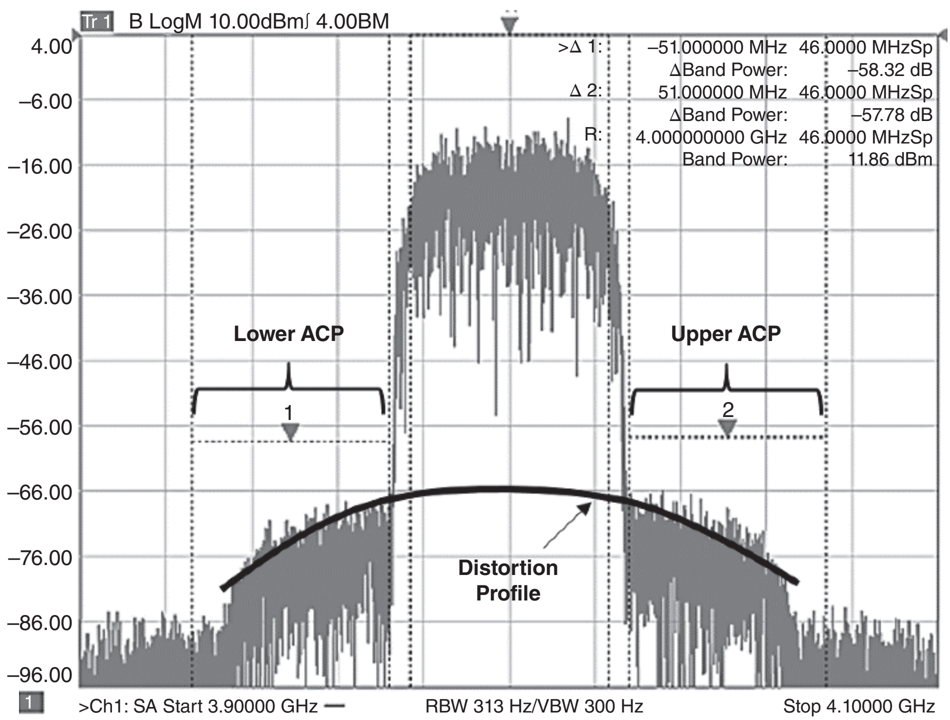

Figure 1.9 Spectral regrowth causing ACP in a 16 QAM signal.

in the adjacent channel where there is no signal. However, the distortion occurs in-channel as well as out-of-channel and is usually a bit higher in the center of the channel. The sloped response of the distortion profile is typical and can be understood by considering the density of signals that can create the intermodulation distortion. Note the outer edges of the adjacent channel where the distortion signal is lowest; only the outermost tones of the main signal can intermodulate to create a signal at these outer reaches of the adjacent channel. At the edge of the adjacent channel nearest the main signal, any two signals that are separated by one-half the main-signal bandwidth can intermodulate to create a signal here. The density of these signals is quite high, roughly half the power of the main signal. In the center of the main signal, where the distortion is not apparent because it is masked by the main-signal power, it is outlined by the distortion profile curve in the figure; any two closely spaced signals can cause distortion power here. The density of such signals is over the whole bandwidth, so the distortion level here is roughly twice that at the close-in edge of the ACP signal. Even though this distortion is masked by the main signal, it is still present and causes errors in the transmitted signal.

The total integrated power is the ACP. The ratio of the ACP to the total power in the main channel is the ACLR, shown by the Markers 1 and 2 in the figure (they are set to be a delta-marker with respect to the reference Marker R, which shows the main tone absolute power). Often, test system noise can mask the ACP or ACLR to some extent and becomes the limitation of the measurement. Details of the ACP and ACLR measurements are found in Chapter 8.

1.6.5 Noise Power Ratio (NPR)

Widely found in the satellite communications industry, noise power ratio (NPR) is a measure of distortion, and not of noise at all. In the early days of satellite development, the industry needed a measure of distortion for satellite components but could not use the more common IMD or ACP. Most satellite systems have strongly channelized amplifiers, where the communication signals fill an entire channel and are filtered at the output so adjacent channel distortion would be filtered away, and could not be used as a figure of merit for the in-channel distortion. Furthermore, the communications protocols for satellites could change over the life of the satellite, and often many different communication methods could be used in the same channel. NPR was developed to emulate a densely loaded communications channel but still provide a means to determine distortion.

In the early days, NPR signals were generated by using a noise diode followed by a filtered amplifier. This would produce a noise signal at high power, of the specified channel. This was followed by a narrow band-stop filter, which blocked the noise signal in the middle of the channel. When this signal was applied to the system component, distortion of the amplifier could be seen in the notch of the NPR signal. Figure 1.10 shows an example NPR signal, after passing through an amplifier. This is not one created by a noise diode, but rather using an arbitrary waveform generator, which is programmed to produce an additive-white-gaussian-noise (AWGN) signal with a notch at its center. In fact, the use of noise diodes to produce NPR signals has been essentially replaced throughout the industry with arbitrary waveform-generated signals. In this example, the AWGN signal is created in a low-frequency baseband generator

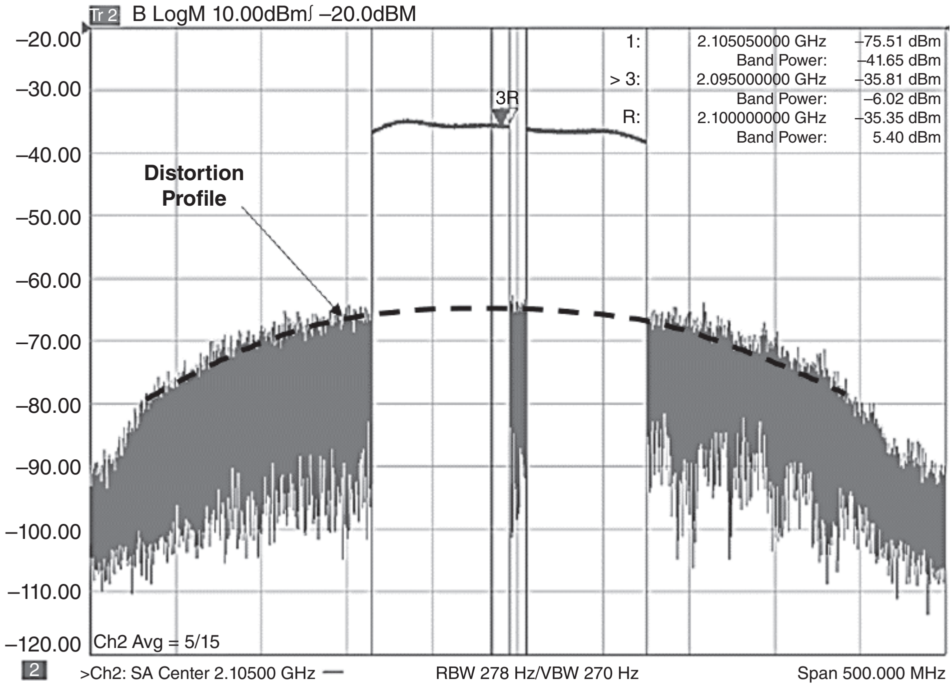

Figure 1.10 An NPR signal showing the total power and ratios of band power.

and then upconverted inside the signal source to the desired center frequency. Some unflatness is apparent in the passband of the signal due to frequency response of the signal source.

Apparent in the figure is also the ACLR level, which is nearly the same at the edge of the main signal as the NPR signal in the middle. It is clear from this figure that ACP and NPR are closely related. Imagine, though, if the DUT is followed by a sharp channelizing filter; the ACLR would be removed by the filter and could not be used to determine the distortion but the NPR signal allows one to see the in-channel distortion. NPR measurements are covered extensively in Chapter 8.

1.6.6 Error Vector Magnitude (EVM)

Error vector magnitude (EVM) is a figure of merit used in communications systems to describe the quality of a modulated signal compared to an idealized signal. In most cases, it is a measure in the so-called IQ plane of the vector difference between the measured signal and the idealized signal, which is determined by recovering the modulation pattern from the measured signal and re-creating the idealized signal. It is used when the errors are small and becomes inaccurate with large errors as the recovered signal may not be the correct signal when the EVM is quite large.

The sample point for determining the error is determined by the signaling method, and the idealized signal is time-shifted to line up with the measured signal to find the difference at the sample point. In some signaling methods, the EVM is determined by taking the fast Fourier transform (FFT) of the modulated signal and idealized signal and measuring the vector difference in the frequency domain.

EVM is affected primarily by distortion of the channel (usually in the transmitter amplifier), nonuniform frequency response (ripples or roll-off in the channel components), and noise in the system. For a transmitter component, which is the principal contributor to EVM, the noise contribution is generally not significant. In many modulation schemes, such as orthogonal frequency domain multiplexing (OFTM), the signal is broken into many narrow channels, such that the frequency response changes are small over each channel, and thus frequency flatness errors don't contribute to the EVM in these modulation schemes. In other cases, the measurement receiver has the ability to apply frequency response compensation, a kind-of inverse filtering, to remove the effects of the nonideal frequency response. This is sometimes called *equalization*, and the EVM measurement is called equa*lized EVM*. After equalization, the frequency response does not contribute significant errors to the EVM signal.

This leaves only distortion as the predominant contribution to EVM, and as such EVM has become a common figure of merit for distortion in these systems. EVM measurements generally require a full demodulation to evaluate the signal quality, and at this time such capabilities are not generally available in VNAs, but this is likely to change as EVM becomes a significant figure of merit in more systems.

Recently, several papers have been presented (Sombrin 2011; Freiberger et al. 2017) that demonstrate a corresponding relationship between EVM and NPR. These works are compelling and lead one to infer that with further development, the time is near when EVM can be determined without the need for full demodulation, as illustrated in Chapter 8.

1.7 Characteristics of Microwave Components

Microwave components differ from other electrical devices in a few respects. The principal discerning attribute is the fact that the components' size cannot be ignored. In fact, the size of many components is a significant portion of a wavelength at the frequency of interest. This size causes the phase of the signals incident on the device to vary across the device, implying that microwave devices must be treated as distributed devices. A second, related attribute is that the reference ground for the device is not defined by a point but is distributed as well. Indeed, in many cases the ground is not well defined. In some situations, grounds for a device are isolated by sufficient distance that signal propagation can occur from one device ground to another. Further, even if devices are defined as series only (with no ground contact), one must realize that there is always an earth ground available so there can always be some impedance to this ground. In practice, the earth ground is actually the chassis or package of the device, or a power or other ground plane on a printed circuit board (PCB).

Finally, only in microwave components can one find the concept of wave propagation. In waveguide components, there is no "signal" and no "ground." Rather, a wave of electric-magnetic (EM) field is guided into and out of the device without regard to a specific ground plane. For these devices, even the transmission structures, waveguide for example, are a large percentage of a signal wavelength. Common concepts such as impedance become ambiguous in the realm of waveguide measurements and must be treated with special care.

1.8 Passive Microwave Components

1.8.1 Cables, Connectors, and Transmission Lines

1.8.1.1 Cables

The simplest and most ubiquitous microwave components are transmission lines. These can be found in a variety of forms and applications, and they provide the essential glue that connects the components of a microwave system. RF and microwave cables are often the first exposure an engineer has to microwave components and transmission systems, the most widespread example being a coaxial cable used for cable television (CATV, aka Community Antenna TeleVison).

The key characteristics of coaxial cables are their impedance and loss. The characteristics of coaxial cables are often defined in terms of their equivalent distributed parameters (Magnusson 2001), as shown in Figure 1.11, described by the *telegraphers' equation*

$$\frac{dv(z)}{dz} = -(r + j\omega\, l) \cdot i(z) \tag{1.73}$$

$$\frac{di(z)}{dz} = -(g + j\omega\, c) \cdot v(z) \tag{1.74}$$

where $v(z)$, $i(z)$ are the voltage and current along the transmission line, and r, l, g, c are the resistance, inductance, conductance, and capacitance per unit length.

For a lossless cable, the impedance can be computed as simply

$$Z = \sqrt{\frac{l}{c}} \tag{1.75}$$

but it becomes more complicated when loss is introduced, becoming

$$Z_{lossy} = \sqrt{\frac{r + j\omega\, l}{g + j\omega\, c}} \tag{1.76}$$

In many applications, the conductance of the cable is negligible, particularly at low frequencies, so that the only loss element is the resistance per unit length, yielding

$$Z_{lossy} = \sqrt{\frac{r + j\omega\, l}{j\omega\, c}} \tag{1.77}$$

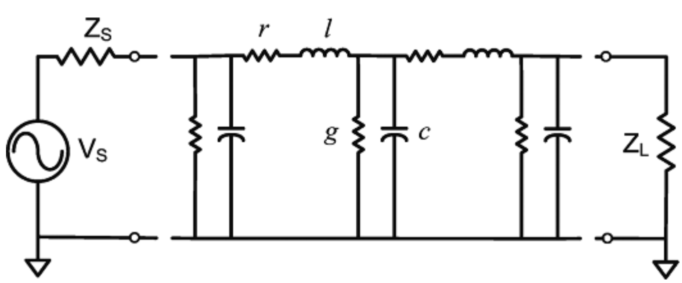

Figure 1.11 A transmission line modeled as distributed elements.

Inspection of Eq. (1.77) shows that the impedance of a cable must increase as the frequency goes down toward DC. Figure 1.12 demonstrates this with a calculation the impedance of a nominal 75 Ω cable, with a 0.0001 Ω mm^{-1} loss and capacitance of 0.07 pF mm^{-1} (typical for RG 6 CATV coax). In this case, the impedance deviates from the expected value at 300 kHz by over 10 Ω; and by 1 Ω at 1 MHz.

This low-frequency response of impedance for any real transmission line is often unexpected by those unfamiliar with Eq. (1.77), and it is sometimes assumed that this is a result of measurement error. However, all real transmission lines must show such a low frequency characteristic, and verification methods must take into account this effect.

An “airline” coax consists of a cable with an air dielectric, sometimes supported by dielectric beads at either end or sometimes supported only by the center conductor of the adjacent connectors, as shown in Figure 1.13. This type of cable has virtually no conductance, so series

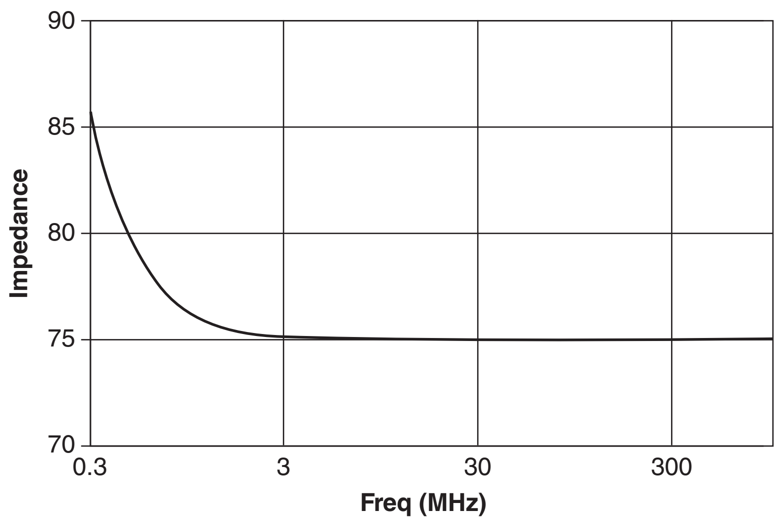

Figure 1.12 Impedance of a real transmission line at low frequency.

Figure 1.13 An airline coaxial transmission line.

resistive loss is the only loss element. The small white ring on the airline sometimes used to prevent sagging at the male end of the pin so that it may be more easily mated.

In some special applications, such as using measurements of a transmission line loaded with some material to determine the properties of the material, none of the elements of the telegraphers' equation can be ignored.

At higher frequencies, the loss of a cable is increased due to skin effect, which can be shown to increase as the square root of frequency (Collin 1966).

$$r = \sqrt{\frac{\omega\mu}{2\sigma}} \tag{1.78}$$

Thus, the insertion loss of an airline coaxial cable depends only upon the resistance per unit length of the cable, and so the insertion loss (in dB) per unit length, as a function of frequency, can be directly computed as

$$\begin{aligned} Loss(f) &= 8.68\frac{r}{4\pi Z_0}\left(\frac{1}{R_a}+\frac{1}{R_b}\right) \\ &= A \cdot f^{1/2} \end{aligned} \tag{1.79}$$

where R_a and R_b are the inner and outer conductor radius and r contains the square root of frequency. Thus, all the attributes can be lumped into a simple single loss-term, A. Figure 1.14 shows the loss of a 10 cm airline as well as the idealized loss, as described in Eq. (1.79), where good agreement to theory is seen. However, the introduction of dielectric loading of the coaxial line will add some additional loss due to the loss tangent of the dielectric.

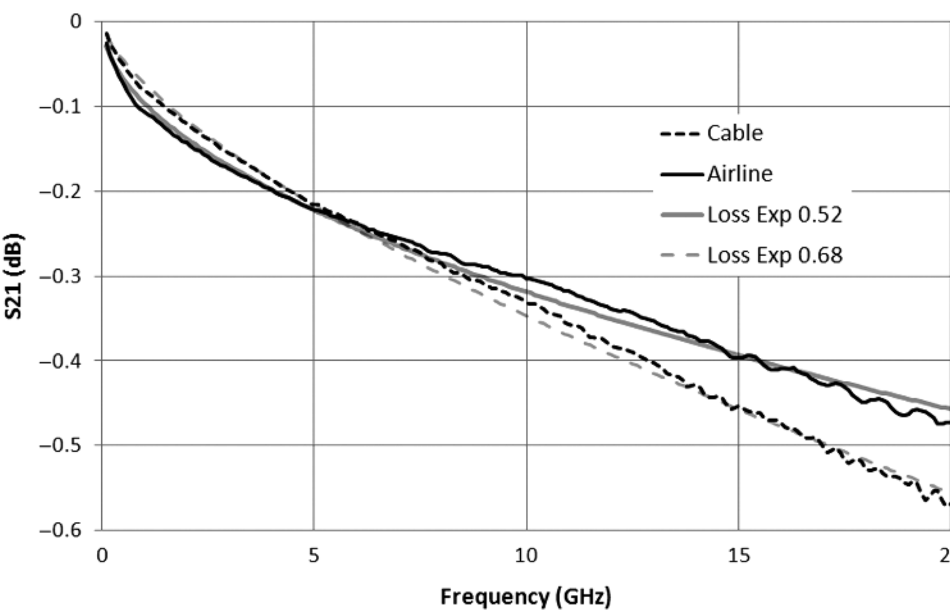

Figure 1.14 Loss of a 15 cm airline and a 15 cm semi-rigid Teflon-loaded coaxial line.

This additional loss often presents itself as an equivalent conductance per unit length, and this loss is often more significant that the skin-effect loss. Because of dielectric loss, the computed loss of (1.79) fails to fit many cables. The equation can be generalized to account for differing losses by modifying the exponent to obtain

$$Loss(f) = A \cdot f^b \tag{1.80}$$

where the loss is expressed in dB, and A and b are the loss factor and loss exponent. From the measured loss at two frequencies, it is possible to find the loss factor and loss exponent directly, although better results can be obtained by using a least-squares fit to many frequency points. Figure 1.14 shows the loss of a 15 cm section of 0.141 in. semi-rigid coaxial cable. The values for the loss at one-fourth and three-fourths of the frequency span are recorded. From these two losses, the loss factor and exponent are computed as

$$L_1 = A \cdot (f_1)^b, L_2 = A \cdot (f_2)^b$$

Taking the log of both sides, this can be turned in to a linear system as

$$\begin{aligned} \log(L_1) &= \log(A) + b \cdot \log(f_1) \\ \log(L_2) &= \log(A) + b \cdot \log(f_2) \end{aligned} \tag{1.81}$$

This system of linear equations can be solved for the loss factor A and the loss exponent b.

$$A = \exp\left(\frac{\log(f_1) \cdot \log(L_2) - \log(f_2) \cdot \log(L_1)}{\log(f_1) - \log(f_2)}\right) \tag{1.82}$$

$$b = \frac{\log(L_1) - \log(L_2)}{\log(f_1) - \log(f_2)} \tag{1.83}$$

The computed loss for all frequencies from Eq. (1.80) is also shown, with remarkably good agreement to the measured values over a wide range. Ripples in the measured response are likely due to small calibration errors, as discussed in Chapter 5.

The insertion phase of a cable can likewise be computed; in practice, a linear approximation is typically sufficient, but the phase of a cable will vary with frequency beyond the linear slope due to loss as well.

The velocity of propagation for a lossless transmission line is

$$v = \frac{1}{\sqrt{l \cdot c}} \tag{1.84}$$

The impedance of a lossy cable *must* be complex from Eq. (1.77), and thus the phase response must deviate from a pure linear phase response, due to the phase velocity changing with frequency at lower frequencies. A special case for airlines, which have no dielectric loss, is

$$v_{prop} \approx \left.\sqrt{\frac{2\omega}{rc}}\right|_{\omega \cdot l \ll r} \tag{1.85}$$

For cables in general, the dielectric loss will cause a deviation in the velocity of propagation similar to that seen for loss. So far the discussion has focused on ideal low-loss cables,

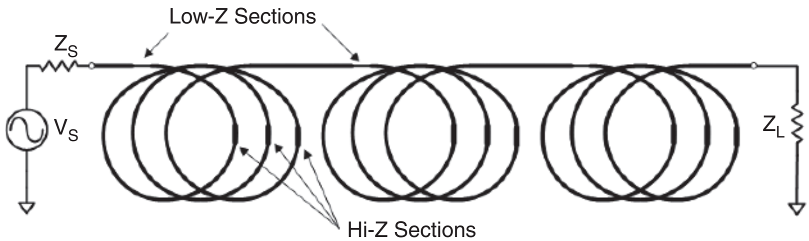

Figure 1.15 A model of a coax line with periodic impedance disturbances.

but in practice cables have defects that cause the impedance of the cable to vary along the cable. If these defects are occasional, they cause little concern and are typically overlooked unless they are so large as to cause a noticeable discrete reflection (more of that in Chapter 5). However, during cable manufacturing it is typical that the processing equipment contains elements such as spooling machines or other circular equipment (e.g. pulleys, spindles). If these have any defects in the circularity, or even a discrete flaw like a dimple, it can cause minute but periodic changes in the impedance of the cable. A flaw that causes even a one-tenth Ω deviation of impedance periodically over a long cable can cause substantial system problems called *structural return loss* (SRL), as shown in Figure 1.15. These periodic defects add up all at one frequency and can cause very narrow (as low as 100 kHz BW) very high return loss peaks, and thereby cause insertion loss dropouts at these same frequencies. In practice, the SRL test is the most difficult for low-loss, long-length cables such as those used in the CATV industry. Figure 1.16 shows a simulation of a SRL response caused by a 15 mm long, 0.1 Ω impedance variation, every 30 cm, and another −0.1 Ω variation every 2.7 m, each on the same 300 m coaxial cable with an insertion loss typical for main-line CATV cables. In the figure, two SRL effects are shown; a smaller effect every 50 MHz or so, due to the 2.7 m periodic variation, and a much higher effect every 500 MHz or so due to the 30 cm impedance variation. The higher impedance variation occurs more often, and so the periodic error will have a greater cumulative effect resulting in a nearly full reflection, as shown in the figure.

1.8.2 Connectors

Connectors provide the means to transition from one transmission media to another. They are often not considered as part of the device or measurement system, but their effects can dominate the results of a measurement, particularly for low-loss devices. Connectors can be distinguished by the quality and application. One remarkable aspect of connectors is the great difficulty in measuring them with any kind of accuracy. This is because most connectors provide a transition between different media, such as from a coaxial cable to a connector interface or from a PC board to a connector interface. While the connector interface is often well defined, the "back-end" of the connector is poorly defined.

Connectors that are "in series" provide transitions from male to female and provide interconnections between components. These are easiest to characterize as the ports are well defined and typically calibration kits are available and calibration methods are well understood. Connectors that are "between series" are equally well defined, but until recently they have been difficult to characterize as there were not well-defined standards for between-series

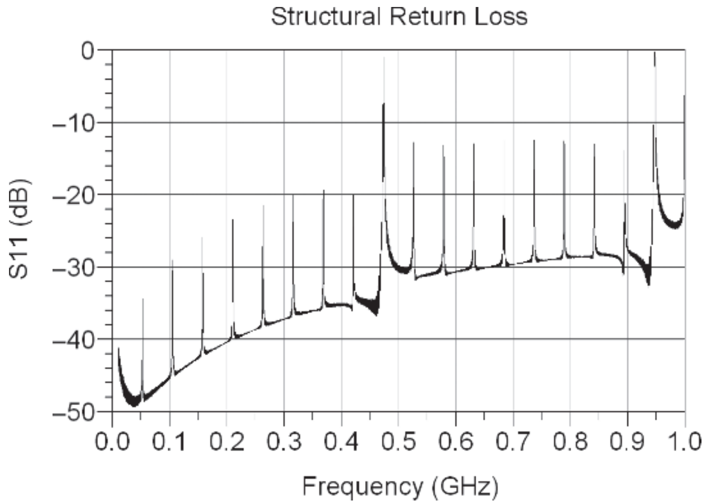

Figure 1.16 The return loss of a line with structural return loss.

Figure 1.17 In-series and between-series connectors.

adapters. Recent improvements in calibration algorithms have essentially eliminated any difficulty with characterizing these between-series adapters. Figure 1.17 shows some examples of in-series and between-series connectors.

For microwave work, there are some commonly utilized connector types that are found on the majority of components and equipment. Table 1.1 lists these common connectors along

Table 1.1 Test connectors used for RF and microwave components

Name	Outer Conductor Diameter (mm)	Rated Frequency (GHz)	First Mode (GHz)	Maximum Usable Frequency (GHz)
Type-N (50 Ω) precision	7	18	18.6	26.5[a]
Type-N (50 Ω) commercial	7	12	12.5	15
Type-N (75 Ω) precision	7	18	18.6	18
Type-N (75 Ω) commercial	7	12	12.5	15
7 mm	7	18	18.6	18
SMA	3.5	18	19	22
3.5 mm	3.5	26.5	28	33
2.92 mm (“K”)	2.92	40	44	44
2.4 mm	2.4	50	52.5	55
1.85 mm (“V”)	1.85	67	68.5	70
1 mm	1	110	120	130

[a]Some instrument manufacturers place this connector on 26.5 GHz instruments because it is rugged; it has the same first modes as Type-N and 7 mm.

with their normal operating frequency range. These are divided into three broad categories: precision sexless connectors, precision male/female connectors, and general-purpose or utility connectors. These connectors are typically 50 Ω, but a few can be found as 75 Ω versions as well.

From Table 1.1 one can see that there are three frequencies associated with connectors: the generally understood operating frequency (often dictated by the calibration kit’s maximum certified frequency), the frequency of the first mode, and the maximum frequency determined by the waveguide propagating mode of the outer conductor. The operating frequency is always below the first mode and usually by several percent. The first mode in many connectors is due to the support structure for the center pin. It is often made of some plastic material and thus has higher dielectric constant and a lower frequency to support a mode. In connectors and cables, *modes* is the term used to refer to non-transverse-electromagnetic (non-TEM) propagation that can occur in a circular waveguide mode defined by the inside dimension of the outer conductor. Adding dielectric in the bead that supports the center pin can theoretically lower the mode frequency, but if the bead is short, the mode will be evanescent (non-propagating) and may not affect the quality of the measurement. At a somewhat higher frequency, there will be a propagating mode in air for the diameter of the center conductor, but if the cable attached to the connector is sufficiently small, this mode may not propagate as well. It is the propagating modes that cause the significant dips in the response, and more importantly, these dips cannot be removed with calibration because they are not localized and because reflections in the mode of transmission far removed for the connector interface can interact with these connector modes, causing the frequency response of the mode effect to change when different devices are connected. If the response of the mode does not change when other devices are connected, it can be calibrated out.

The precision sexless connectors are now found only in metrology labs. Their chief benefit was a repeatable connector that has identical characteristics for each connector. As such, it was easy to create a system calibration, and any part with such connectors could be inserted between two cables in either direction. This was important because in the past it was difficult to deal with "non-insertable" devices from a calibration sense (a non-insertable device is one with the same sexed connector on each port, e.g. female-female). The 7 mm connector is often found on precision attenuators and airlines used as transfer standards. The 7 mm connector is also known as the GPC-7 for general precision connector, and often as the APC-7™ for amphenol precision connector. Because these connectors are sexless, there is no need for adapters to provide interconnections between devices or between devices and cables.

1.8.2.1 7 mm Connector (APC-7, GPC-7)

The 7 mm connector has a couple of interesting attributes: the center pin has no slots but contains spring-loaded center collets that protrude slightly from the mating surface, as shown in Figure 1.18.

When mated, the collet from each connector floats against each other, providing a good center contact. There is a slight gap in the slotless outer sleeve of the center pin. As with almost all RF connectors, the outer conductor forms the physical mating plane. On most connectors, there is a slip-ring-threaded sleeve surrounded by a coupling nut. To mate, the threaded sleeve is extended on one connector and retracted on the other. On the retracted connector, the coupling nut is extended to engage the other's sleeve and is tightened. Only one coupling nut should be tightened, although it is common but incorrect practice to tighten the other coupling nut. In fact, tightening both coupling nuts can result in the center pins pulling apart and a poorly matched contact. Occasionally, one sees parts that contain only a solid threaded outer conductor (serving the purpose of the threaded sleeve) and no coupling nut. These are more common on older test fixtures intended to mount directly the 7 mm connectors of network analyzer test sets.

1.8.2.2 Type-N 50 Ω Connector

The Type-N connector is common in lower-frequency and higher-power radio frequency (RF) and microwave work. It has the same outer diameter (7 mm) as the 7 mm connector but is sexed. In fact, this connector has the unusual attribute of having the mating surface for the

Figure 1.18 A 7 mm connector.

outer conductor (which is almost always the electrical reference plane) recessed for the female connector. Thus, the female pin protrudes (in an electrical sense) from the reference plane, and the male pin is recessed. Thus, the calibration standards associated with Type-N connectors have electrical models that are highly asymmetric for male and female standards.

The Type-N connector has precision forms, including ones with slotless connectors (metrology grade), ones with precision six-slotted collets and solid outer conductor sleeves (found on most commercial test equipment), and commercial forms with slotted outer conductor sleeves and four or even two slotted female collets. Slotless connectors have a solid hollow cylinder for the female connector with an internal four- or six-finger spring contact that takes up tolerances of the male center pin. As such, the diameter of the female center pin does not depend at all on the radius of the male pin. Typical female contacts with collets expand or contract to accept the male pin, and thus their outer dimension (and thereby their impedance) varies with the diameter tolerance of the male pin.

The commercial forms are found on a variety of devices and interconnect cables. The male version of these commercial-grade parts present two common and distinct problems: there is often a rubber "weather-seal" o-ring in the base of the connector, and the outer nut of the male connector is knurled but has no flats to allow using a torque wrench. The first problem exacerbates the second, as the mating surface of the outer conductor of the male connector is often prevented from contacting the base of the female connector because the outer (supposedly non-mating) surface of the female connector touches the rubber o-ring and prevents the male outer conductor from making full contact. If one can fully torque a Type-N connector, the rubber o-ring would compress, and the contact of the male outer conductor would occur, but as there are no flats for a torque wrench, it is difficult to sufficiently torque the Type-N connector to get good repeatable connections. This one issue is the cause of hundreds of hours of retest when components don't pass their return-loss specs. The solution is quite simple: remove the rubber o-ring from the base of the male connector, always, before any measurement. A pair of tweezers and a needle-nose pliers are indispensable for the process of removing this annoying o-ring. One will note that none of the precision versions of Type-N connectors contains such an o-ring. Figure 1.19 shows some examples of Type-N connectors; the upper two are commercial grade, and the lower two are precision grade. Figure 1.20 shows the insertion loss measurement of a male-to-male Type-N adapter mated to a female-to-female Type-N adapter for a precision pair and a commercial-grade pair, where the loss is normalized to the length of the adapter. The commercial-grade pair is operational only to about 12 GHz, due to moding in the connector. The precision N is mode free beyond 18 GHz.

1.8.2.3 Type-N 75 Ω Connector

Type-N connectors also have a 75 Ω version, which has the same outer dimensions but a smaller center conductor. This is in some ways unfortunate as the smaller female collet of the 75 Ω version can be damaged when inserted with a 50 Ω male pin. There are a couple of versions of the 75 Ω female collet, one with short slots and six fingers, and one with long slots and four fingers. A precision slotless version is also available. The short slot version has the potential for better measurements, as the slots expand less so there is less uncertainty of the open capacitance. However, on many products with 75 Ω N-connectors, the long slot connector is used; the long slots were designed to accept a 50 Ω male pin, at least for a few

Figure 1.19 Examples of Type-N connectors: commercial (upper) and precision (lower).

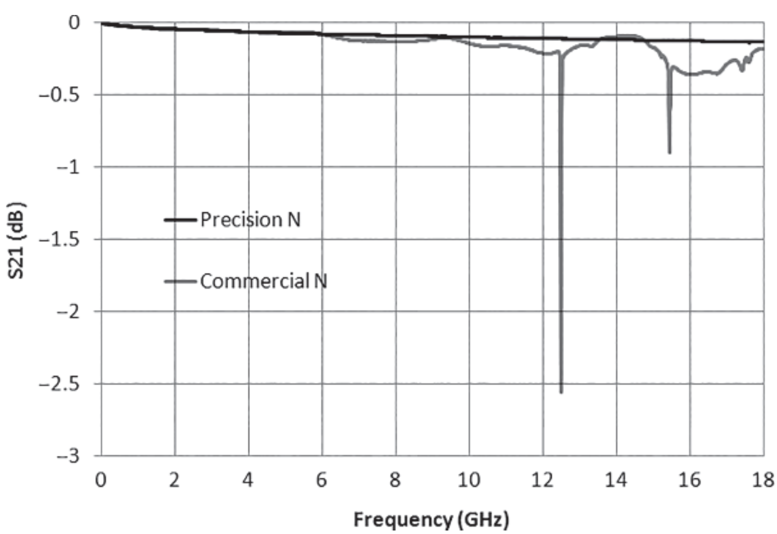

Figure 1.20 Performance of a precision and a standard Type-N connector.

Figure 1.21 75 Ω Type-N connectors: commercial (upper) and precision (lower).

insertions, without damage. Often the 75 Ω components have an extra machined ring or line on the outer nut to help identify it. Versions of 75 Ω Type-N connectors are shown in Figure 1.21. An example of the insertion loss measurement of a mated pair of a male-to-male adapter with a female-to-female adapter is shown in Figure 1.22, where the loss is normalized for length of the adapter. The frequency limit of Type-N 75 is often stated as 2 or 3 GHz, but that is because the commercially available calibration kits were rated only to those frequencies. In practice, these connectors could be used up to 7 or 8 GHz without difficulty. The response of the commercial-grade connector is likely limited not due to moding (since the loss signature is quite low Q) but rather due to poor impedance control in the center pin support bead, causing impedance mismatch.

1.8.2.4 3.5 mm and SMA Connectors

The 3.5 mm connector is in essence half the scale of the N connector and provides higher-frequency coverage. The center pin of the 3.5 mm connector is supported by a plastic bead, rather than solid dielectric, meaning it has mode-free operation to a much higher frequency than Type-N. Traditionally, 3.5 mm connectors are specified up to 26.5 GHz,

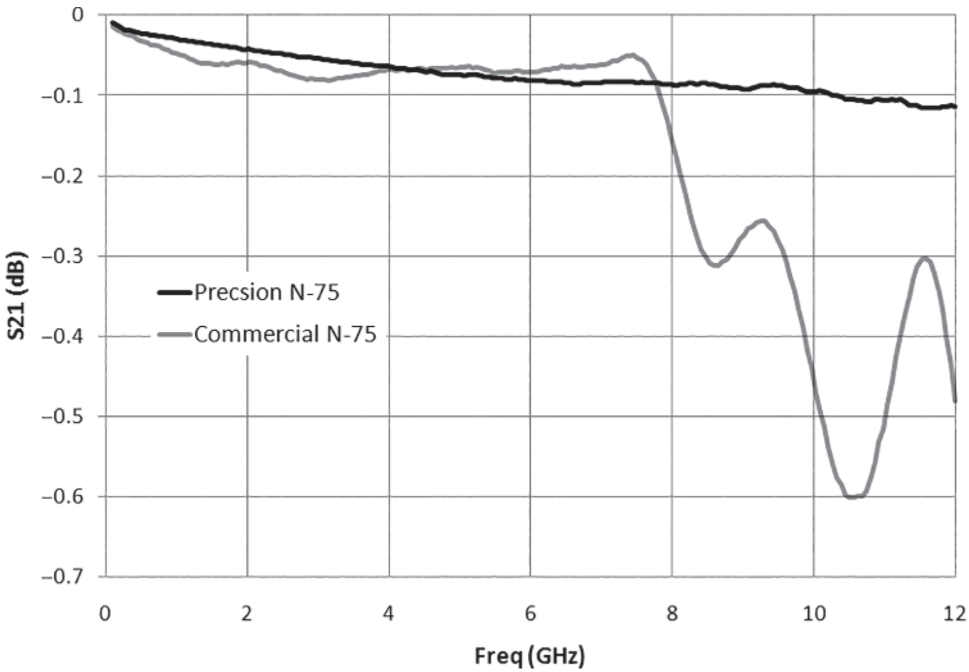

Figure 1.22 Insertion loss of 75 Ω connectors.

but their first mode is nearly 30 GHz, and they are functional up to about 38 GHz. An interesting aspect of modes is that the first mode of a 3.5 mm connector is due to the bead (and its increased effective dielectric), but this mode is non-propagating, so it is reasonable to use these connectors to even higher frequencies. The 3.5 mm female connector comes with several versions of center pin, the main varieties being a four-slot collect and a slotless precision connection, found now on most calibration kits. Interestingly, even though the slotless connectors may have the center spring contact damaged due to oversized or misaligned male pins (under the microscope one or more fingers may be crushed back into the hollow of the female pin), the RF performance is almost unaffected due the robust solid outer conductor. In fact, one typically can tell if a slotless connector is damaged only by visual inspection, as the RF performance is substantially unchanged, as long as even one finger is left to make contact.

The SMA connector is mechanically compatible with the 3.5 mm connector but has a solid Teflon dielectric and thus a lower operating frequency due to moding. SMA is traditionally considered to be an 18 GHz connector, but the first propagating mode is well above 20 GHz, depending upon the type of cable that is connected to the SMA connector. The chief advantage of SMA connectors is low cost, especially when mounted to semi-rigid coaxial cables. The dimensions are such that the center wire of the coax can be used a connector pin for SMA, and only an outer conductor sleeve needs to be added to the coax outer conductor to form a male connector, shown in the lower-right picture of Figure 1.23. But these cables are notoriously bad at maintaining the proper dimensions for the center pin, and often the center pins are poorly trimmed and improperly chamfered so that they cause mating problems with their

Figure 1.23 3.5 mm (f) and (m) (upper left); SMA (f) and (m) connectors (upper right); 3.5 mm (lower left) and SMA adapters (lower right).

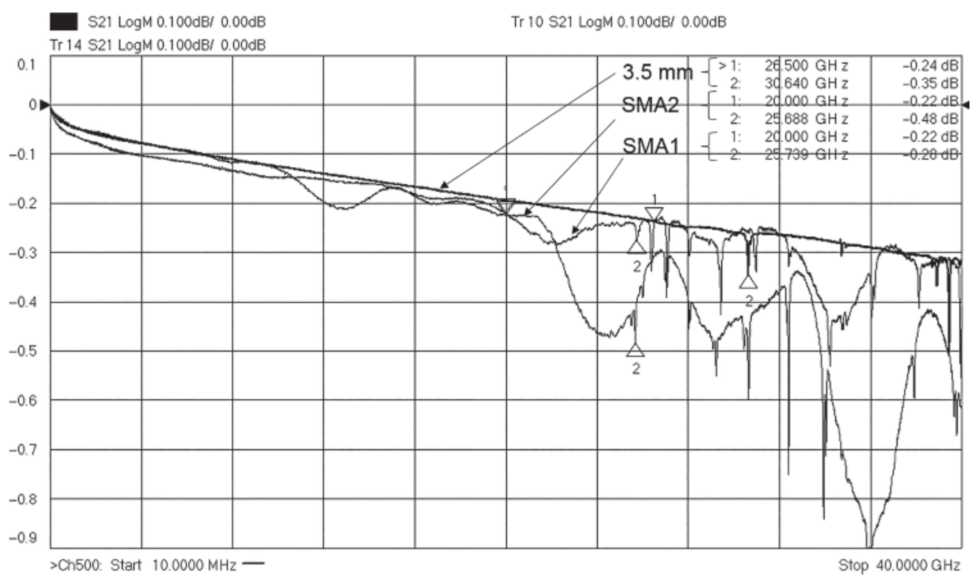

Figure 1.24 Performance of SMA and 3.5 mm mated-pair connectors.

female counterparts. This is particularly true when mating them to 3.5 mm female connectors, slotless ones in particular. Figure 1.23 shows examples of 3.5 mm and SMA connectors, with 3.5 mm on the left and SMA on the right.

Figure 1.24 shows measurement plots of a mated pair of 3.5 mm male-to-male with a 3.5 mm female-to-female, as well as two SMA examples. The moding of the SMA connector is clearly seen above 25 GHz (Marker 2 on the SMA1 and SMA2 trace). The moding of the 3.5 mm connector is seen just above 30 GHz (Marker 2) and again at 34 and 38 GHz. There two typical construction types for SMA, one with a press-fit of the Teflon and center conductor (SMA1 in the measured response) and one where the Teflon is held in with a small dot of epoxy through a hole in the outer conductor (SMA2 in the measured response). The second method usually gives a poorer match, and we can see that with the small dip in the S21 response of SMA2 near 12 GHz and the larger dip just above 20 GHz.

1.8.2.5 2.92 mm Connector

The 2.92 mm connector is scaled down from the 3.5 mm connector and can be mechanically mated to both the 3.5 mm and the SMA connectors. The smaller diameter outer conductor means that its mode-free operation extends proportionally higher, to 40 GHz, and is usable to perhaps 46 GHz. The female connector has a two-slot collet that provides sufficient compliance to mate with the center pin of the larger 3.5 mm and SMA connectors but that makes it less suitable for precision measurements due to increased uncertainty of the contact point on the center pin radius, which now depends upon the radius of the pin that is inserted. A further point is that the metal wall of the female collet on the 2.92 connector is quite thin and prone to damage if the mating pin is not well aligned or oversize. It's not uncommon to find 2.92 female adapters missing one of the collet fingers. The 2.92 mm connector was popularized by the Anristu company (formally Wiltron), which introduced it as the K connector, and it is common to hear any 2.92 mm connectors referred to by that name.

Figure 1.25 shows some examples of 2.92 connectors. The key difference is in the diameter of the inside of the outer conductor. Figure 1.26 shows the insertion loss of a mated pair of 2.92 mm female-to-female adapters with a 2.92 male-to-male adapter, along with an example of a 3.5 mm mated adapter pair. The moding of the 3.5 mm pair is clearly seen above 30 GHz, but the connector is generally usable up to 38 GHz as the first small modes are bead modes and are able to be calibrated out as they generally don't propagate through the cable.

1.8.2.6 2.4 mm Connector

The 2.4 mm connector is essentially a scaled version of the 3.5 mm connector, with an associated scaling in maximum frequency. It is used extensively on 50 GHz applications, though it can be used up to 60 GHz. This connector cannot be mated to any of the SMA, 3.5 mm or 2.92, and in fact was designed to prevent damage if one tried to mate to these types. It comes with both slotted and slotless female center pins, much like the 3.5 mm connector.

1.8.2.7 1.85 mm Connectors

There are two variants of the 1.85 mm connectors, designed originally by Anritsu and Agilent. The Anritsu variety is called the V connector, and the Agilent variety is called the 1.85 mm connector. They are mechanically compatible and were originally designed for 67 GHz operation, usable to above 75 GHz. These connectors are mechanically compatible

Figure 1.25 A 3.5 mm connector compared with 2.92 mm female (upper) and male (lower).

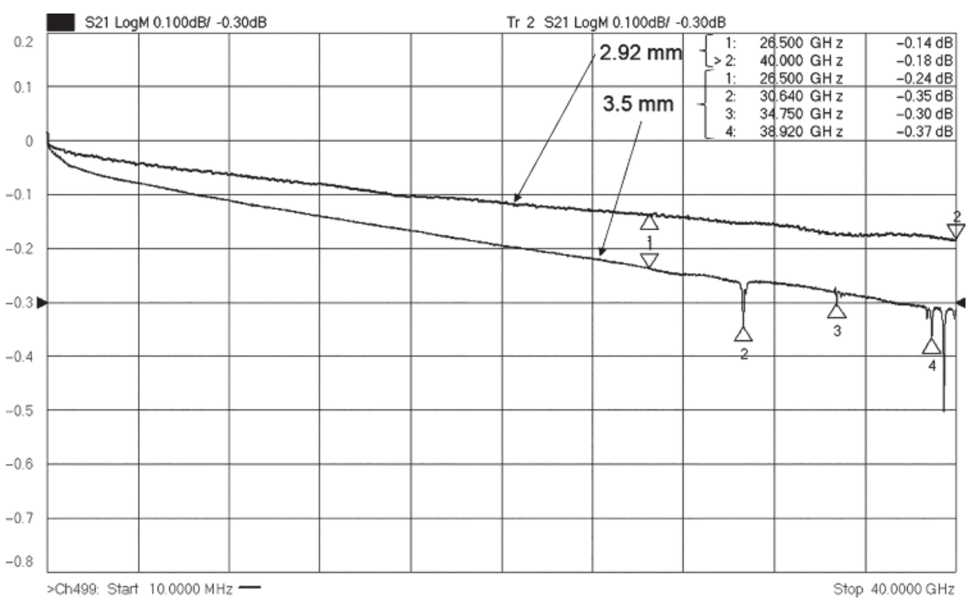

Figure 1.26 Performance of a mated pair, 2.92 compared with 3.5 mm.

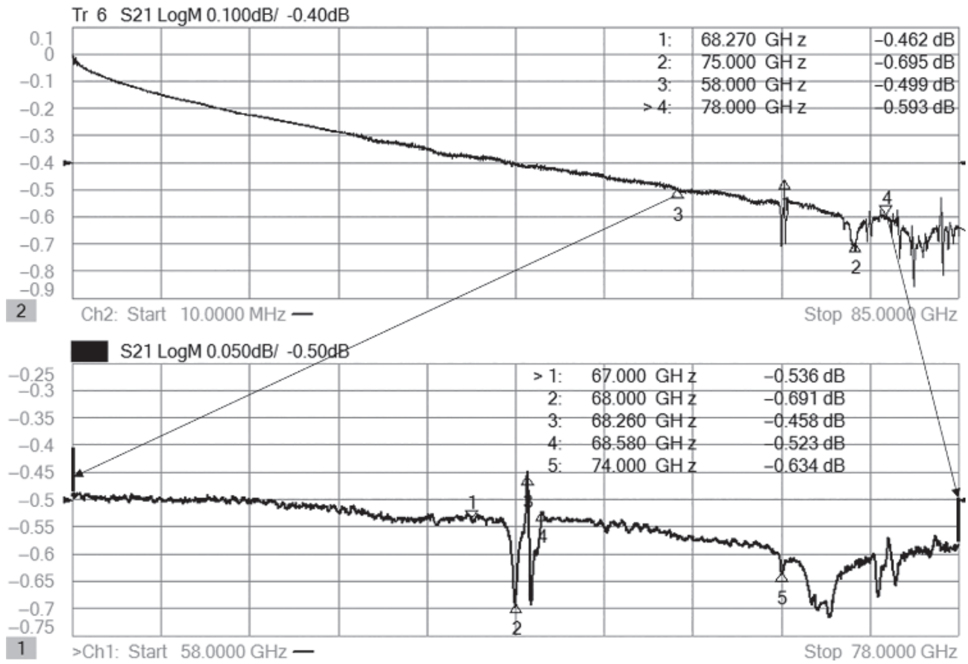

Figure 1.27 Response of mated pair of male-to-male and female-to-female 1.85 mm connectors (upper), with zoomed-in view of the first mode (lower).

with the 2.4 mm connector. Figure 1.27 shows the wideband response to 85 GHz of the 1.85 mm connector in the upper plot and a zoomed-in view of the first mode in the lower plot. This first mode is a "bead-mode" and is caused by the increased dielectric constant of the bead, which holds the center pin, lowering the first mode of the 1.85 mm coaxial line. In general, this mode is non-propagating (since it is contained in the bead) and may be calibrated out in some circumstances. For example, if this connector is used with an on-wafer probe and the coax from the connector to the probe-tip is mode free, then the bead mode will act like a small, stationary resonance that can be removed with a calibration. If the mode is propagating, then changes in the termination impedance change the effects of the mode, and it cannot be calibrated out (it is not stationary with respect to an external impedance); but if it is non-propagating and there is a sufficient length of mode-free line (such a cable) between this bead mode and the reference plane, the evanescent fields associated with the mode will die off before arriving at the reference plane and thus will not couple to the terminating impedance. As this first mode is less than 0.2 dB, in many cases it is not significant. The 1.85 mm connector has been used out past 75 GHz.

1.8.2.8 1 mm Connector

The 1 mm connector is essentially a scaled version of the 1.85 mm connector but cannot be mated to it. It is typically specified to 110 GHz performance but is usable to above 120 GHz,

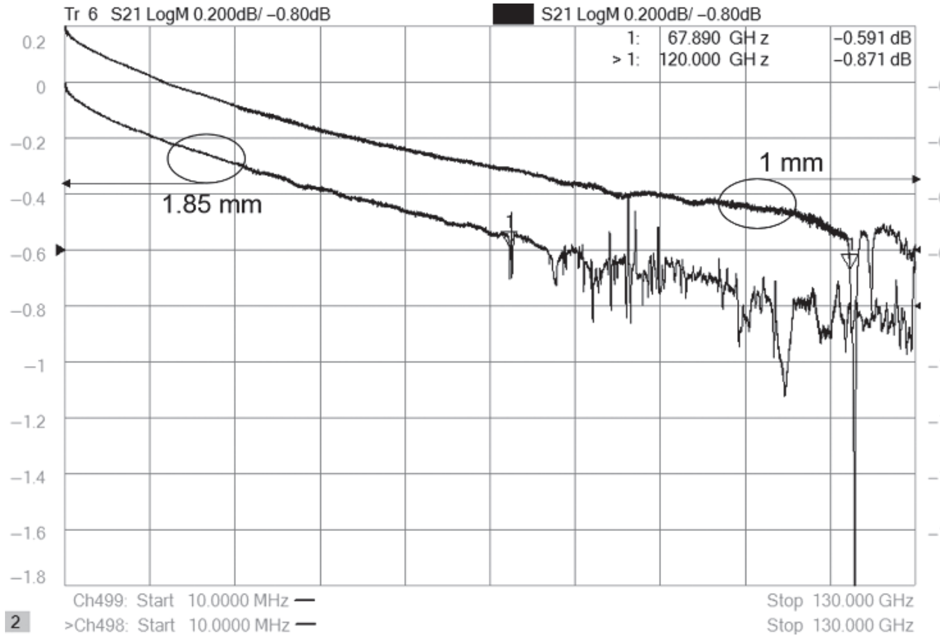

Figure 1.28 Response of a 1 mm mated pair and a 1.85 mm mated pair.

with some versions being specified up to 120 GHz and used up to 140 GHz. Figure 1.28 shows the response of a mated pair of 1 mm male-to-male and female-to-female adapters (right scale) along with the same 1.85 mm mated pair of Figure 1.27 (left scale). The reference line is offset by 1 division to make it easier to see the traces; without the offset, the traces would lie nearly on top of each other. The mode for 1 mm is a bit of a deeper mode but is now out past 120 GHz. There is a second mode at 124 GHz, but both are also non-propagating, so it may be possible to remove them with calibration. The depth of the mode does imply that it may be from the multiple beads in the mated pair and not be very stable with changes in temperature.

1.8.2.9 PC Board Launches and Cable Connectors

For many design and measurement applications, the circuit of interest is embedded in a PC board. There are many types and styles of PC board launches, which typically have an SMA connector on one end and PC board contacts at the other, as well as miniature versions such as the QMA connector. These can come in edge launch as well as right angle, and their performance depends greatly upon the mounting pattern on the PC board trace. These can be difficult to characterize because only one end is available in a standard connector. An example of a common PC board launch is shown in Figure 1.29. Measurement techniques for these devices, as well as methods to remove their effects from the measurement of on-board PC components, are discussed in Chapter 11.

Connectors designed for coaxial-cables provide similar challenges, as the cable to which they are attached affects the quality of the connection, and the common practice of attaching

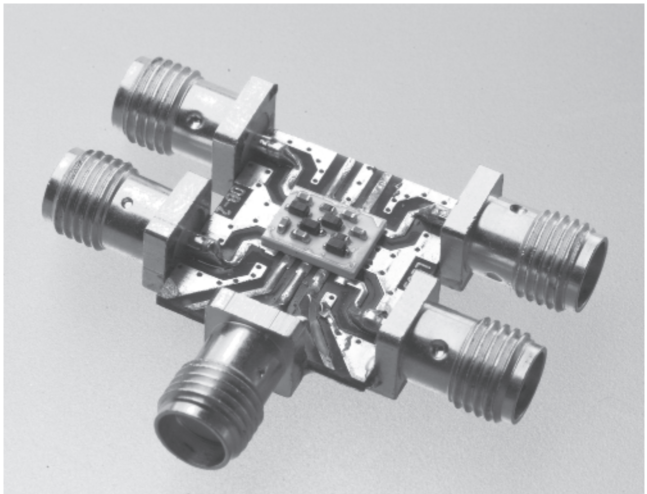

Figure 1.29 PC board SMC launches.

two connectors to each end of cable makes it difficult to separate the effects of one from the other. Time-domain techniques can be applied to remove these unwanted effects, as described in Chapter 5.

1.8.3 Non-coaxial Transmission Lines

Transmission lines provide the interconnection between components, typically in a microcircuit or a PC board. These are distinguished from a measurement perspective because they are typically much shorter, often not shielded, and the interface to them is not easy to make and sometimes not well defined. While there have been whole books written on the subject, a short review of some common transmission line structures and their attributes are described next, with a focus on attributes important to measurement. Transmission lines are characterized by the same three parameters: impedance, effective dielectric constant, and loss.

1.8.3.1 Microstrip

Certainly the most widespread transmission line must be the microstrip line, shown in Figure 1.30. This is found in planar structures such as PC boards and micro-circuits. Consisting of a thin strip of metal on a dielectric substrate, over a ground plane, it is used for connection between components as well as creating transmission line components such as couplers and filters (Hong and Lancaster 2001).

The computation of the transmission parameters has been fully documented in many forms, but for measurement purposes these lines are typically 50 Ω (or the equivalent system impedance) even though as a design element they can take on any value. For most applications, the dielectric constant is 10 or less, so the *w/h* ratio is greater than 1 for 50 Ω.

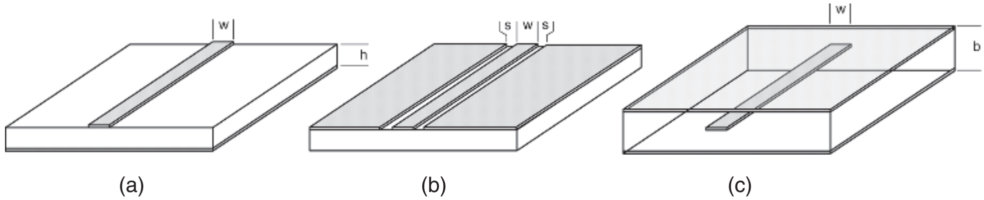

Figure 1.30 Planer transmission lines: microstrip (a), coplanar waveguide (b), strip line (c).

The approximate impedance can be computed as (Pozar 1990)

$$Z_{\mu strip} = \begin{cases} \dfrac{60}{\sqrt{\varepsilon_{re}}} \ln\left(\dfrac{8h}{w} + \dfrac{w}{4h}\right) & for\ \dfrac{w}{d} \leq 1 \\ \dfrac{377}{\sqrt{\varepsilon_{re}}\left[\dfrac{w}{h} + 1.393 + 0.677 \ln\left(\dfrac{w}{h} + 1.444\right)\right]} & for\ \dfrac{w}{d} \leq 1 \end{cases} \tag{1.86}$$

where ε_{re} is the effective relative-dielectric-constant, found from

$$\varepsilon_{re} = \left(\frac{\varepsilon_r + 1}{2}\right) + \left(\frac{\varepsilon_r - 1}{2}\right) \cdot \left(1 + 12\frac{h}{w}\right)^{-1/2} \tag{1.87}$$

The effective relative-dielectric constant sets the velocity factor of the transmission line, but in microstrip, some of the fields travel in the substrate and some in air. Therefore, the transmission is not purely transverse-electromagnetic (TEM), and some structures become more difficult to design, particularly coupled lines, the even and odd mode velocity factors of which are not the equal. Since the line is not pure TEM, at high frequency, dispersion effects will become apparent where the effective delay of the line is not constant with frequency.

The loss of microstrip lines is difficult to compute accurately because it depends upon many factors including the conductivity of the microstrip line and the ground plan, the dielectric loss of the substrate, radiated loss to the housing or shield, and losses related to both surface roughness and edge roughness. These roughness losses can be significant in PC board and low-temperature cofired-ceramic (LTCC) applications and are dependent upon the particular processes used. While there are high-quality PC board materials (Duriod™ or GTEK™ are common trade names), the material known as FR4 is most common, and the dielectric constant and loss of this PC board material can be uncertain. The finished substrate can be comprised of layers of board material sandwiched together with glue, and the final thickness can depend upon processing steps, so it is best when evaluating microstrip transmission lines to produce sample structures that can help determine the exact nature of the material.

One high-performance material used is single-crystal sapphire, and it has the unusual property of having a dielectric constant that has a directionality, with a higher constant of 10.4 in one of the three dimensions, and a lower constant of 9.8 in the other two. A second, common high-performance dielectric is ceramic found in thin-film, thick-film, and LTCC applications. It has a uniform dielectric constant typically between 9.6 and 9.8 depending upon the purity and grain structure of the ceramic.

1.8.3.2 Other Quasi-Microstrip Structures

For many applications, the size of 50 Ω microstrip line is not suitable for connections to very large devices. Some common modifications are *suspended substrate* microstrip line, where the ground plan has been removed some distance from the dielectric. This has the effect of lowering the effective dielectric constant and raising the impedance of the line. In this way, a wider line can be used to connect to a wide component and still maintain a matched impedance. A *shielded* microstrip line is entirely enclosed (the theoretical models of microstrip lines assume no top shield), and the top metal tends to lower the impedance of the line. This is particularly true for suspended microstrip lines.

1.8.3.3 Coplaner Waveguide

One difficulty with microstrip transmission lines is that the ground and signal conductors are on different physical planes. Coplanar waveguide (CPW), as the name implies, provides a coplanar structure of ground-signal-ground, as shown in Figure 1.30b. An alternative is grounded coplanar where the backside is a conductor as well, and in practice, all coplanar lines have associated package ground, but the ground may be ignored if there is a substantial air-gap between the substrate and the package ground. The references provide some computations of coplanar waveguide impedance for various configurations (Wen 1969; Simons 2001). In microwave measurements, CPW is used extensively as a contacting means for on-wafer measurements and is used to provide extremely low ground inductance for measuring microwave transistors and circuits, as shown in Figure 1.31, with either topside grounds (left) or backside grounds (right). Note that since the impedance depends only upon the scale of width to space, this allows contacts of large scale (such as probes) to be transitioned to small scale such as IC devices.

CPW has some inherent problems due to the ground being on a surface plane or sheet. In many instances the CPW line is mounted in a metal package, and the ground plane is grounded at the package wall. If the distance from the package wall to the ground plane edge approaches a quarter-wavelength at the frequency of interest, or multiples thereof, then a transmission line mode can form such that the ground of the CPW appears as an open relative to the package ground. This concept of "hot grounds" for CPW has been observed in many situations and

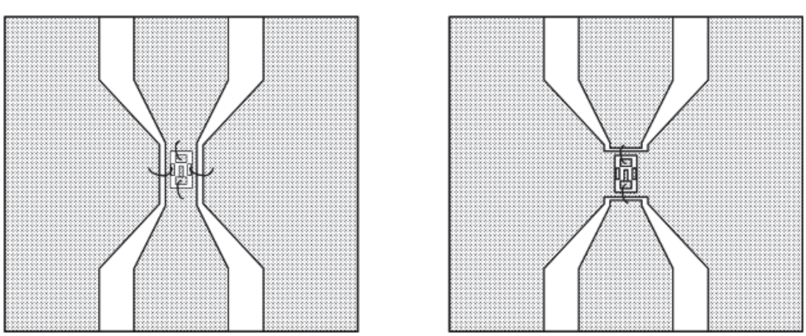

Figure 1.31 CPW-mounted IC.

is sometimes avoided by periodically grounding one side to the other through a small cross connection on the backside of the CPW. Another method is to provide a lossy connection to the sidewall ground through absorptive material or thin film resist material to suppress energy in the unwanted mode. Another alternative is treating the CPW as a suspended substrate only under the gap between ground and conductor and “stitching” the CPW ground to the backside ground through a series of conductive vias. The impedance of these structures is lowered by the added ground, so an adjustment of the center line width is usually made to accommodate the additional ground paths.

1.8.3.4 Stripline

More common as a transmission line on an inner layer of a PC board, strip line consists of a thin strip or rectangle of metal sandwiched between two ground planes embedded in a uniform dielectric constant, as shown in Figure 1.30c. The impedance of these lines is much lower than the equivalent-width microstrip line, but they have an advantage of being fully TEM in nature and so often the design of components such as coupled lines is easier as the even- and odd-mode velocity factors are the same. An approximate formula for computing the value of a stripline impedance with a zero thickness strip is (Pozar 1990)

$$Z_{stripline} = \frac{30\pi}{\sqrt{\varepsilon_r}} \frac{b}{W_e + 0.441b} \tag{1.88}$$

$$W_e = \begin{cases} w & for\ w/b > 0.35 \\ (w - (0.35 - w/b)^2 \cdot b & for\ w/b < 0.35 \end{cases} \tag{1.89}$$

More complex formulas that include a broad range of applicability and include effects for finite strip thickness and asymmetric placement of the strip can be found in many references (IPC 2004; Cohn 1954).

1.9 Filters

Filters come in a variety of types including low pass, band pass, high pass, and band stop. Multiport filters form diplexers or multiplexers, which are used to separate or combine signals of different frequency from a common port to a port associated with the different frequencies of interest. Diplexers are sometimes called *duplexers*, but duplexing is a function of the operation of a communication system. That is, a system that can transmit and receive at the same time is said to operating in a duplex mode. A diplexer is used to support the duplex operation by keeping the transmit signal from saturating the receiver.

The structure and variety of filters are almost endless, but they all share these common attributes: low loss in the pass band, low reflection in the pass band, high reflection, and high loss in the stop band. In nearly every case, the goal of the design is to minimize unwanted loss, and this quality of a filter is often referred to as the *Q* of the filter. In microwave cases, filters are designed to operate into a matched impedance, so there is always loss associated with power from the source being absorbed by the load. The Q of a filter in operation is fixed by the loading of the ports and can never be infinite. The quality of a filter is usually defined by its unloaded Q, which accounts for the (desired) power loss from the source to the load.

For many filters, the desired qualities are a trade-off between creating a maximally flat passband and creating a maximally sharp cutoff. Thus, the measurement of the transmission response of the filter is critical in evaluating the quality of a filter design. For most filters used in communications, the transmission responses is desired to be equally flat (rather than maximally flat) across the passband, resulting in filters that have Chevyshev-type response (equal ripple) in the passband (Zverev 1967). The desire for sharp cutoffs has led to many filters employing an elliptic response, which provides for finite zeros in the transmission response. Stopband performance of high-performance filters can also require careful consideration in measuring, with some requirements going beyond 130 dB of isolation over selected regions of the stop band. These extreme isolation requirements put tremendous burdens on the design of the filter, as well as the design and use of the measurement systems.

In modern communications systems using complex modulation, the phase response of the filters is also critical, and a significant design parameter is controlling the phase of the filter to follow a linear response, with a key measurement parameter being deviation from linear phase. Closely aligned to that is maintaining a constant group delay through the passband. Equalization techniques are utilized that can remove higher-order phase responses, such that another measure of filter phase response is deviation from parabolic phase, where the phase is fitted to a second-order response, and the deviation of the phase from this second-order response is the measurement criteria. Some filters are used as part of a feed-forward or matched system network where their phase response as well as absolute phase and delay must be carefully controlled.

The reflection response of filters is also a key measurement parameter. To the first order, any signal that is reflected is not transmitted so that high reflections lead to high transmission loss. However, the loss due to reflection for most well-matched filters is much less than the dissipation loss. Still, low reflections at the test ports are required to avoid excess transmission ripple from concatenated components, and even moderate reflections from filters in a high-power transmission path can cause damage to the preceding power amplifier. Thus, very low return loss is often a critical parameter of filters and also a difficult parameter to measure well. This becomes especially true in the case of diplex and multiplex filters, where the loading of any port affects the return loss of the common port.

For high-power applications, the filter itself can become a source of IM distortion, and the attribute passive inter-modulation (PIM) has become common in the measurement of these high-power filters. Poor mechanical contacts between components in a filter, poor plating on a filter, or the use of magnetic materials in the plating or construction of the filter can lead to hysteresis effects that cause IMD to be created in an otherwise passive structure. The level of IMD typically found in these filters is less than −155 dBc, but this can be a difficult spec to meet without careful design and assembly.

Most of these high-performance communication filters are designed using coupled-resonator designs (Cameron et al. 2007; Hunter 2001). Because of manufacturing tolerances, these filters cannot be manufactured to specification from the start; they require tuning of the resonators as well as the inter-resonator couplings. Techniques to optimize the response of these filters are highly sought and a key aspect of the filter measurement task, requiring fast precise response of the transmission and reflection response in real time.

Another type of filter commonly found in the intermediate frequency (IF) paths of receivers is a surface acoustic wave (SAW) filter. The frequency of these SAW filters has been steadily increasing, and they are sometimes found in the front end of a receiver. SAW filters can be

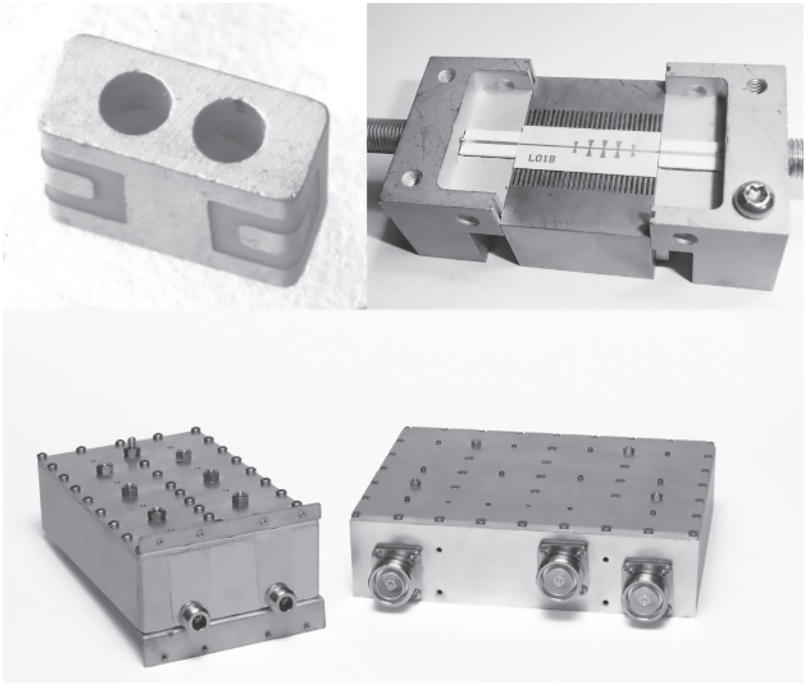

Figure 1.32 Examples of microwave filters: cellular phone handset filter (upper left), thin film filter (upper right), and cellular phone base station filters (bottom).

made to high orders and can have large delays (in the order of microseconds). Because of these long delays, special measurement techniques are required when attempting high-speed measurements. Another type of acoustic wave filters are the film bulk acoustic resonator (FBAR) filters, which are small in size and have been used as RF/TX duplexers in handset cell phones.

Ceramic coupled resonator filters are also used extensively in cell phone and radio applications. Because of manufacturing tolerances, the filters are often required to be tuned as part of the manufacturing process, and tuning consists of grinding or laser-cutting electrodes until the proper filter shape is obtained. This presents some difficulty in coupled resonator filters as the tuning is often "one way," and once the resonator frequency has been increased, it cannot be reduced again. This has led to the need for high-speed measurements to ensure that the latency between measurement and tuning is as small as possible.

Some examples of filters are shown in Figure 1.32.

1.10 Directional Couplers

Directional couplers separate the forward and reverse waves in a transmission system (see Section 1.3). A directional-coupler is classically defined as a 4-port device, often with a good load on the fourth port, as shown in Figure 1.33; but in practice a load element is almost always permanently attached. The directional-coupler has four key characteristics: insertion

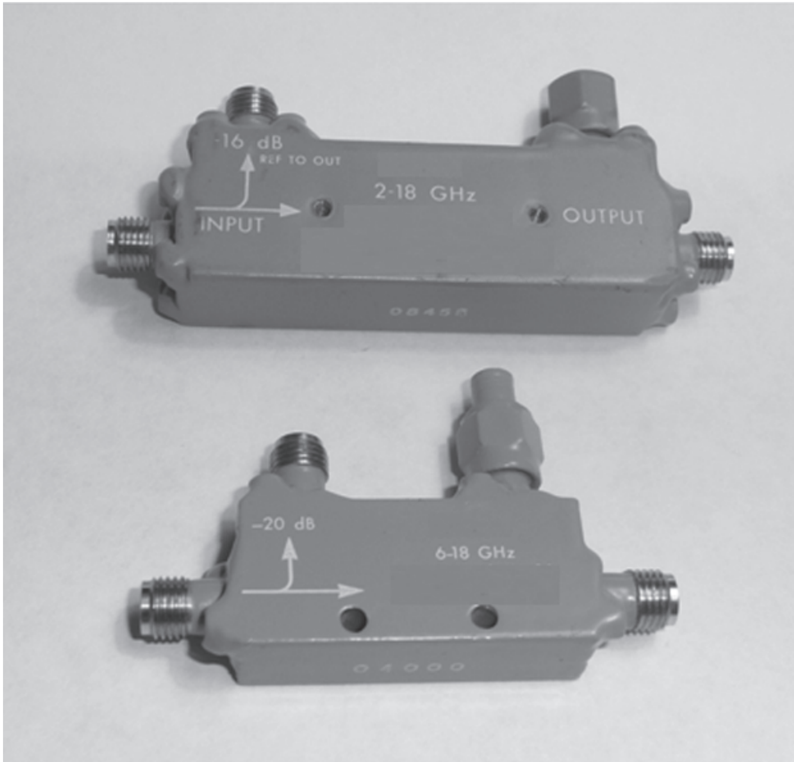

Figure 1.33 Directional couplers.

loss, coupling factor, isolation, and directivity. In fact, directivity is related to the other three factors in a specific way.

$$Directivity = \frac{Isolation}{Coupling \cdot Loss} \tag{1.90}$$

Most couplers have a nearly lossless structure so that the directivity is nearly equal to the isolation/coupling, but for lossy structures, such as directional bridges, the earlier definition provides the proper description. In fact, consider the case of a directional-coupler with 20 dB of coupling, 50 dB of isolation, and 0.05 dB of insertion loss, setting the directivity at nearly 30 dB. If a 10 dB pad is added to the input, as shown in Figure 1.34, the isolation is increased by 10 dB, the loss is increased by 10 dB, and the coupling stays the same. Thus, the simple but incorrect definition of directivity as isolation/coupling would yield an increase of 10 dB.

In fact, a better way of looking at directivity is the ability of the power at the coupled port to represent a change in reflection at the test port. Again considering Figure 1.34, if a signal of 0 dBm is injected into the input port and a full reflection (an open or short) is applied to the test port, the coupled port will show a power of about −30 dB (10 dB loss, plus a full reflection, plus 20 dB coupling; here the isolation term is ignored for the moment). If a load is applied to the test port, the signal at the input sees a 10 dB loss and 50 dB isolation for a

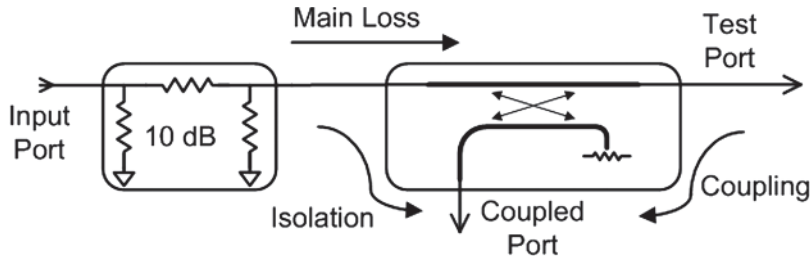

Figure 1.34 The effect of attenuation at the input of a coupler.

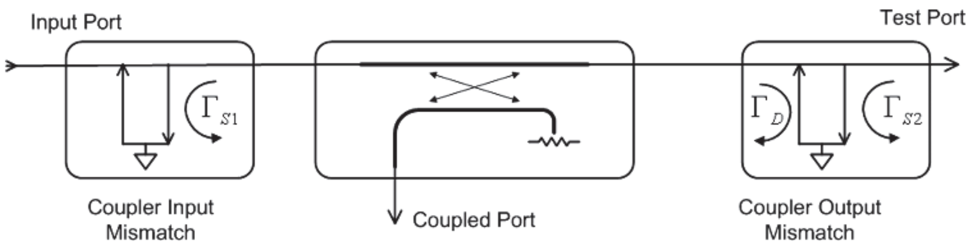

Figure 1.35 Coupler with mismatch after the test port flow graph.

value at the coupled port of −60 dBm. The difference between the open and the load is 30 dB; hence, the directivity is 30 dB, and adding the pad at the input has no effect.

In practice, the output match of a directional-coupler is critical, and the test port mismatch can dominate the directivity. This signal flow is demonstrated in Figure 1.35. Mismatch at the output of the directional-coupler affects directivity on a one-for-one basis. This mismatch is combined with the coupler input mismatch to create the overall source-match. The source-match affects the power measured at the coupled port when measuring large reflections at the test port. This "source mismatch" causes some reflected signal from the test port to re-reflect from the input port, reflect a second time off the test port termination, and add or subtract to the main reflection, causing error in the coupled port power.

However, the output mismatch is a direct error and causes reflection back into the coupler, thereby adding directly to the coupler directivity.

1.11 Circulators and Isolators

While most passive components are linear and bilateral (that is, the forward loss equals the reverse loss), a particular class of devices based on the ferromagnetic effect doesn't follow this rule. These devices comprise circulators and isolators. A circulator is a 3-port device, with low loss in one direction between ports, say from port 1 to 2, port 2 to 3, and port 3 to 1. But it has high loss (called the *isolation*) in the reverse direction, from port 2 to 1, 3 to 2, or 3 to 1. An isolator is a special case of a circulator with a good load applied to port 3, such that it becomes a 2-port device. Circulators pose a particular measurement difficulty as the isolation

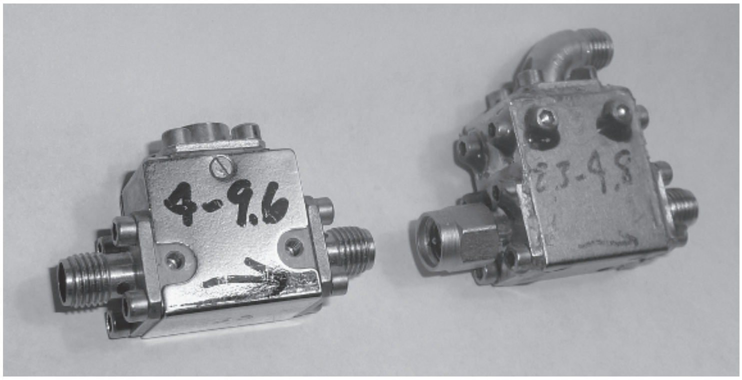

Figure 1.36 Isolator (left) and circulator (right).

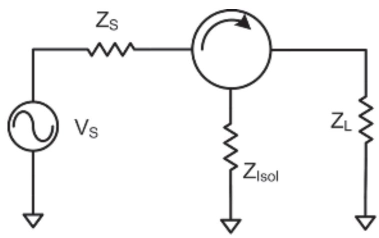

Figure 1.37 Schematic representation of a circulator.

between ports depends upon the match applied to a third port. Thus, for good measurement quality, the isolation measurement requires a good effective-match on the ports.

Further, circulators are often tuned by magnetizing a permanent magnet attached to the circulator, and it's desired that the measurement system can determine the isolation of all three ports in a single connection step, and with good speed. Thus, multiport (more than two ports) systems were developed to simplify the connections, and multiport calibration techniques were developed to satisfy the need for high-quality correction.

Even though they are passive devices, circulators and isolators are sometimes tested for their high-power response, such as compression and IMD. The ferromagnetic effect has hysteresis properties that can produce IMD and compression when driven with sufficiently high powers. Figure 1.36 shows an isolator (left) and a circulator (right). Figure 1.37 shows the signal flow for a circulator. In the figure, the isolator on the left has an internal load mounted at the top, and the circulator on the right has three ports, with an SMA connector in place of the isolator load.

1.12 Antennas

As the air interface for all communications systems, antenna performance is the first (in a receiver) and the last (in a transmitter) characteristic that affects the overall system performance. An antenna can be small and simple, such as a whip antenna found on a handset, or quite complicated such as those found in phased-array radar systems. Antennas have two key attributes: reflection and gain pattern.

Antenna reflection is essentially a measure of the power transfer efficiency from the transmitter to the over-the-air signal. Ideally, the antenna should be impedance matched to the transmitter's output impedance. In fact, it is typically the case that the antenna is matched to some reference impedance, typically $50\,\Omega$, while the transmitter is likewise matched to the same reference impedance. This implies that while the two may be matched, in many cases they can be exactly mismatched if the phase of the antenna mismatch is not the conjugate of the phase of the transmitter's mismatch. The tighter the mismatch specification is for each, the less variation in transmitter power one sees when phasing causes the two mismatches to be on opposite sides of the reference impedance.

Further, simple antennas are matched to a rather narrow range of frequencies, and it is a significant aspect of antenna design to extend the impedance match across a broad range of frequencies. One common form is a bi-conical antenna, often found for use in testing the radiated emissions from electrical components. On the other end of the spectrum is the desire for a narrowband antenna to have a low return loss over a small frequency range to minimize reflected power back to the high-power transmitter.

Antenna gain, or antenna gain pattern, describes the efficiency of an antenna in radiating into the desired direction (or beam) relative to a theoretical omni-directional antenna, often referred to as an *isotropic radiator*. This figure of merit is known as dBi, or decibels relative to an isotropic antenna.

Antenna pattern measurements are the measurement of the antenna radiation pattern, typically plotted as a contour of constant dBi on a polar plot, where the polar angle is relative to the main beam or "bore-sight" of the antenna. Antenna pattern measurements can range from simple gain measurements on an antenna on a turntable to near-field probing of complex multi-element phased array structure. While these complex measurements are beyond the scope of the book, many aspects of antenna return loss measurement, including techniques to improve these measurements, will be covered.

1.13 PC Board Components

While a wide-ranging topic, the measurement of passive PC board components is focused on the measurement of surface mount technology (SMT) resistors, SMT capacitors, and SMT inductors. These components comprise the majority of passive elements used in radio circuits and also create some of the most undesirable side effects in circuits because of the nature of their parasitic elements. Here is a review of the models of these elements; during measurement, the difficulty is in understanding the relative importance of aspects of these models and extracting the values of the model elements.

1.13.1 SMT Resistors

Resistors are perhaps the simplest of electronic elements to consider, and Ohms law is often the first lesson of an electronic text.

$$R = \frac{V}{I} \tag{1.91}$$

However, the model of an RF resistor becomes much more complex as frequencies rise and distributed effects and parasitic elements become dominant. In this discussion the focus will be on surface-mount PC board components, as they are used almost exclusively today in modern circuits. Thin film or thick film hybrid resistors have similar effects, and although the parasitic and distributed effects tend to hold off until higher frequencies, much of this discussion applies to them as well.

A good model for a resistor consists of a resistive value in series with an inductance, both shunted by a capacitance. This is a reasonable model for an SMT resistor in isolation, but the values and effects of the model are modified greatly by the mounting scheme of the component. For example, if it is mounted in series with a microstrip transmission line and the impedance is such that the resistor is much narrower than the transmission line, then this model works well for predicting circuit behavior; on the other hand, if it is mounted on a narrow line, then the contact pads will provide additional shunt capacitance to ground, and the model must include some element to account for this effect. At lower frequencies, some shunt capacitance will do well, but at higher frequencies, a length of low impedance transmission line might be a better choice.

A resistor used in shunt mode to ground can have an entirely different model when it comes to parasitic effects from that of a series resistance. While the RF value of the resistive element may stay almost the same as the series value (close to the DC value), the effective inductance can be substantially higher as the inductance of the ground via adds to that of the resistor in a microstrip configuration. A larger pad on the ground via, surprisingly, can add even more effective inductance as it resonates with the inductance of the via to increase the apparent inductance of the pair. Meanwhile, the shunt capacitance of the resistor may be absorbed in the transmission line width. Figure 1.38 shows a model for a resistor mounted in the series and shunt configurations. Measurement examples to illustrate extracting these values will be shown in Chapter 11.

In many instances, one of the two parasitic elements will dominate the model for first-order high-frequency effects. In fact, one can use some simple calculations to estimate a rough order of magnitude for these parasitic elements. Take, for example, an 0603 resistor, which has dimensions of approximately 0.6 mm width, 0.4 mm height (considering some excess plating, and some edge effect), and 0.76 mm length. If one considers the contact of the resistor wrapped around the body, one might reasonably divide the effective length by 3, to about 0.25 mm. Remembering that SMT resistors are often constructed on ceramic substrates, with a relative dielectric constant of about 10, then the capacitance can be computed as

$$C = \varepsilon_r \varepsilon_0 \frac{W \cdot H}{L} = 10 \cdot 8.85 \times 10^{-15} (F/mm) \cdot \frac{0.6 \cdot 0.4}{.25} = 0.085 pF \tag{1.92}$$

The actual value may be substantially greater or less depending upon the exact attributes of the electrodes, but this gives a starting estimate. For inductance one can look to the formula

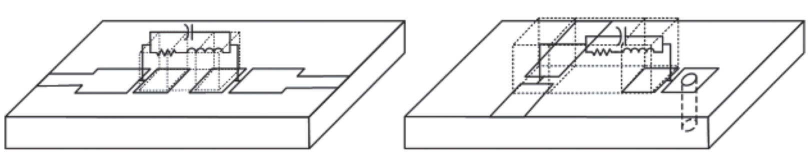

Figure 1.38 Models for a series resistor (left) and shunt resistor (right).

of a transmission line, and assuming that the resistor is mounted on a narrow line, such that its impedance is high, the inductance is computed as

$$l = \mu_0 \cdot L = 4\pi 10^{-10}(H/mm) \cdot 0.8mm = 0.8nH \tag{1.93}$$

Thus, from the model one can compute the values of resistance for which the inductance or capacitive term dominates, at some frequency. For example, at 3 GHz, the inductance has a value of about 15 Ω reactance in series with the resistive element; the capacitance has a value of about 1250 Ω reactive in shunt. At 50 Ω, the inductance value dominates, at 300 Ω, the capacitive value is the dominant parasitic effect. For low values of resistance and high frequency, the inductance becomes dominant, and the series impedance is larger than expected, causing the loss through the resistance to be larger than expected because of this effect. At high values of resistance, the parasitic capacitance reduces the series impedance, and the expected loss through the resistor is less than expected. The values change with the physical size of the component; thus, cross-over points differ in resistance and frequency, but with similar effects. This can be used to advantage as there exists a crossover point where the inductive and capacitive effects cancel somewhat, and the resistor behaves in a more ideal way, to higher frequencies, than for values above or below the crossover value. Using this value of resistance can, in series or parallel arrangements, provide a range of resistances that avoid parasitic effects until higher frequencies. For these values, a 50 Ω resistor terminated to ground will have about −18 dB return loss at 3 GHz; however, two 100 Ω resistors terminated to ground will have about −36 dB return loss, thus providing a better RF resistance of 50 Ω than a single resistor. Thus, characterization of the parasitic effects, and proper compensation, can allow use of SMT parts to much higher than expected frequencies (Dunsmore 1988). Figure 1.39 shows the effective impedance of a single 50 Ω SMT resistor and two 100 Ω SMT resistors in parallel, when used as a 50 Ω load. This effect also occurs for SMT inductors and capacitors.

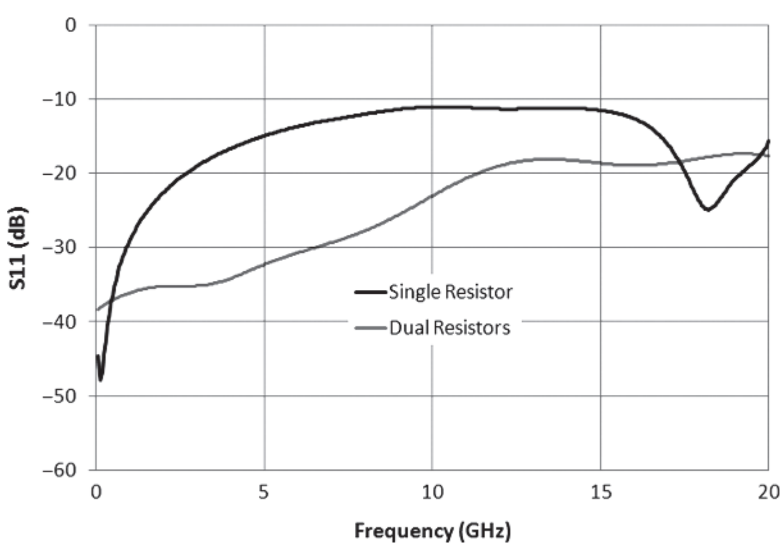

Figure 1.39 Input match of a single SMT resistor and two in parallel.

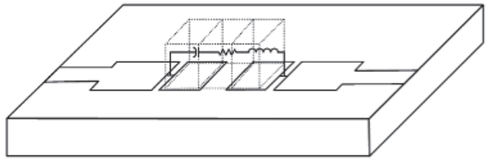

Figure 1.40 Model of an SMT capacitor.

1.13.2 SMT Capacitors

SMT capacitors have a different model from resistors. To a first order, their parasitic effects tend to be all in series, as shown in the model of Figure 1.40. The series inductance is due to primarily to the package size and is similar to that of a resistor. The series resistance is due to the manufacturing characteristics of the capacitor and thus cannot be easily estimated. If an SMT capacitor is used in a resonant structure, this resistance will have the principal effect on the Q of the resonator. However, its effect is typically small in most wideband applications, where the capacitor is used as a series DC blocking capacitor or a shunt RF bypass capacitor. This is because the series inductance will dominate the series resistance in these use cases and cause the impedance of the capacitor to rise with frequency (rather than go to zero). At high frequencies, there may also exist a parasitic shunt capacitance across the entire package, which may cause the impedance to fall again.

The case where the series resistance is of consequence is when the capacitor is used in a turned circuit, where the package inductance may be subsumed in the resonating inductor and thus at resonance the series resistance adds to a degradation of the Q of the capacitor. With careful design, the capacitance value may be compensated for by the including the effects of the series inductance; this effect is to make the capacitor look larger than its prescribed value. In fact, where the reactance of the parasitic inductance equals the reactance of the capacitance, the effective value of capacitance goes to infinite and the series impedance becomes just the parasitic resistance. So, for characterizing capacitors for use in tuned circuits, one must really assess their value near the frequency on which they will be having the most effect on a circuit. Consider a one-pole filter, where the cutoff of the filter starts to occur when the reactance of the capacitor reaches 50 Ω. In many cases, the inductance is quite significant and already altering the effective value of the capacitor. Thus, it is important to evaluate the effective capacitance near this point. A good rule of thumb is to evaluate a capacitor where the reactance is j50 Ω.

A further characteristic of capacitors that is significant is the internal assembly structure. Capacitors are typically formed by a set of interleaved parallel plates with alternate plates connected at each end to the terminals. The plates can be parallel to or vertical to the PC board. For some cases, the capacitor body itself can form a dielectric resonator at high frequency, but below that the capacitor can act as a single, large conductive block on a PC board trace, typically resulting in a model that might best be considered a transmission line of somewhat lower impedance than the mounting line.

Capacitors used as bypass capacitors have an additional parasitic effect from the series inductance of the ground via, and from the pad above the ground via.

1.13.3 SMT Inductors

Inductors are perhaps the most complicated of the simple passive components. Because they are constructed of coils of very fine wire, sometimes multiple layers of coils, their parasitic elements are greatly affected by the details of their construction. Some inductors have the axis of the coil parallel to the PC board, and some are wound with the axis perpendicular. In both cases, the model for the inductor is essentially the same as the resistor, as shown Figure 1.39, but with the value of the series inductance equal to the DC value of the inductor, and the series resistance equal to the DC resistance. Inductors, because of the nature of their construction, have very large relative parasitic capacitances. In cases where an inductor is used for a bias element (relying on its impedance to be high at high frequencies) one often finds that the parasitic capacitance will become the main effect over the band of interest. Thus, in many cases the value of inductance used is carefully selected based on the overall effective inductance and sometimes utilizes the shunt capacitance to provide a high impedance at a particular frequency of interest. It may quite difficult to make a single inductor provide good RF performance over a wide band.

When inductors are used as elements in filters, the parasitic capacitance can often have significant effects for use in band-pass filters, and the inductance must be evaluated for each use to find the effective value considering the parasitic capacitance.

A common figure of merit for inductors is the self-resonant frequency (SRF), above which they act more like a capacitor (impedance goes lower with increasing frequency) than an inductor. The value of the SRF can be estimated in one way by looking at the length of the wire used in making the inductor. The SRF will be less than the frequency for which the wire is one-quarter wavelength.

1.13.4 PC Board Vias

The PC board via is perhaps the most common PC board component, and often the most overlooked. The effect of a via depends greatly upon how it is structured in the circuit. A single via to ground in the center of a transmission line appears as almost a pure inductance. However, a via between RF traces can have aspects of inductance and some parasitic capacitance (due to pads around the via) that can cancel, in part or all, the inductive effect. When a via is used in a mounting pad for a shunt element, such as a resistor used as a load, or a bypass capacitor, the mounting pad and via form a resonant structure such that the size of the mounting pad can increase the effective impedance of the via. Further, several vias are often used in parallel to ground devices, sometimes to lower their effective inductance and sometimes to provide greater heat sinking of an active device. Putting vias in parallel does lower their effective inductance, but not in a simple way. Rather than halving the inductance, mutual inductance between vias means that the value of effective inductance doesn’t reduce as expected. For example, putting two 100 Ω resistors at the end of a line to ground, placed in parallel, may show much larger inductive effect the same two 100 Ω resistors place in a T pattern, where the ground vias are separated and the mutual inductance is less.

1.14 Active Microwave Components

With a few exceptions, passive components follow some fundamental rules that greatly simplify their characterization; principally, they are linear, so their characterization doesn't depend on the power of the signal used to characterize them, but only on the frequency. Active components, on the other hand, are sensitive to power, and their responses to both frequency stimulus and power stimulus are important. Often, passive components are operated well below any power level that causes a change in their response, but more and more active components are being driven into higher-power operation to optimize their efficiency.

1.14.1 Linear and Non-linear

In the measurement sense, one definition of linear devices is that they are devices in which the output power is a linear function of the input power. If the input power is doubled, the output power is doubled. Almost all passive devices follow this rule, and many active devices as well. An alternative definition of linear is one for which only frequencies that are available at the input appear at the output. In practice, the first definition is more useful for system response. Some important characteristics of active components are discussed next.

1.14.2 Amplifiers: System, Low-Noise, High Power

1.14.2.1 System Amplifiers

System amplifiers are simply gain blocks used to boost signal levels in a system, while providing reverse isolation. They can have higher noise figures than LNAs as they are used in signal paths where the signal is well above the noise floor. They often follow an LNA stage, frequently after some pre-filtering. They are also often used in the frequency converters as a local oscillator (LO) amplifier to isolate the RF signal from leaking out the LO port, or as an isolation amplifier to prevent LO leakage out the RF port. These tend to be broad band amplifiers, with good input and output match, emulating an idealized gain block. The important figures of merit for such amplifiers are gain (S21), input and output match (S11, S22), and isolation (S12). Occasionally, directivity of an amplifier is defined as isolation (a positive number in dB) minus the gain (in dB), or S12/S21. It is a measure of the effects of a load apparent at the input of the amplifier, or how the output impedance is affected by the source impedance (Mini-circuits n.d.) and is important in cases where other system components have a poor or unstable match. Since these amplifiers have wide bandwidths, it is important that they have good stability as they can have a variety of load impedances applied. Other figures of merit for system amplifiers can include gain flatness (deviation of the gain from nominal value), 1 dB compression point (the power at which the gain drops by 1 dB), harmonic distortion, and two-tone third-order IM, sometimes expressed as third-order intercept point (see Section 1.3).

1.14.2.2 Low-Noise Amplifiers

Low-noise amplifiers are found in the front end of communications systems and are particularly designed to provide signal gain without a lot of added noise power. The key figure of merit is noise figure, along with gain. But, in a system sense, the noise parameters (see Section 1.3) are quite important as they describe the way in which the noise figure changes with source impedance. LNAs are used in low-power applications, and their 1 dB compression point is not a key spec; however, distortion can still be a limiting factor in their use, so a common specification is the input referred intercept point. A key trade-off made in LNAs is between lowest noise figure and good input match. The source impedance where the LNA provides the lowest noise may not be the same as the system impedance, so a key design task for LNAs is to optimize this trade-off.

1.14.2.3 Power Amplifiers

Many of the figures of merit for power amplifiers are the same as system and LNAs, but with an emphasis on power handling. In addition, the efficiency of the amplifier is one of the key specifications that one finds primarily with power amplifiers, implying that the DC drive voltage and current must also be characterized. Because power amplifiers are often used with pulsed RF stimulus, the pulse characteristics, such as pulse profile including pulse amplitude and phase droop, are key parameters.

Power amplifiers are often driven into a non-linear region, so the common linear S-parameters may not apply well to predict matching. Therefore, load-pull characterization is often performed on power amplifiers. Gain compression and output-referred intercept point are common for power amplifiers. Some amplifier designs such as traveling-wave-tube amplifiers (TWT) have a characteristic that causes the output power to reach a maximum and then decrease with increasing drive power, and the point of maximum power is called *saturation*. Gain at rated output power is another form of a compression measurement where rather than specifying a power for which the gain is reduced by 1 dB, it specifies a fixed output power at which the gain is measured.

Power amplifiers are often specified for their distortion characteristics including IMD and harmonic content. In the case of modulated drive signals, other related figures of merit are ACPR and adjacent channel power level (ACPL). A figure of merit that combines many others is EVM, which is influenced by a combination of compression, flatness, and inter-modulation distortion among other effects.

1.14.3 Mixers and Frequency Converters

Another major class of components is mixers and frequency converters. Mixers convert an RF signal to an IF frequency (aka down-converters) or IF frequencies to RF (up-converters) through the use of a third signal, known as the *local oscillator* (LO). Typically, the output or input that is lower in frequency is called the IF (for *intermediate frequency*). The LO provided to the mixer drives some non-linear aspect of the circuit, typically diodes or transistors that are switched on and off at the LO rate. Using the second definition, frequency converters and mixers would not be considered linear. In fact, in their normal operation, they are linear (under the first definition) with respect the desired information signal, and ideally the frequency conversion does not change the linearity of the input/output transfer function.

The input is transferred to the output through this time-varying conduction, and the output signal includes the sum and difference of the input and the LO. Mixers tend to be fundamental building blocks that are combined with filters and amplifiers and sometimes other mixers to form frequency converters. In practice, frequency converters are typically filtered at the input to prevent unwanted signals from mixing with the LO and creating a signal in the output band, and they are typically filtered at the output to eliminate one of the plus or minus products of the mixing process. Some converters, known as *image reject mixers*, have circuitry that suppresses the unwanted product, sometimes called the *sideband*, without filtering. These are typically created by two mixers driven with RF and LO signals phased such that the output signals of the desired sideband are in phase and added together to produce a higher output, and the undesired sideband is out of phase; thus they cancel and produce a smaller output. Monolithic microwave integrated circuits (MMICs) sometimes blur the line between converters and mixers as they may contain several amplifier and mixing stages, but they typically don't contain filtering.

Mixers have fundamental parameters that include conversion loss (or gain in the case of MMICs with amplifiers), isolation (of which there are 12 varieties, to be discussed in Chapter 7), compression level, noise figure (which for passive mixers is typically just the conversion loss), input and output match, and most of the other parameters found for amplifiers.

Mixers are sometimes referred to as passive mixers or active mixers. Passive mixers don't contain amplification circuits, are typically constructed of diodes (with a ring or star configuration being most common), and have baluns in some paths to provide improved isolation and reduced higher-order products. Higher-order mixing products, also called *spurious* mixing products or *spurs*, refer to signals other than the simple sum and difference of the input and LO frequencies. They are often referred to by the order of harmonic related to creating them; for example, a 2:1 spur is created at the frequency of two times the LO plus and minus one times the input. The level of the mixer spurious products, which are sometimes called *mixer intermod products* (even though the two tones are not applied to the input), change with respect to the RF drive signal at the rate of the order of the RF portion of the spur. For example, a 2:1 spur will increase 1 dB in power for each dB in power the RF signal increases. However, since the LO is creating the non-linearity in the mixer, as well as the spurious signal, it is difficult to predict how the spurious power will change with respect to LO power. In many cases, driving the LO power higher produces a higher spurious signal, as the relative magnitude of the RF signal to the LO power is reduced; in other cases, the transfer impedance of the non-linear element becomes more consistent across the RF drive level. The spurious higher-order products of a mixer are sometimes defined as a spur table, which shows the dBc values of higher-order products relative to the desired output. Chapter 7 has more details on measuring these behaviors.

Mixers with baluns can suppress some of the higher-order products, with baluns on the RF port suppressing products that have even-order LO spurious, and baluns on the LO port suppressing products that have even-order RF spurious. These are called *single-balance mixers*, and it is typical that the LO port is balanced. Double-balanced mixers have baluns on the RF or IF ports. Refer to Chapter 7 for more details on mixer configurations. Triple-balanced mixers are usually comprised of a pair of double-balanced mixers, adding a balun to the IF port. Their main advantage is to divide the RF signal power between the two diode quads, lowering the RF relative to the LO, whereby the spurious signals created will be lower; then the outputs are combined to recover the power. This provides a mixer with the same conversion loss and lower spurious products at the same output power level. The disadvantage is that since the LO drive is also divided, higher LO drive power is needed to achieve the same linearity for each diode quad.

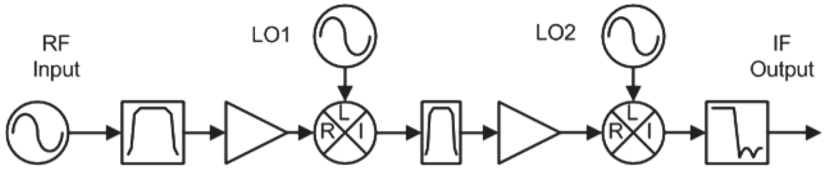

Figure 1.41 Dual-LO frequency converter.

The creation of spurious mixer products is a key aspect of system design, with the goal of eliminating spurious signals from the IF output. Unfortunately, some frequency plans are such that the spurious products must fall into the band of interest. In such cases, system designers move to multiple conversion stages to create a first stage, which produces an output free of spurious signals over the range of input signals of interest, and then has a second conversion stage that produces the frequency of interest at the output. This multiple conversion or "dual-LO" system is typically called a *frequency converter* and often contains additional filtering and amplification, as shown in Figure 1.41.

Because of the multiple components used, the frequency response of converters often has gain ripple and phase ripple, which can distort the information signal. Key figures of merit for converters are gain flatness, group delay flatness, and the related phase flatness, which is also known as deviation from linear phase, and represents residual ripple after fitting the phase data to a straight line. Modern systems employ equalization techniques that can remove some of the flatness effects provided they follow simple curvatures; as such, another specification found on converters is deviation from parabolic phase, which is the residual ripple in the phase data when it is fitted to a second order curve.

Mixers often have quite poor input or output match because of the switching nature of their operation, so their effect on system flatness when assembled into a converter can be quite dramatic. Until recently, it was difficult to predict the effects of output load of a mixer on its input match, but the mathematical tools to model mixers as system elements have been developed (Williams et al. 2005) that describe these relationships. Mixers that produce an output that is the sum of the input and output signals are relatively simple to describe, but mixers that produce an output that is the difference between input and LO have a more complicated behavior in the case where the input frequency is less the LO frequency. These are sometimes called *image mixers*, and their unusual characteristic is that as the input frequency goes up, the output frequency goes down; this also applies to phase: a negative phase shift of the input signal results in a positive phase shift of the output signal. How these special cases affect the system performance will be described in more precise mathematical terms in Chapter 7.

1.14.4 Frequency Multiplier and Limiters and Dividers

Mixers are not the only way to create new frequencies at the output; frequency multipliers are also used to generate high-frequency signals, particularly when creating mm-wave sources. Frequency multipliers produce harmonics by changing a sine-wave input signal into non-linear wave. The basic doubler is a half-wave or full-wave rectifier, such as a diode bridge. A pair of back-to-back diodes turns a sine wave into a square wave, which is rich in odd harmonic content. This is essentially the same as a limiter.

The key figure of merits of a multiplier is the conversion loss from fundamental drive to the desired harmonic. Other important characteristics are fundamental feed-through and higher-order harmonics.

Limiters have the key characteristic of maximum output power; that is the power at which they limit. Also important is the onset of limiting and the compression point. Ideal limiters are linear until the onset of limiting, and then they effectively clip the output voltage above that level.

Other multiplier types are step-recovery diodes, and non-linear transmission lines that, when driven with a sine wave, effectively "snap" on to produce a sharp edge. Depending upon the design, the on-time can be short, which produces an output rich in harmonics. Some digital circuits can also be used to create narrow pulses from a sine-wave input as a pulse generator. Such a pulse will also be rich in harmonics.

One aspect of a multiplier that is not easily discerned is the group delay through it. That is because for some change in input frequency one will see a multiplied change in output frequency. FM that passes through a doubler will have the same rate as the original FM, but twice the deviation. For this reason, doublers or multipliers are seldom used in the signal or communication path of RF or microwave systems but can be used in the base carrier paths and in the LOs of many systems.

1.14.4.1 Frequency Dividers

Frequency dividers provide for a lower value of frequency than the input frequency. Like multipliers they are highly non-linear and often produce square-wave outputs. Some key specifications of dividers are minimum and maximum input power to ensure proper operation of the divider, output power and harmonics, and additive phase noise. Typically, for each divide by two stage, the phase noise is reduced by 6 dB. But noise or jitter in the divide circuitry can add noise to the signal; the added phase noise at the output relative to the phase noise at the input is called *additive phase noise*. This is also a concern with mixers, where the LO phase noise can be added to the output signal, and to a lesser extent amplifiers.

1.14.5 Oscillators

Oscillators in some ways represent the most non-linear of electrical circuits with frequencies created at the output with no input (other than noise). Oscillators have a wide variety of characteristics that are important to characterize, including output frequency, output power, harmonics, phase noise, frequency pushing (change in frequency with change in DC power), frequency pulling (change in frequency due to change in load impedance), and output match.

Voltage-controlled oscillators (VCOs) have the additional ability to control the frequency of the output due to a voltage change at the input. The voltage-to-frequency control factor is a key attribute of a VCO. A related microwave component is a yttrium-iron-garnet (YIG) oscillator, which uses a spherical YIG resonator as the frequency control element of the YIG-tuned-oscillator (YTO). The YIG resonator has the characteristic that the resonant frequency changes with magnetic field. YTOs have wide tuning bandwidth (up to 10:1) and low phase noise. Tuning is performed by changing the current in an electromagnet but can be very low bandwidth due to the large inductance of the magnet. YTOs often have a second, lower inductance coil (called the *FM coil*), which provides small change to frequency but with high bandwidth.

As the focus of this book is stimulus/response measurements, the measurement of oscillators will not be covered.

1.15 Measurement Instrumentation

1.15.1 Power Meters

Perhaps the simplest and most common of microwave instruments is the power meter. It consists of a sensor, which absorbs or detects RF power and converts it to a DC signal, and measurement circuitry, which accurately measures this DC signal and applies correction and calibration factors to it, to produce a reading of the RF power level. Power meters come in a variety of forms and complexities, some of which are noted next.

1.15.1.1 Calorimeters

Often considered the most accurate and traceable of power measurement systems, the calorimeter consists of a thermally isolated RF load that absorbs the RF energy. This load is kept in a heat exchanger, and a thermopile is used to sense the change in temperature. Since the fundamental measurement is temperature, the traceability of these systems to fundamental SI units is good. These systems can handle large power but are slow to respond, are heavy, and are typically not used by practicing RF engineers, except in special cases.

1.15.1.2 RF Bolometers and Thermistor

An RF bolometer or thermistor is a system where the RF measuring element is a thermally sensitive resistor used as part of a DC bridge system. The DC bridge is electrically balanced, and when an RF signal is applied to the bolometer element, the element heats and its DC resistance changes. The DC bridge is nulled using an offset voltage, and the measure of the offset voltage can be related directly to the power absorbed by the bolometer. The key aspect of the bolometer is that it is equally sensitive to RF or DC power; thus, a precision DC source can be used to produce a known power at the bolometer, and the balancing circuit is thus calibrated relative to the DC power absorbed. The heating effect of the RF power produces the same offsets as the DC power and thus is easily calibrated. Bolometers have a relatively small dynamic range (the range of input powers over which they operate) but have linearity (the ability to correctly measure differences in input power) derived from a bridge circuit using DC substitution. Typically, bolometers are found only in precision metrology laboratories and are not in common use among RF engineers.

1.15.1.3 RF Thermocouples

Until recently, RF thermocouples were the most common type of power sensor used. These thermocouples convert heating directly into a DC voltage and, because their small size and thus small thermal mass, are much faster responding and have a larger dynamic range than either thermistors or calorimeters. As with other sensors, these require calibration with a precision source but are typically DC blocked, so the source must be a low-frequency AC source. These sensors are commonly used throughout the RF industry but have the detriment of being somewhat slow-responding (with response times in the several to tens of millisecond range), but they are extremely linear and relatively non-responsive to harmonics. That is, harmonic power will be detected as an RMS error of the power of desired signal. Since harmonics 20 dBc or lower represent less than 1% of the power of the main signal, the error due to harmonics is quite low.

1.15.1.4 Diode Detectors

For modern power meter applications, the diode or multi-diode power sensor is often the preferred choice. These sensors employ one or more diodes that rectify the RF signal and produce an equivalent DC signal. Occasionally, the DC signal is “chopped” or modified in such a way as to produce a square wave to the measurement portion of the power meter, typically a precision analog-to-digital converter (ADC). Chopping the signal helps compensate for DC offsets in the ADC input.

Older diode detectors used only a single diode, and the top 20 dB of the detector range was often described as the “linear” range; below that range, the diode would operate in “square-law” mode where the output voltage would be a function of the square of the input RF signal. In the low-power range, the output voltage would be linearly related to the square of the input voltage of the RF signal, thus be linearly proportional to the detected power. In such a region, they operated almost as well as the thermistor sensors but with much faster speeds and much wider dynamic range. At the top of their measurement range, in the linear region, the output circuitry and measurement algorithms are adjusted to compensate for the change to the linear mode of operation. However, in the linear mode, the power in the harmonics has a much greater effect, and a 20 dBc harmonic signal can have up to a 10% change in the measured power of the fundamental, even though it contains only 1% of the power. This is due to the peaking effect that the harmonic can have on the RF voltage. Out of the square-law region (also known as the *linear region*, which is in fact where the power meter is not as linear in the usual sense of the word), the power meter may not give accurate readings for complex modulated signal or signal with high harmonic content or high peak-to-average envelope power.

More modern diode sensors use a multitude (two or more) of embedded diode elements, some of which are padded with larger attenuation to allow them to operate at higher powers and still be in the square-law region. Complex algorithms in the power meter instrumentation detect when the power from one sensor exceeds the square-law region and change to take their readings for power from one of the attenuated diodes. This extends the useful range of the power sensor over more common older diode sensors.

1.15.2 Signal Sources

1.15.2.1 Analog Sources

While not a measurement instrument in their own right, signal sources or signal synthesizers, or simply sources, are used as accessory equipment in a variety of measurement tasks. They can provide CW signals in place of a mixer LO or provide an input signal to an amplifier or filter. These are typically called *analog sources*, and their key attributes are frequency range, output power range (minimum and maximum), phase noise and spectral purity, and frequency switching speed.

While the first two attributes are obvious, the phase noise and spectral purity are key attributes when making measurements close to the carrier such as IMD measurements or when making other distortion measurements such as harmonics.

Switching speed becomes important in automated test systems (ATSs) when using the source as a swept frequency stimulus. Commonly, stand-alone signal sources make a trade-off between lower-phase noise and slower switching speeds.

1.15.2.2 Vector Sources

Another class of signal sources are vector signal generators, which have an internal I/Q modulator that allows an almost infinite variety of signals to be created. Some of these vector sources (also called *digital sources* because they can create signals using digital modulation techniques) have built-in arbitrary-waveform generators (AWGs, or *arbs* in the vernacular), while others have broadband I/Q inputs to allow external AWGs to drive their vector modulators directly.

With vector sources, the AWGs can be used to create a wide variety of signals including extremely fast switching CW sources (within the bandwidth of the I/Q modulator), two-tone or multitone signals, pseudo-random noise waveforms, and complex modulated signals following the formats used in digital communications and cellular phones.

Some key attributes of vector sources are the modulation bandwidth of the I/Q inputs, the modulation bandwidth or speed of the arbitrary waveform generator (if it is built-in), the memory of the AWG (which affects the length of signals that can be created), and the I/Q fidelity or linearity of the modulator. This linearity limits the ability of the vector source to produce clean signals. For example, a two-tone signal can be created by doing a double sideband suppressed carrier modulation, but if there is imbalance or non-linearity in the modulator, there will be carrier leakage between the two tones.

The output power amplifiers of vector sources are important as their distortion will directly affect the modulated signal, causing TOI, and the spectral spreading of modulated signals.

1.15.3 Spectrum Analyzers

A spectrum analyzer (SA) is a specialized type of receiver, which displays the power of a signal on the y-axis versus the frequency of the signal on the x-axis. As such, it could be considered a frequency-sensitive power meter.

The key attributes of a spectrum analyzer are its displayed average noise level (DANL) and its maximum input power. The maximum input is set by the compression of the input mixer in the SA and can be increased by adding input attenuation. However, adding attenuation degrades the DANL by an equal amount. A further limitation in measuring signals, for example, TOI, is the self-generated distortion of the input mixer, which will generate TOI signals at the same frequencies as that as the TOI from the signal under test. The data sheet for an SA will typically specify the distortion in dBc relative to some input level at the mixer.

This, coupled with the noise floor, will set the measurement range of the SA. Lower-resolution bandwidth will lower the noise floor at the cost of speed of measurement. Similar effects are present for the measurement of harmonics.

Another key attribute of a spectrum analyzer is its frequency flatness and power linearity specifications. Flatness specifications of a spectrum analyzer are usually quite large, as much as ±2.0 dB for 26 GHz microwave version, although typical performance is much better, and this flatness can be compensated for with an amplitude calibration. Microwave and mm-wave systems can be even worse. The large value for frequency response comes from the interactions of the pre-selector (which is usually a swept YIG filter) and the first converter in the SA. To a first order, these are stable and can be corrected for, but there will still be a residual flatness error even after calibration related to the post-tuning drift of the pre-selector; that is, it does not always tune its peak value to the same frequency for the same settings. Another source of uncorrected error is the mismatch between the SA input and the output match of the signal source being measured. In some cases, this can be a quite high number, up to ±1 dB or more.

As the name implies, the key role of spectrum analyzers is in determining the quality of unknown spectrums. The use of spectrum analyzers in microwave component test applications is primarily as a means to the frequency response or distortion response of a system as a known stimulus is applied.

In recent times, these applications of spectrum analyzers are being augmented by advanced VNAs, which have higher-speed receivers, built-in sources, and advanced calibration capabilities, as well as PXI-based spectrum analyzers. Spectrum analyzers are available in modular (PXI) formats and can be configured as multi-channel spectrum analyzers. Further, some spectrum analysis has moved into the mm-wave range, typically with external down-converters, but some manufacturers are producing spectrum analyzers with internal mm-wave converters to provide broadband (up to 110 GHz) capabilities. These newer implementations may not have hardware preselection for image rejection, but rather use advanced forms of digital image-rejection based on high-speed FFT processing or other methods. More about this appears in Chapter 8.

1.15.4 Vector Signal Analyzers

With the advent of digitally modulated signals for RF and microwave communications, spectrum analyzers have evolved into much more complex systems that include the ability to do wideband de-modulation of these signals. These specialized spectrum analyzers are often called *vector signal analyzers* (VSAs) and play an important role in component tests.

For many active components, a key figure of merit is the distortion that they apply to the vector modulated signal in the form of amplitude or phase error relative to an ideal signal. The composite of all errors over a set of digital symbols is called the *error vector*, and the average magnitude of this error is the error-vector-magnitude (EVM). While EVM is a signal figure of merit, as it is compared to an ideal waveform, the EVM from an amplifier is a combination of the EVM of the input signal and the errors added by the amplifier. From this it is clear that EVM is not a microwave-component parameter; but a related value, residual or added EVM, is and is described as the EVM at the output signal relative to the input signal. In practice, high-quality sources are used to produce the digitally modulated signals, so the input effect on EVM is small, but with higher data rates and wider bandwidths of modulation, these input effects are becoming more important. Thus, there is a need for a multi-channel VSA that can compare input to output signals. A normal SA with a VSA capability does not provide such dual-channel capability, but some manufacturers supply a specialized dual-channel receiver for the VSA, while other implementations of a VSA use a wideband digitizer, or even a digital oscilloscope, to do direct digitization of the modulated signal. To date, up to four simultaneous channels have been reported in such a VSA.

1.15.5 Noise Figure Analyzers

An offshoot of a spectrum analyzer, a noise figure analyzer (NFA) is a specialized test instrument that is designed particularly for measuring noise figure. NFAs started out as specialized spectrum analyzers, with improvements in the quality of the receiver and with electronically switched gains to allow the noise figure of the test equipment to be minimized relative to the signal being measured. Some of the things needed to accomplish this, such as adding high-gain LNAs in front of the first converter, reduced the maximum input power of the instrument so that it was no longer suitable for general-purpose SA applications.

On the other hand, several spectrum analyzer manufacturers have added a noise figure personality to their SA offerings so that there is quite a lot of overlap between the capabilities of the two systems. However, most SA implementations require the use of an added LNA, at least over some of the band. Newer SAs have an IF structure almost as flexible as an NFA to optimize the performance of the system.

All of these systems of NFA utilize the “hot/cold” or “Y-factor” method of measuring noise figure (more about this in Chapter 9) using as an input to the DUT a noise source that can be turned on and off. From careful measurement of the output noise, the gain and noise figure of the DUT can be discerned.

More recently, VNAs have been modified to operate as NFAs, utilizing an entirely different technique called the *cold-noise* method. In this method, the output noise power is measured, along with the gain using the normal VNA measurement of gain, and the noise figure is computed from these values. No noise source is used in the measurement. This has an advantage of being faster (only one noise measurement is needed) but does have the disadvantage of being sensitive to drift in the gain of the VNA noise receiver. The Y-factor method does not depend upon the gain of the NFA receiver, but this advantage is often offset by the fact that the gain measurements of the NFA are sensitive to match errors as are the noise measurements, and these are not compensated for.

The ultimate in noise figure analysis is a noise parameter test system. This system properly accounts for all mismatch effects. Some systems use both a VNA and an NFA to measure the gain and noise power, respectively. All noise parameter systems include an input impedance tuner to characterize the change in noise power versus impedance value. Recently, tuners have been combined with VNA-based NFAs to produce compact, high-speed noise parameter test systems. These newer systems provide the ultimate in speed and accuracy available today.

1.15.6 Network Analyzers

Network analyzers combine the attributes of a source, and a tracking spectrum analyzer, to produce a stimulus/response test system ideally suited to component tests. These systems have been commercialized for more than 40 years and provide some of the highest-quality measurements available today. While there are many distinct manufacturers and architectures, network analyzers broadly fall into two categories: scalar network analyzers (SNAs) or VNAs.

1.15.6.1 Scalar Network Analyzers

These instruments were some of the earliest implementations of stimulus/response testing and often consisted only of a sweeping signal source (sometimes called a *sweeper*) and a diode detector, the output of which was passed through a "log-amplifier" that produced an output proportional the power (in dBm) at the input. This was sent to the y-axis of a display, with the sweep tune-voltage of the sweeper sent to the x-axis, thus producing frequency response trace. Later, the signal from the detector and the sweeper were digitized and displayed on more modern displays with marker readouts and numerical scaling.

Other SNA systems were developed by putting a tracking generator into a spectrum analyzer so that the source signal followed the tuned filter of the SA. This produced a frequency response trace on the SA screen.

SNAs had the attribute of being simple to use, with almost no setup or calibration required. The scalar detectors were designed to be quite flat in frequency response, and a system typically consisted of one at the input and one at the output of a DUT. However, for measurements of input and output match, or impedance, the SNA relied on a high-quality coupler or directional bridge. If there was any cabling, switching, or other test system fixturing between the bridge and the DUT, the composite match of all were measured. There was no additional calibration possible to remove the effects of mismatch. As test systems became more complex and integrated, scalar analyzers started to fall from favor, and there are virtually none sold today by commercial instrument manufacturers.

1.15.6.2 Vector Network Analyzers

For microwave component test, the quintessential instrument is a VNA. These products have been around in a modern form since the mid-1980s, and there are many units from that time still in use today. The modern VNA consists of several key components, all of which contribute to making it the most versatile, as well as most complicated, of test instruments; these are as follows:

RF or Microwave Source: This provides the stimulus signal to the DUT. RF sources in a VNA have several important attributes including frequency range, power range (absolute maximum and minimum powers), automatic-level-control (ALC) range (the range over which power can be changed without changing the internal step attenuators), harmonic and spurious content, and sweep speed. In the most modern analyzers, there may be more than one source, up to one source per port of the VNA. Older VNAs required that the source be connected to the reference channel in some way, as either the receiver was locked to the source (e.g. the HP 8510) or the source was locked to the receiver (e.g. the HP 8753). Modern VNAs, for the most part, have multiple synthesizers so that the source and receiver can be tuned completely independently.

RF test set: In older-model VNAs, the test set was a separate instrument with a port switch (for switching the source from port 1 to port 2), a reference channel splitter, and directional-couplers. The test set provided the signal switching and signal separation to find the incident and reflected waves at each port. Most modern VNAs have the test set integrated with the rest of the components in a single frame, but for some high-power cases, it is still necessary to use external components for the test set.

Receivers: A key attribute of VNAs is the ability to measure the magnitude and phase of the incident and reflected waves at the same instant. This requires sets of phase synchronous receivers, which implies that all the receivers must have a common LO. In older, RF VNAs, the reference channel was common to ports 1 and 2, and the port switch occurred after the reference channel tap. Most modern analyzers have a receiver per port, which is required for some of the more sophisticated calibration algorithms. More about that appears in Chapter 3.

Digitizer: After the receiver converts the RF signals into an IF baseband signal, they pass to a multi-channel phase-synchronous digitizer that provides the detection method. Very old VNAs used analog amplitude and phase detectors, but since at least 1985, all VNAs utilize a fully digital IF. In modern VNAs, the digital IF allows complete flexibility to change IF detection bandwidths, modify gains based on signal conditions, and detect overload conditions. Deep memory on the IF allows complicated signal processing, and sophisticated triggering allows synchronization with pulsed RF and DC measurements.

CPU: The main processor of a VNA used to be custom-built micro-controllers, but most modern VNAs take advantage of Windows™-based processors and provide rich programming environments. These newer instruments essentially contain a PC inside, with custom programming, known as *firmware*, which is designed to maximize the capability of the instruments' intrinsic hardware.

Front Panel: The front panel provides the digital display as well as the normal user interface to the measurement functions. Only the spectrum analyzer comes close to the sophistication of the VNA, and in more modern systems, the VNA essentially contains all the functions of each of the instruments mentioned so far. Thus, its user interface is understandably more complex. Significant research and design effort goes into streamlining the interface, but as the complexity of test functions increases, with more difficult and divergent requirements, it is natural that the user interface of these modern systems can be quite complex.

Rear Panel: Often overlooked, much of the triggering, synchronization, and programming interface is accomplished through rear-panel interface functions. These can include built-in voltage sources, voltmeters, general-purpose input/output (GPIO) busses, pulse generators and pulse gating, as well as LAN interfaces, USB interfaces, and video display outputs.

The detailed operation of a VNA is described in Chapter 2.

Extensions to traditional VNAs allow them to create multiple signals for two-tone measurements and to have very low noise figures for noise figure measurements. But the main attraction of VNAs is calibration. A key attribute is that since they measure the magnitude and phase of waves applied to their ports, they can use mathematical correction to remove the effects of their own impedance mismatch and frequency response in a manner that makes their measurements nearly ideal. The details of VNA calibration are covered in depth in Chapter 3.

Thus, even though there is a wide variety of test equipment available for microwave component measurement, by far the most widely used is the VNA, and while many of the topics of component measurements in this book are extensible to any of the previous instruments, the specific implementation and examples will be illustrated primarily using the VNA, as that has become the predominant component test analyzer in use today.

References

Cameron, R.J., Kudsia, C.M., and Mansour, R.R. *Microwave Filters for Communication Systems: Fundamentals, Design, and Applications*. Hoboken, NJ: Wiley-Interscience. Print.

Cohn, S.B. (1954). Characteristic impedance of the shielded-strip transmission line. *Transactions of the IRE Professional Group on Microwave Theory and Techniques* **2** (2): 52–57.

Collier, R.J. and Skinner, A.D. (2007). *Microwave Measurements*. London: Institution of Engineering and Technology. Print.

Collin, R. (1966). *Foundations for Microwave Engineering*. New York: McGraw-Hill. Print.

Dunsmore, J. (1988). Utilize an ANA to model lumped circuit elements. *Microwaves and RF*: 11.

Freiberger, K., Enzinger, H., and Vogel, C. (2017). A noise power ratio measurement method for accurate estimation of the error vector magnitude. *IEEE Transactions on Microwave Theory and Techniques* **65** (5): 1632–1645.

Hong, J.-S. and Lancaster, M.J. (2001). *Microstrip Filters for RF/Microwave Applications*. New York: Wiley. Print.

Hunter, I.C. (2001). *Theory and Design of Microwave Filters*. London: Institution of Electrical Engineers. Print.

IPC (2004). *IPC-2141A: Design Guide for High-Speed Controlled Impedance Circuit Boards*. Northbrook, IL: IPC Print.

Keysight Technologies. (1968). HP Application Note AN-95-1 [Online]. Available at: http://literature.cdn.keysight.com/litweb/pdf/5952-0918.pdf (Accessed 20 January 2020).

Kurokawa, K. (1965). Power waves and the scattering matrix. *IEEE Transactions on Microwave Theory and Techniques* **13** (2): 194–202.

Magnusson, P. (2001). *Transmission Lines and Wave Propagation*. Boca Raton, FL: CRC Press. Print.

Marks, R.B. and Williams, D.F. (1992). A general waveguide circuit theory. *Journal of Research of the National Institute of Standards and Technology* **97** (5): 535–562.

Mini-Circuits. (n.d.) Amplifier Terms Defined AN-60-038 [Online]. Available at: http://www.minicircuits.com/app/AN60-038.pdf. (Accessed 11 Feb. 2012).

Pozar, D.M. (1990). *Microwave Engineering*. Reading, MA: Addison-Wesley. Print.

Simons, R.N. (2001). *Coplanar Waveguide Circuits Components and Systems*. Wiley. Print.

Sombrin, Jacques, 2011. On the formal identity of EVM and NPR measurement methods: Conditions for identity of error vector magnitude and noise power ratio, Microwave Conference (EuMC), 2011 41st European, IEEE, 337–340.

Wen, C.P. (1969). Coplanar waveguide: a surface strip transmission line suitable for nonreciprocal gyromagnetic device applications. *IEEE Transactions on Microwave Theory and Techniques* **17** (12): 1087–1090.

Williams, D.F., Ndagijimana, F., Remley, K.A. et al. (2005). Scattering-parameter models and representations for microwave mixers. *IEEE Transactions on Microwave Theory and Techniques* **53** (1): 314–321.

Zverev, A.I. (1967). *Handbook of Filter Synthesis*. New York: Wiley. Print.