Misura dei Parametri d'ingresso delle Antenne

Richiami sui Parametri S
L'analizzatore di Rete

13

S-Parameters

13.1 Scattering Parameters

Linear two-port (and multi-port) networks are characterized by a number of equivalent circuit parameters, such as their transfer matrix, impedance matrix, admittance matrix, and scattering matrix. Fig. 13.1.1 shows a typical two-port network.

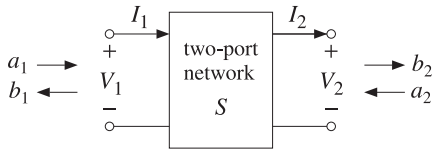


Fig. 13.1.1 Two-port network.

The *transfer matrix*, also known as the ABCD matrix, relates the voltage and current at port 1 to those at port 2, whereas the *impedance matrix* relates the two voltages V_1, V_2 to the two currents I_1, I_2:[†]

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \quad \text{(transfer matrix)}$$
$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ -I_2 \end{bmatrix} \quad \text{(impedance matrix)} \tag{13.1.1}$$

Thus, the transfer and impedance matrices are the 2×2 matrices:

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \qquad Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \tag{13.1.2}$$

The *admittance matrix* is simply the inverse of the impedance matrix, $Y = Z^{-1}$. The *scattering matrix* relates the *outgoing* waves b_1, b_2 to the *incoming* waves a_1, a_2 that are *incident* on the two-port:

[†]In the figure, I_2 flows out of port 2, and hence $-I_2$ flows into it. In the usual convention, both currents I_1, I_2 are taken to flow into their respective ports.

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad \text{(scattering matrix)} \tag{13.1.3}$$

The matrix elements $S_{11}, S_{12}, S_{21}, S_{22}$ are referred to as the *scattering parameters* or the *S-parameters*. The parameters S_{11}, S_{22} have the meaning of reflection coefficients, and S_{21}, S_{12}, the meaning of transmission coefficients.

The many properties and uses of the S-parameters in applications are discussed in [980–1019]. One particularly nice overview is the HP application note AN-95-1 by Anderson [995] and is available on the web [1354].

We have already seen several examples of transfer, impedance, and scattering matrices. Eq. (10.7.6) or (10.7.7) is an example of a transfer matrix and (10.8.1) is the corresponding impedance matrix. The transfer and scattering matrices of multilayer structures, Eqs. (6.6.23) and (6.6.37), are more complicated examples.

The traveling wave variables a_1, b_1 at port 1 and a_2, b_2 at port 2 are defined in terms of V_1, I_1 and V_2, I_2 and a real-valued positive reference impedance Z_0 as follows:

$$\boxed{\begin{aligned} a_1 &= \frac{V_1 + Z_0 I_1}{2\sqrt{Z_0}} \\ b_1 &= \frac{V_1 - Z_0 I_1}{2\sqrt{Z_0}} \end{aligned}} \qquad \boxed{\begin{aligned} a_2 &= \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} \\ b_2 &= \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} \end{aligned}} \quad \text{(traveling waves)} \tag{13.1.4}$$

The definitions at port 2 appear different from those at port 1, but they are really the same if expressed in terms of the incoming current $-I_2$:

$$a_2 = \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} = \frac{V_2 + Z_0(-I_2)}{2\sqrt{Z_0}}$$
$$b_2 = \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} = \frac{V_2 - Z_0(-I_2)}{2\sqrt{Z_0}}$$

The term *traveling waves* is justified below. Eqs. (13.1.4) may be inverted to express the voltages and currents in terms of the wave variables:

$$\boxed{\begin{aligned} V_1 &= \sqrt{Z_0}(a_1 + b_1) \\ I_1 &= \frac{1}{\sqrt{Z_0}}(a_1 - b_1) \end{aligned}} \qquad \boxed{\begin{aligned} V_2 &= \sqrt{Z_0}(a_2 + b_2) \\ I_2 &= \frac{1}{\sqrt{Z_0}}(b_2 - a_2) \end{aligned}} \tag{13.1.5}$$

In practice, the reference impedance is chosen to be $Z_0 = 50$ ohm. At lower frequencies the transfer and impedance matrices are commonly used, but at microwave frequencies they become difficult to measure and therefore, the scattering matrix description is preferred.

The S-parameters can be measured by embedding the two-port network (the device-under-test, or, DUT) in a transmission line whose ends are connected to a network analyzer. Fig. 13.1.2 shows the experimental setup.

A typical network analyzer can measure S-parameters over a large frequency range, for example, the HP 8720D vector network analyzer covers the range from 50 MHz to

40 GHz. Frequency resolution is typically 1 Hz and the results can be displayed either on a Smith chart or as a conventional gain versus frequency graph.

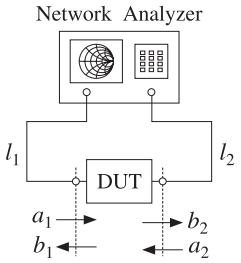


Fig. 13.1.2 Device under test connected to network analyzer.

Fig. 13.1.3 shows more details of the connection. The generator and load impedances are configured by the network analyzer. The connections can be reversed, with the generator connected to port 2 and the load to port 1.

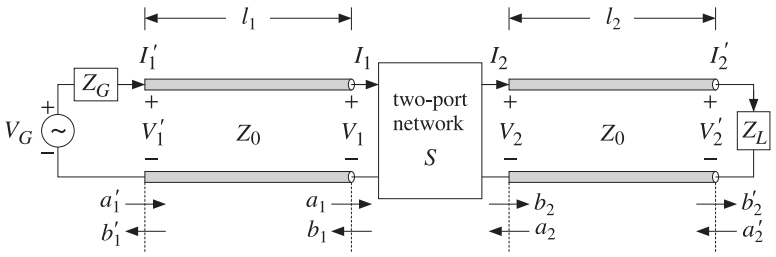


Fig. 13.1.3 Two-port network under test.

The two line segments of lengths l_1, l_2 are assumed to have characteristic impedance equal to the reference impedance Z_0. Then, the wave variables a_1, b_1 and a_2, b_2 are recognized as normalized versions of forward and backward *traveling waves*. Indeed, according to Eq. (10.7.8), we have:

$$\begin{aligned} a_1 &= \frac{V_1 + Z_0 I_1}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{1+} \qquad a_2 = \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{2-} \\ b_1 &= \frac{V_1 - Z_0 I_1}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{1-} \qquad b_2 = \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{2+} \end{aligned} \tag{13.1.6}$$

Thus, a_1 is essentially the incident wave at port 1 and b_1 the corresponding reflected wave. Similarly, a_2 is incident from the right onto port 2 and b_2 is the reflected wave from port 2.

The network analyzer measures the waves a_1', b_1' and a_2', b_2' at the generator and load ends of the line segments, as shown in Fig. 13.1.3. From these, the waves at the inputs of the two-port can be determined. Assuming lossless segments and using the propagation matrices (10.7.7), we have:

$$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} e^{-j\delta_1} & 0 \\ 0 & e^{j\delta_1} \end{bmatrix} \begin{bmatrix} a_1' \\ b_1' \end{bmatrix}, \qquad \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} e^{-j\delta_2} & 0 \\ 0 & e^{j\delta_2} \end{bmatrix} \begin{bmatrix} a_2' \\ b_2' \end{bmatrix} \tag{13.1.7}$$

where $\delta_1 = \beta l_l$ and $\delta_2 = \beta l_2$ are the phase lengths of the segments. Eqs. (13.1.7) can be rearranged into the forms:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = D \begin{bmatrix} b_1' \\ b_2' \end{bmatrix}, \quad \begin{bmatrix} a_1' \\ a_2' \end{bmatrix} = D \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad D = \begin{bmatrix} e^{j\delta_1} & 0 \\ 0 & e^{j\delta_2} \end{bmatrix}$$

The network analyzer measures the corresponding S-parameters of the primed variables, that is,

$$\begin{bmatrix} b_1' \\ b_2' \end{bmatrix} = \begin{bmatrix} S_{11}' & S_{12}' \\ S_{21}' & S_{22}' \end{bmatrix} \begin{bmatrix} a_1' \\ a_2' \end{bmatrix}, \quad S' = \begin{bmatrix} S_{11}' & S_{12}' \\ S_{21}' & S_{22}' \end{bmatrix} \quad \text{(measured } S\text{-matrix)} \tag{13.1.8}$$

The S-matrix of the two-port can be obtained then from:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = D \begin{bmatrix} b_1' \\ b_2' \end{bmatrix} = DS' \begin{bmatrix} a_1' \\ a_2' \end{bmatrix} = DS'D \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad \Rightarrow \quad S = DS'D$$

or, more explicitly,

$$\begin{aligned} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} &= \begin{bmatrix} e^{j\delta_1} & 0 \\ 0 & e^{j\delta_2} \end{bmatrix} \begin{bmatrix} S_{11}' & S_{12}' \\ S_{21}' & S_{22}' \end{bmatrix} \begin{bmatrix} e^{j\delta_1} & 0 \\ 0 & e^{j\delta_2} \end{bmatrix} \\ &= \begin{bmatrix} S_{11}' e^{2j\delta_1} & S_{12}' e^{j(\delta_1+\delta_2)} \\ S_{21}' e^{j(\delta_1+\delta_2)} & S_{22}' e^{2j\delta_2} \end{bmatrix} \end{aligned} \tag{13.1.9}$$

Thus, changing the points along the transmission lines at which the S-parameters are measured introduces only phase changes in the parameters.

Without loss of generality, we may replace the extended circuit of Fig. 13.1.3 with the one shown in Fig. 13.1.4 with the understanding that either we are using the extended two-port parameters S', or, equivalently, the generator and segment l_1 have been replaced by their Thévenin equivalents, and the load impedance has been replaced by its propagated version to distance l_2.

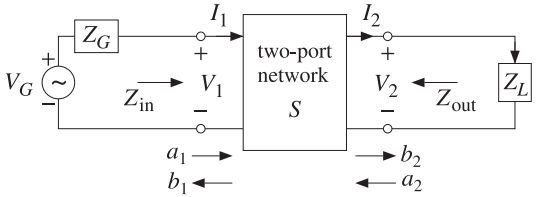


Fig. 13.1.4 Two-port network connected to generator and load.

The actual measurements of the S-parameters are made by connecting to a matched load, $Z_L = Z_0$. Then, there will be no reflected waves from the load, $a_2 = 0$, and the S-matrix equations will give:

$$b_1 = S_{11}a_1 + S_{12}a_2 = S_{11}a_1 \quad \Rightarrow \quad S_{11} = \left.\frac{b_1}{a_1}\right|_{Z_L=Z_0} = \text{reflection coefficient}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 = S_{21}a_1 \quad \Rightarrow \quad S_{21} = \left.\frac{b_2}{a_1}\right|_{Z_L=Z_0} = \text{transmission coefficient}$$

Reversing the roles of the generator and load, one can measure in the same way the parameters S_{12} and S_{22}.

13.2 Power Flow

Power flow into and out of the two-port is expressed very simply in terms of the traveling wave amplitudes. Using the inverse relationships (13.1.5), we find:

$$\begin{aligned} \frac{1}{2}\,\text{Re}[V_1^* I_1] &= \frac{1}{2}|a_1|^2 - \frac{1}{2}|b_1|^2 \\ -\frac{1}{2}\,\text{Re}[V_2^* I_2] &= \frac{1}{2}|a_2|^2 - \frac{1}{2}|b_2|^2 \end{aligned} \tag{13.2.1}$$

The left-hand sides represent the power flow *into* ports 1 and 2. The right-hand sides represent the difference between the power incident on a port and the power reflected from it. The quantity $\text{Re}[V_2^* I_2]/2$ represents the power transferred to the load.

Another way of phrasing these is to say that part of the incident power on a port gets reflected and part enters the port:

$$\begin{aligned} \frac{1}{2}|a_1|^2 &= \frac{1}{2}|b_1|^2 + \frac{1}{2}\,\text{Re}[V_1^* I_1] \\ \frac{1}{2}|a_2|^2 &= \frac{1}{2}|b_2|^2 + \frac{1}{2}\,\text{Re}[V_2^* (-I_2)] \end{aligned} \tag{13.2.2}$$

One of the reasons for normalizing the traveling wave amplitudes by $\sqrt{Z_0}$ in the definitions (13.1.4) was precisely this simple way of expressing the incident and reflected powers from a port.

If the two-port is lossy, the power lost in it will be the difference between the power entering port 1 and the power leaving port 2, that is,

$$P_{\text{loss}} = \frac{1}{2}\,\text{Re}[V_1^* I_1] - \frac{1}{2}\,\text{Re}[V_2^* I_2] = \frac{1}{2}|a_1|^2 + \frac{1}{2}|a_2|^2 - \frac{1}{2}|b_1|^2 - \frac{1}{2}|b_2|^2$$

Noting that $\mathbf{a}^\dagger\mathbf{a} = |a_1|^2 + |a_2|^2$ and $\mathbf{b}^\dagger\mathbf{b} = |b_1|^2 + |b_2|^2$, and writing $\mathbf{b}^\dagger\mathbf{b} = \mathbf{a}^\dagger S^\dagger S\mathbf{a}$, we may express this relationship in terms of the scattering matrix:

$$P_{\text{loss}} = \frac{1}{2}\mathbf{a}^\dagger\mathbf{a} - \frac{1}{2}\mathbf{b}^\dagger\mathbf{b} = \frac{1}{2}\mathbf{a}^\dagger\mathbf{a} - \frac{1}{2}\mathbf{a}^\dagger S^\dagger S\mathbf{a} = \frac{1}{2}\mathbf{a}^\dagger (I - S^\dagger S)\mathbf{a} \tag{13.2.3}$$

For a lossy two-port, the power loss is positive, which implies that the matrix $I - S^\dagger S$ must be positive definite. If the two-port is lossless, $P_{\text{loss}} = 0$, the S-matrix will be *unitary*, that is, $S^\dagger S = I$.

We already saw examples of such unitary scattering matrices in the cases of the equal travel-time multilayer dielectric structures and their equivalent quarter wavelength multisection transformers.

13.3 Parameter Conversions

It is straightforward to derive the relationships that allow one to pass from one parameter set to another. For example, starting with the transfer matrix, we have:

$$\begin{aligned} V_1 &= AV_2 + BI_2 \\ I_1 &= CV_2 + DI_2 \end{aligned} \quad \Rightarrow \quad \begin{aligned} V_1 &= A\left(\frac{1}{C}I_1 - \frac{D}{C}I_2\right) + BI_2 = \frac{A}{C}I_1 - \frac{AD-BC}{C}I_2 \\ V_2 &= \frac{1}{C}I_1 - \frac{D}{C}I_2 \end{aligned}$$

The coefficients of I_1, I_2 are the impedance matrix elements. The steps are reversible, and we summarize the final relationships below:

$$\begin{aligned} Z &= \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = \frac{1}{C}\begin{bmatrix} A & AD-BC \\ 1 & D \end{bmatrix} \\ T &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{Z_{21}}\begin{bmatrix} Z_{11} & Z_{11}Z_{22} - Z_{12}Z_{21} \\ 1 & Z_{22} \end{bmatrix} \end{aligned} \tag{13.3.1}$$

We note the determinants $\det(T) = AD - BC$ and $\det(Z) = Z_{11}Z_{22} - Z_{12}Z_{21}$. The relationship between the scattering and impedance matrices is also straightforward to derive. We define the 2×1 vectors:

$$\boldsymbol{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad \boldsymbol{I} = \begin{bmatrix} I_1 \\ -I_2 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{13.3.2}$$

Then, the definitions (13.1.4) can be written compactly as:

$$\begin{aligned} \mathbf{a} &= \frac{1}{2\sqrt{Z_0}}(\boldsymbol{V} + Z_0\boldsymbol{I}) = \frac{1}{2\sqrt{Z_0}}(Z + Z_0 I)\,\boldsymbol{I} \\ \mathbf{b} &= \frac{1}{2\sqrt{Z_0}}(\boldsymbol{V} - Z_0\boldsymbol{I}) = \frac{1}{2\sqrt{Z_0}}(Z - Z_0 I)\,\boldsymbol{I} \end{aligned} \tag{13.3.3}$$

where we used the impedance matrix relationship $\boldsymbol{V} = Z\boldsymbol{I}$ and defined the 2×2 unit matrix I. It follows then,

$$\frac{1}{2\sqrt{Z_0}}\boldsymbol{I} = (Z + Z_0 I)^{-1}\mathbf{a} \quad \Rightarrow \quad \mathbf{b} = \frac{1}{2\sqrt{Z_0}}(Z - Z_0 I)\,\boldsymbol{I} = (Z - Z_0 I)(Z + Z_0 I)^{-1}\mathbf{a}$$

Thus, the scattering matrix S will be related to the impedance matrix Z by

$$\boxed{S = (Z - Z_0 I)(Z + Z_0 I)^{-1}} \quad \Leftrightarrow \quad \boxed{Z = (I - S)^{-1}(I + S)\,Z_0} \tag{13.3.4}$$

Explicitly, we have:

$$S = \begin{bmatrix} Z_{11}-Z_0 & Z_{12} \\ Z_{21} & Z_{22}-Z_0 \end{bmatrix} \begin{bmatrix} Z_{11}+Z_0 & Z_{12} \\ Z_{21} & Z_{22}+Z_0 \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} Z_{11}-Z_0 & Z_{12} \\ Z_{21} & Z_{22}-Z_0 \end{bmatrix} \frac{1}{D_z} \begin{bmatrix} Z_{22}+Z_0 & -Z_{12} \\ -Z_{21} & Z_{11}+Z_0 \end{bmatrix}$$

where $D_z = \det(Z+Z_0 I) = (Z_{11}+Z_0)(Z_{22}+Z_0) - Z_{12}Z_{21}$. Multiplying the matrix factors, we obtain:

$$S = \frac{1}{D_z} \begin{bmatrix} (Z_{11}-Z_0)(Z_{22}+Z_0) - Z_{12}Z_{21} & 2Z_{12}Z_0 \\ 2Z_{21}Z_0 & (Z_{11}+Z_0)(Z_{22}-Z_0) - Z_{12}Z_{21} \end{bmatrix} \tag{13.3.5}$$

Similarly, the inverse relationship gives:

$$Z = \frac{Z_0}{D_s} \begin{bmatrix} (1+S_{11})(1-S_{22}) + S_{12}S_{21} & 2S_{12} \\ 2S_{21} & (1-S_{11})(1+S_{22}) + S_{12}S_{21} \end{bmatrix} \tag{13.3.6}$$

where $D_s = \det(I-S) = (1-S_{11})(1-S_{22}) - S_{12}S_{21}$. Expressing the impedance parameters in terms of the transfer matrix parameters, we also find:

$$S = \frac{1}{D_a} \begin{bmatrix} A + \dfrac{B}{Z_0} - CZ_0 - D & 2(AD-BC) \\ 2 & -A + \dfrac{B}{Z_0} - CZ_0 + D \end{bmatrix} \tag{13.3.7}$$

where $D_a = A + \dfrac{B}{Z_0} + CZ_0 + D$.

13.4 Input and Output Reflection Coefficients

When the two-port is connected to a generator and load as in Fig. 13.1.4, the impedance and scattering matrix equations take the simpler forms:

$$\boxed{\begin{aligned} V_1 &= Z_{\text{in}} I_1 \\ V_2 &= Z_L I_2 \end{aligned}} \quad \Leftrightarrow \quad \boxed{\begin{aligned} b_1 &= \Gamma_{\text{in}} a_1 \\ a_2 &= \Gamma_L b_2 \end{aligned}} \tag{13.4.1}$$

where Z_{in} is the input impedance at port 1, and $\Gamma_{\text{in}}, \Gamma_L$ are the reflection coefficients at port 1 and at the load:

$$\Gamma_{\text{in}} = \frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0}, \quad \Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{13.4.2}$$

The input impedance and input reflection coefficient can be expressed in terms of the Z- and S-parameters, as follows:

$$\boxed{Z_{\text{in}} = Z_{11} - \frac{Z_{12}Z_{21}}{Z_{22}+Z_L}} \quad \Leftrightarrow \quad \boxed{\Gamma_{\text{in}} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-S_{22}\Gamma_L}} \tag{13.4.3}$$

The equivalence of these two expressions can be shown by using the parameter conversion formulas of Eqs. (13.3.5) and (13.3.6), or they can be shown indirectly, as follows. Starting with $V_2 = Z_L I_2$ and using the second impedance matrix equation, we can solve for I_2 in terms of I_1:

$$V_2 = Z_{21}I_1 - Z_{22}I_2 = Z_L I_2 \quad \Rightarrow \quad I_2 = \frac{Z_{21}}{Z_{22}+Z_L} I_1 \tag{13.4.4}$$

Then, the first impedance matrix equation implies:

$$V_1 = Z_{11}I_1 - Z_{12}I_2 = \left(Z_{11} - \frac{Z_{12}Z_{21}}{Z_{22}+Z_L} \right) I_1 = Z_{\text{in}} I_1$$

Starting again with $V_2 = Z_L I_2$ we find for the traveling waves at port 2:

$$\begin{aligned} a_2 &= \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} = \frac{Z_L - Z_0}{2\sqrt{Z_0}} I_2 \\ b_2 &= \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} = \frac{Z_L + Z_0}{2\sqrt{Z_0}} I_2 \end{aligned} \quad \Rightarrow \quad a_2 = \frac{Z_L - Z_0}{Z_L + Z_0} b_2 = \Gamma_L b_2$$

Using $V_1 = Z_{\text{in}} I_1$, a similar argument implies for the waves at port 1:

$$\begin{aligned} a_1 &= \frac{V_1 + Z_0 I_1}{2\sqrt{Z_0}} = \frac{Z_{\text{in}} + Z_0}{2\sqrt{Z_0}} I_1 \\ b_1 &= \frac{V_1 - Z_0 I_1}{2\sqrt{Z_0}} = \frac{Z_{\text{in}} - Z_0}{2\sqrt{Z_0}} I_1 \end{aligned} \quad \Rightarrow \quad b_1 = \frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0} a_1 = \Gamma_{\text{in}} a_1$$

It follows then from the scattering matrix equations that:

$$b_2 = S_{21}a_1 + S_{22}a_2 = S_{22}a_1 + S_{22}\Gamma_L b_2 \quad \Rightarrow \quad \boxed{b_2 = \frac{S_{21}}{1-S_{22}\Gamma_L} a_1} \tag{13.4.5}$$

which implies for b_1:

$$b_1 = S_{11}a_1 + S_{12}a_2 = S_{11}a_1 + S_{12}\Gamma_L b_2 = \left(S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-S_{22}\Gamma_L} \right) a_1 = \Gamma_{\text{in}} a_1$$

Reversing the roles of generator and load, we obtain the impedance and reflection coefficients from the output side of the two-port:

$$\boxed{Z_{\text{out}} = Z_{22} - \frac{Z_{12}Z_{21}}{Z_{11}+Z_G}} \quad \Leftrightarrow \quad \boxed{\Gamma_{\text{out}} = S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1-S_{11}\Gamma_G}} \tag{13.4.6}$$

where

$$\Gamma_{\text{out}} = \frac{Z_{\text{out}} - Z_0}{Z_{\text{out}} + Z_0}, \quad \Gamma_G = \frac{Z_G - Z_0}{Z_G + Z_0} \tag{13.4.7}$$

The input and output impedances allow one to replace the original two-port circuit of Fig. 13.1.4 by simpler equivalent circuits. For example, the two-port and the load can be replaced by the input impedance Z_{in} connected at port 1, as shown in Fig. 13.4.1.

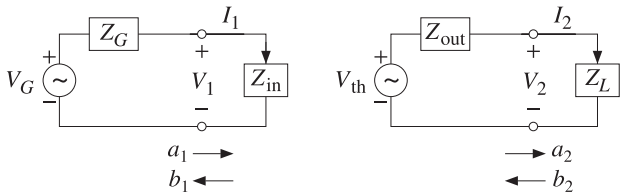


Fig. 13.4.1 Input and output equivalent circuits.

Similarly, the generator and the two-port can be replaced by a Thévenin equivalent circuit connected at port 2. By determining the open-circuit voltage and short-circuit current at port 2, we find the corresponding Thévenin parameters in terms of the impedance parameters:

$$\boxed{V_{\rm th} = \frac{Z_{21}V_G}{Z_{11}+Z_G}\,, \quad Z_{\rm th} = Z_{\rm out} = Z_{22} - \frac{Z_{12}Z_{21}}{Z_{11}+Z_G}} \tag{13.4.8}$$

13.5 Stability Circles

In discussing the stability conditions of a two-port in terms of S-parameters, the following definitions of constants are often used:

$$\begin{aligned}
\Delta &= \det(S) = S_{11}S_{22} - S_{12}S_{21} \\
K &= \frac{1-|S_{11}|^2-|S_{22}|^2+|\Delta|^2}{2|S_{12}S_{21}|} \qquad \text{(Rollett stability factor)} \\
\mu_1 &= \frac{1-|S_{11}|^2}{|S_{22}-\Delta S_{11}^*|+|S_{12}S_{21}|} \qquad \text{(Edwards-Sinsky stability parameter)} \\
\mu_2 &= \frac{1-|S_{22}|^2}{|S_{11}-\Delta S_{22}^*|+|S_{12}S_{21}|} \\
B_1 &= 1+|S_{11}|^2-|S_{22}|^2-|\Delta|^2 \\
B_2 &= 1+|S_{22}|^2-|S_{11}|^2-|\Delta|^2 \\
C_1 &= S_{11}-\Delta S_{22}^*\,, \quad D_1 = |S_{11}|^2-|\Delta|^2 \\
C_2 &= S_{22}-\Delta S_{11}^*\,, \quad D_2 = |S_{22}|^2-|\Delta|^2
\end{aligned} \tag{13.5.1}$$

The quantity K is the Rollett stability factor [991], and μ_1, μ_2, the Edwards-Sinsky stability parameters [994]. The following identities hold among these constants:

$$\begin{aligned}
&B_1^2 - 4|C_1|^2 = B_2^2 - 4|C_2|^2 = 4|S_{12}S_{21}|^2(K^2-1) \\
&|C_1|^2 = |S_{12}S_{21}|^2 + (1-|S_{22}|^2)D_1 \\
&|C_2|^2 = |S_{12}S_{21}|^2 + (1-|S_{11}|^2)D_2
\end{aligned} \tag{13.5.2}$$

For example, noting that $S_{12}S_{21} = S_{11}S_{22} - \Delta$, the last of Eqs. (13.5.2) is a direct consequence of the identity:

$$|A-BC|^2 - |B-AC^*|^2 = (1-|C|^2)\,(|A|^2-|B|^2) \tag{13.5.3}$$

We define also the following parameters, which will be recognized as the centers and radii of the source and load stability circles:

$$c_G = \frac{C_1^*}{D_1}\,, \quad r_G = \frac{|S_{12}S_{21}|}{|D_1|} \qquad \text{(source stability circle)} \tag{13.5.4}$$

$$c_L = \frac{C_2^*}{D_2}\,, \quad r_L = \frac{|S_{12}S_{21}|}{|D_2|} \qquad \text{(load stability circle)} \tag{13.5.5}$$

They satisfy the following relationships, which are consequences of the last two of Eqs. (13.5.2) and the definitions (13.5.4) and (13.5.5):

$$\begin{aligned}
1-|S_{11}|^2 &= (|c_L|^2 - r_L^2)D_2 \\
1-|S_{22}|^2 &= (|c_G|^2 - r_G^2)D_1
\end{aligned} \tag{13.5.6}$$

We note also that using Eqs. (13.5.6), the stability parameters μ_1, μ_2 can be written as:

$$\begin{aligned}
\mu_1 &= (|c_L| - r_L)\,\text{sign}(D_2) \\
\mu_2 &= (|c_G| - r_G)\,\text{sign}(D_1)
\end{aligned} \tag{13.5.7}$$

For example, we have:

$$\mu_1 = \frac{1-|S_{11}|^2}{|C_2|+|S_{12}S_{21}|} = \frac{D_2(|c_L|^2-r_L^2)}{|D_2||c_L|+|D_2|r_L} = \frac{D_2(|c_L|^2-r_L^2)}{|D_2|(|c_L|+r_L)} = \frac{D_2}{|D_2|}(|c_L|-r_L)$$

We finally note that the input and output reflection coefficients can be written in the alternative forms:

$$\begin{aligned}
\Gamma_{\rm in} &= S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-S_{22}\Gamma_L} = \frac{S_{11}-\Delta\Gamma_L}{1-S_{22}\Gamma_L} \\
\Gamma_{\rm out} &= S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1-S_{22}\Gamma_G} = \frac{S_{22}-\Delta\Gamma_G}{1-S_{11}\Gamma_G}
\end{aligned} \tag{13.5.8}$$

Next, we discuss the stability conditions. The two-port is *unconditionally stable* if any generator and load impedances with positive resistive parts R_G, R_L, will always lead to input and output impedances with positive resistive parts $R_{\rm in}, R_{\rm out}$.

Equivalently, unconditional stability requires that any load and generator with $|\Gamma_L| < 1$ and $|\Gamma_G| < 1$ will result into $|\Gamma_{\rm in}| < 1$ and $|\Gamma_{\rm out}| < 1$.

The two-port is termed *potentially or conditionally unstable* if there are $|\Gamma_L| < 1$ and $|\Gamma_G| < 1$ resulting into $|\Gamma_{\rm in}| \geq 1$ and/or $|\Gamma_{\rm out}| \geq 1$.

The *load stability region* is the set of all Γ_L that result into $|\Gamma_{\rm in}| < 1$, and the *source stability region*, the set of all Γ_G that result into $|\Gamma_{\rm out}| < 1$.

In the unconditionally stable case, the load and source stability regions contain the entire unit-circles $|\Gamma_L| < 1$ or $|\Gamma_G| < 1$. However, in the potentially unstable case, only

portions of the unit-circles may lie within the stability regions and such Γ_G, Γ_L will lead to a stable input and output impedances.

The connection of the stability regions to the stability circles is brought about by the following identities, which can be proved easily using Eqs. (13.5.1)–(13.5.8):

$$\begin{aligned} 1 - |\Gamma_{\text{in}}|^2 &= \frac{|\Gamma_L - c_L|^2 - r_L^2}{|1 - S_{22}\Gamma_L|^2} D_2 \\ 1 - |\Gamma_{\text{out}}|^2 &= \frac{|\Gamma_G - c_G|^2 - r_G^2}{|1 - S_{11}\Gamma_G|^2} D_1 \end{aligned} \tag{13.5.9}$$

For example, the first can be shown starting with Eq. (13.5.8) and using the definitions (13.5.5) and the relationship (13.5.6):

$$\begin{aligned} 1 - |\Gamma_{\text{in}}|^2 &= 1 - \left| \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} \right|^2 = \frac{|S_{11} - \Delta\Gamma_L|^2 - |1 - S_{22}\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \\ &= \frac{(|S_{22}|^2 - |\Delta|^2)|\Gamma_L|^2 - (S_{22} - \Delta S_{11}^*)\Gamma_L - (S_{22}^* - \Delta^* S_{11})\Gamma_L^* + 1 - |S_{11}|^2}{|1 - S_{22}\Gamma_L|^2} \\ &= \frac{D_2|\Gamma_L|^2 - C_2\Gamma_L - C_2^*\Gamma_L^* + 1 - |S_{11}|^2}{|1 - S_{22}\Gamma_L|^2} \\ &= \frac{D_2(|\Gamma_L|^2 - c_L^*\Gamma_L - c_L^*\Gamma_L^* + |c_L|^2 - r_L^2)}{|1 - S_{22}\Gamma_L|^2} = \frac{D_2(|\Gamma_L - c_L|^2 - r_L^2)}{|1 - S_{22}\Gamma_L|^2} \end{aligned}$$

It follows from Eq. (13.5.9) that the load stability region is defined by the conditions:

$$1 - |\Gamma_{\text{in}}|^2 > 0 \quad \Leftrightarrow \quad (|\Gamma_L - c_L|^2 - r_L^2) D_2 > 0$$

Depending on the sign of D_2, these are equivalent to the outside or the inside of the load stability circle of center c_L and radius r_L:

$$\boxed{\begin{aligned} |\Gamma_L - c_L| &> r_L, \quad \text{if } D_2 > 0 \\ |\Gamma_L - c_L| &< r_L, \quad \text{if } D_2 < 0 \end{aligned}} \quad \text{(load stability region)} \tag{13.5.10}$$

The boundary of the circle $|\Gamma_L - c_L| = r_L$ corresponds to $|\Gamma_{\text{in}}| = 1$. The complement of these regions corresponds to the unstable region with $|\Gamma_{\text{in}}| > 1$. Similarly, we find for the source stability region:

$$\boxed{\begin{aligned} |\Gamma_G - c_G| &> r_G, \quad \text{if } D_1 > 0 \\ |\Gamma_G - c_G| &< r_G, \quad \text{if } D_1 < 0 \end{aligned}} \quad \text{(source stability region)} \tag{13.5.11}$$

In order to have unconditional stability, the stability regions must contain the unit-circle in its entirety. If $D_2 > 0$, the unit-circle and load stability circle must not overlap at all, as shown in Fig. 13.5.1. Geometrically, the distance between the points O and A in the figure is $(OA) = |c_L| - r_L$. The non-overlapping of the circles requires the condition $(OA) > 1$, or, $|c_L| - r_L > 1$.

If $D_2 < 0$, the stability region is the inside of the stability circle, and therefore, the unit-circle must lie within that circle. This requires that $(OA) = r_L - |c_L| > 1$, as shown in Fig. 13.5.1.

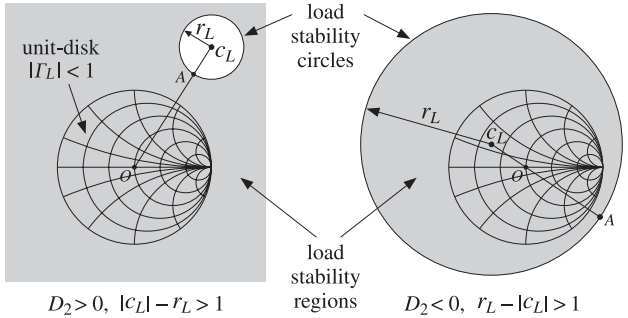


Fig. 13.5.1 Load stability regions in the unconditionally stable case.

These two conditions can be combined into $\text{sign}(D_2)\,(|c_L| - r_L) > 1$. But, that is equivalent to $\mu_1 > 1$ according to Eq. (13.5.7). Geometrically, the parameter μ_1 represents the distance (OA). Thus, the condition for the unconditional stability of the input is equivalent to:

$$\boxed{\mu_1 > 1} \quad \text{(unconditional stability condition)} \tag{13.5.12}$$

It has been shown by Edwards and Sinsky [994] that this single condition (or, alternatively, the single condition $\mu_2 > 1$) is necessary and sufficient for the unconditional stability of both the input and output impedances of the two-port. Clearly, the source stability regions will be similar to those of Fig. 13.5.1.

If the stability condition is not satisfied, that is, $\mu_1 < 1$, then only that portion of the unit-circle that lies within the stability region will be stable and will lead to stable input and output impedances. Fig. 13.5.2 illustrates such a potentially unstable case.

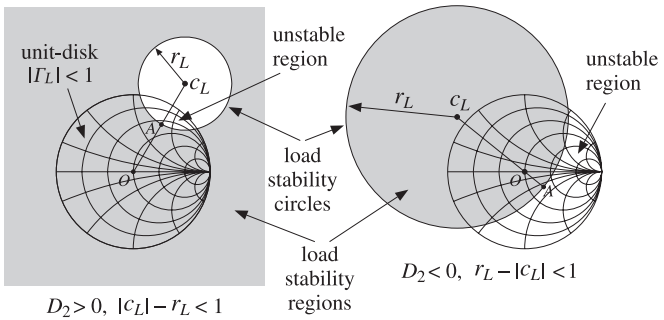


Fig. 13.5.2 Load stability regions in potentially unstable case.

If $D_2 > 0$, then $\mu_1 < 1$ is equivalent to $|c_L| - r_L < 1$, and if $D_2 < 0$, it is equivalent to $r_L - |c_L| < 1$. In either case, the unit-circle is partially overlapping with the stability

circle, as shown in Fig. 13.5.2. The portion of the unit-circle that does not lie within the stability region will correspond to an unstable Z_{in}.

There exist several other unconditional stability criteria that are equivalent to the single criterion $\mu_1 > 1$. They all require that the Rollett stability factor K be greater than unity, $K > 1$, as well as one other condition. Any one of the following criteria are necessary and sufficient for unconditional stability [992]:

$$\begin{aligned} &K > 1 \quad \text{and} \quad |\Delta| < 1 \\ &K > 1 \quad \text{and} \quad B_1 > 0 \\ &K > 1 \quad \text{and} \quad B_2 > 0 \qquad \text{(stability conditions)} \\ &K > 1 \quad \text{and} \quad |S_{12}S_{21}| < 1 - |S_{11}|^2 \\ &K > 1 \quad \text{and} \quad |S_{12}S_{21}| < 1 - |S_{22}|^2 \end{aligned} \tag{13.5.13}$$

Their equivalence to $\mu_1 > 1$ has been shown in [994]. In particular, it follows from the last two conditions that unconditional stability requires $|S_{11}| < 1$ and $|S_{22}| < 1$. These are necessary but not sufficient for stability.

A very common circumstance in practice is to have a potentially unstable two-port, but with $|S_{11}| < 1$ and $|S_{22}| < 1$. In such cases, Eq. (13.5.6) implies $D_2\,(|c_L|^2 - r_L^2) > 0$, and the lack of stability requires $\mu_1 = \text{sign}(D_2)\,(|c_L|^2 - r_L^2) < 1$.

Therefore, if $D_2 > 0$, then we must have $|c_L|^2 - r_L^2 > 0$ and $|c_L| - r_L < 1$, which combine into the inequality $r_L < |c_L| < r_L + 1$. This is depicted in the left picture of Fig. 13.5.2. The geometrical distance $(OA) = |c_L| - r_L$ satisfies $0 < (OA) < 1$, so that stability circle partially overlaps with the unit-circle but does not enclose its center.

On the other hand, if $D_2 < 0$, the two conditions require $|c_L|^2 - r_L^2 < 0$ and $r_L - |c_L| < 1$, which imply $|c_L| < r_L < |c_L| + 1$. This is depicted in the right Fig. 13.5.2. The geometrical distance $(OA) = r_L - |c_L|$ again satisfies $0 < (OA) < 1$, but now the center of the unit-circle lies within the stability circle, which is also the stability region.

We have written a number of MATLAB functions that facilitate working with S-parameters. They are described in detail later on:

`smat`	reshape S-parameters into S-matrix
`sparam`	calculate stability parameters
`sgain`	calculate transducer, available, operating, and unilateral power gains
`smatch`	calculate simultaneous conjugate match for generator and load
`gin,gout`	calculate input and output reflection coefficients
`smith`	draw a basic Smith chart
`smithcir`	draw a stability or gain circle on Smith chart
`sgcirc`	determine stability and gain circles
`nfcirc`	determine noise figure circles
`nfig`	calculate noise figure

The MATLAB function `sparam` calculates the stability parameters $\mu_1, K, |\Delta|, B_1, B_2$, as well as the parameters C_1, C_2, D_1, D_2. It has usage:

```
[K,mu,D,B1,B2,C1,C2,D1,D2] = sparam(S);          % stability parameters
```

The function `sgcirc` calculates the centers and radii of the source and load stability circles. It also calculates gain circles to be discussed later on. Its usage is:


```
[cL,rL] = sgcirc(S,'l');          % load or Zin stability circle
[cG,rG] = sgcirc(S,'s');          % source or Zout stability circle
```

The MATLAB function `smith` draws a basic Smith chart, and the function `smithcir` draws the stability circles:

```
smith(n);                         % draw four basic types of Smith charts, n = 1,2,3,4
smith;                            % default Smith chart corresponding to n = 3

smithcir(c,r,max,width);          % draw circle of center c and radius r
smithcir(c,r,max);                % equivalent to linewidth width=1
smithcir(c,r);                    % draw full circle with linewidth width=1
```

The parameter `max` controls the portion of the stability circle that is visible outside the Smith chart. For example, `max` = 1.1 will display only that portion of the circle that has $|\Gamma| < 1.1$.

Example 13.5.1: The Hewlett-Packard AT-41511 NPN bipolar transistor has the following S-parameters at 1 GHz and 2 GHz [1355]:

$$\begin{aligned} &S_{11} = 0.48\angle{-149^\circ}, \quad S_{21} = 5.189\angle{89^\circ}, \quad S_{12} = 0.073\angle{43^\circ}, \quad S_{22} = 0.49\angle{-39^\circ} \\ &S_{11} = 0.46\angle{162^\circ}, \quad S_{21} = 2.774\angle{59^\circ}, \quad S_{12} = 0.103\angle{45^\circ}, \quad S_{22} = 0.42\angle{-47^\circ} \end{aligned}$$

Determine the stability parameters, stability circles, and stability regions.

Solution: The transistor is potentially unstable at 1 GHz, but unconditionally stable at 2 GHz. The source and load stability circles at 1 GHz are shown in Fig. 13.5.3.

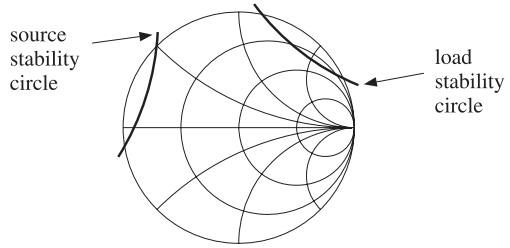


Fig. 13.5.3 Load and source stability circles at 1 GHz.

The MATLAB code used to generate this graph was:

```
S = smat([0.48 -149 5.189 89 0.073 43 0.49 -39]);   % form S-matrix

[K,mu,D,B1,B2,C1,C2,D1,D2] = sparam(S);             % stability parameters

[cL,rL] = sgcirc(S,'l');                            % stability circles
[cG,rG] = sgcirc(S,'s');

smith;                                              % draw basic Smith chart
smithcir(cL, rL, 1.1, 1.5);                         % draw stability circles
smithcir(cG, rG, 1.1, 1.5);
```

The computed stability parameters at 1 GHz were:

$$[K, \mu_1, |\Delta|, B_1, B_2, D_1, D_2] = [0.781,\ 0.847,\ 0.250,\ 0.928,\ 0.947,\ 0.168,\ 0.178]$$

The transistor is potentially unstable because $K < 1$ even though $|\Delta| < 1$, $B_1 > 0$, and $B_2 > 0$. The load and source stability circle centers and radii were:

$$c_L = 2.978\angle 51.75^\circ, \quad r_L = 2.131$$
$$c_G = 3.098\angle 162.24^\circ, \quad r_G = 2.254$$

Because both D_1 and D_2 are positive, both stability regions will be the portion of the Smith chart that lies outside the stability circles. For 2 GHz, we find:

$$[K, \mu_1, |\Delta|, B_1, B_2, D_1, D_2] = [1.089,\ 1.056,\ 0.103,\ 1.025,\ 0.954,\ 0.201,\ 0.166]$$

$$c_L = 2.779\angle 50.12^\circ, \quad r_L = 1.723$$
$$c_G = 2.473\angle -159.36^\circ, \quad r_G = 1.421$$

The transistor is stable at 2 GHz, with both load and source stability circles being completely outside the unit-circle. □

Problem 13.2 presents an example for which the D_2 parameter is negative, so that the stability regions will be the insides of the stability circles. At one frequency, the unit-circle is partially overlapping with the stability circle, while at another frequency, it lies entirely within the stability circle.

13.6 Power Gains

The amplification (or attenuation) properties of the two-port can be deduced by comparing the power $P_{\rm in}$ going into the two-port to the power P_L coming out of the two-port and going into the load. These were given in Eq. (13.2.1) and we rewrite them as:

$$\begin{aligned} P_{\rm in} &= \frac{1}{2}\,\mathrm{Re}[V_1^* I_1] = \frac{1}{2} R_{\rm in} |I_1|^2 \qquad \text{(power into two-port)} \\ P_L &= \frac{1}{2}\,\mathrm{Re}[V_2^* I_2] = \frac{1}{2} R_L |I_2|^2 \qquad \text{(power out of two-port and into load)} \end{aligned} \tag{13.6.1}$$

where we used $V_1 = Z_{\rm in} I_1$, $V_2 = Z_L I_2$, and defined the real parts of the input and load impedances by $R_{\rm in} = \mathrm{Re}(Z_{\rm in})$ and $R_L = \mathrm{Re}(Z_L)$. Using the equivalent circuits of Fig. 13.4.1, we may write I_1, I_2 in terms of the generator voltage V_G and obtain:

$$\begin{aligned} P_{\rm in} &= \frac{1}{2} \frac{|V_G|^2 R_{\rm in}}{|Z_{\rm in} + Z_G|^2} \\ P_L &= \frac{1}{2} \frac{|V_{\rm th}|^2 R_L}{|Z_{\rm out} + Z_L|^2} = \frac{1}{2} \frac{|V_G|^2 R_L |Z_{21}|^2}{\big|\,(Z_{11} + Z_G)\,(Z_{\rm out} + Z_L)\,\big|^2} \end{aligned} \tag{13.6.2}$$

Using the identities of Problem 13.1, P_L can also be written in the alternative forms:

$$P_L = \frac{1}{2} \frac{|V_G|^2 R_L |Z_{21}|^2}{\big|\,(Z_{22} + Z_L)\,(Z_{\rm in} + Z_G)\,\big|^2} = \frac{1}{2} \frac{|V_G|^2 R_L |Z_{21}|^2}{\big|\,(Z_{11} + Z_G)\,(Z_{22} + Z_L) - Z_{12} Z_{21}\,\big|^2} \tag{13.6.3}$$

The *maximum* power that can be delivered by the generator to a connected load is called the *available power* of the generator, $P_{\rm avG}$, and is obtained when the load is conjugate-matched to the generator, that is, $P_{\rm avG} = P_{\rm in}$ when $Z_{\rm in} = Z_G^*$.

Similarly, the *available* power from the two-port network, $P_{\rm avN}$, is the maximum power that can be delivered by the Thévenin-equivalent circuit of Fig. 13.4.1 to a connected load, that is, $P_{\rm avN} = P_L$ when $Z_L = Z_{\rm th}^* = Z_{\rm out}^*$. It follows then from Eq. (13.6.2) that the available powers will be:

$$\begin{aligned} P_{\rm avG} &= \max P_{\rm in} = \frac{|V_G|^2}{8R_G} \qquad \text{(available power from generator)} \\ P_{\rm avN} &= \max P_L = \frac{|V_{\rm th}|^2}{8R_{\rm out}} \qquad \text{(available power from network)} \end{aligned} \tag{13.6.4}$$

Using Eq. (13.4.8), $P_{\rm avN}$ can also be written as:

$$P_{\rm avN} = \frac{|V_G|^2}{8R_{\rm out}} \left| \frac{Z_{21}}{Z_{11} + Z_G} \right|^2 \tag{13.6.5}$$

The powers can be expressed completely in terms of the S-parameters of the two-port and the input and output reflection coefficients. With the help of the identities of Problem 13.1, we find the alternative expressions for $P_{\rm in}$ and P_L:

$$\begin{aligned} P_{\rm in} &= \frac{|V_G|^2}{8Z_0} \frac{(1 - |\Gamma_{\rm in}|^2)\,|1 - \Gamma_G|^2}{|1 - \Gamma_{\rm in}\Gamma_G|^2} \\ P_L &= \frac{|V_G|^2}{8Z_0} \frac{(1 - |\Gamma_L|^2)\,|1 - \Gamma_G|^2 |S_{21}|^2}{\big|\,(1 - \Gamma_{\rm in}\Gamma_G)\,(1 - S_{22}\Gamma_L)\,\big|^2} \\ &= \frac{|V_G|^2}{8Z_0} \frac{(1 - |\Gamma_L|^2)\,|1 - \Gamma_G|^2 |S_{21}|^2}{\big|\,(1 - \Gamma_{\rm out}\Gamma_L)\,(1 - S_{11}\Gamma_G)\,\big|^2} \\ &= \frac{|V_G|^2}{8Z_0} \frac{(1 - |\Gamma_L|^2)\,|1 - \Gamma_G|^2 |S_{21}|^2}{\big|\,(1 - S_{11}\Gamma_G)\,(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L\,\big|^2} \end{aligned} \tag{13.6.6}$$

Similarly, we have for $P_{\rm avG}$ and $P_{\rm avN}$:

$$\begin{aligned} P_{\rm avG} &= \frac{|V_G|^2}{8Z_0} \frac{|1 - \Gamma_G|^2}{1 - |\Gamma_G|^2} \\ P_{\rm avN} &= \frac{|V_G|^2}{8Z_0} \frac{|1 - \Gamma_G|^2 |S_{21}|^2}{(1 - |\Gamma_{\rm out}|^2)\,|1 - S_{11}\Gamma_G|^2} \end{aligned} \tag{13.6.7}$$

It is evident that $P_{\rm avG}$, $P_{\rm avN}$ are obtained from $P_{\rm in}$, P_L by setting $\Gamma_{\rm in} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\rm out}^*$, which are equivalent to the conjugate-match conditions.

Three widely used definitions for the *power gain* of the two-port network are the *transducer* power gain G_T, the *available* power gain G_a, and the power gain G_p, also called the operating gain. They are defined as follows:

$$\begin{aligned} G_T &= \frac{\text{power out of network}}{\text{maximum power in}} = \frac{P_L}{P_{\text{avG}}} \qquad \text{(transducer power gain)} \\ G_a &= \frac{\text{maximum power out}}{\text{maximum power in}} = \frac{P_{\text{avN}}}{P_{\text{avG}}} \qquad \text{(available power gain)} \\ G_p &= \frac{\text{power out of network}}{\text{power into network}} = \frac{P_L}{P_{\text{in}}} \qquad \text{(operating power gain)} \end{aligned} \tag{13.6.8}$$

Each gain is expressible either in terms of the Z-parameters of the two-port, or in terms of its S-parameters. In terms of Z-parameters, the transducer gain is given by the following forms, obtained from the three forms of P_L in Eqs. (13.6.2) and (13.6.3):

$$\begin{aligned} G_T &= \frac{4R_G R_L |Z_{21}|^2}{\big|\,(Z_{22}+Z_L)\,(Z_{\text{in}}+Z_G)\,\big|^2} \\ &= \frac{4R_G R_L |Z_{21}|^2}{\big|\,(Z_{11}+Z_G)\,(Z_{\text{out}}+Z_L)\,\big|^2} \\ &= \frac{4R_G R_L |Z_{21}|^2}{\big|\,(Z_{11}+Z_G)\,(Z_{22}+Z_L)-Z_{12}Z_{21}\,\big|^2} \end{aligned} \tag{13.6.9}$$

And, in terms of the S-parameters:

$$\boxed{\begin{aligned} G_T &= \frac{1-|\Gamma_G|^2}{|1-\Gamma_{\text{in}}\Gamma_G|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \\ &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-\Gamma_{\text{out}}\Gamma_L|^2} \\ &= \frac{(1-|\Gamma_G|^2)|S_{21}|^2(1-|\Gamma_L|^2)}{\big|\,(1-S_{11}\Gamma_G)\,(1-S_{22}\Gamma_L)-S_{12}S_{21}\Gamma_G\Gamma_L\,\big|^2} \end{aligned}} \tag{13.6.10}$$

Similarly, we have for G_a and G_p:

$$\boxed{\begin{aligned} G_a &= \frac{R_G}{R_{\text{out}}}\left|\frac{Z_{21}}{Z_{11}+Z_G}\right|^2 = \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1}{1-|\Gamma_{\text{out}}|^2} \\ G_p &= \frac{R_L}{R_{\text{in}}}\left|\frac{Z_{21}}{Z_{22}+Z_L}\right|^2 = \frac{1}{1-|\Gamma_{\text{in}}|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \end{aligned}} \tag{13.6.11}$$

The transducer gain G_T is, perhaps, the most representative measure of gain for the two-port because it incorporates the effects of both the load and generator impedances, whereas G_a depends only on the generator impedance and G_p only on the load impedance.

If the generator and load impedances are matched to the reference impedance Z_0, so that $Z_G = Z_L = Z_0$ and $\Gamma_G = \Gamma_L = 0$, and $\Gamma_{\text{in}} = S_{11}$, $\Gamma_{\text{out}} = S_{22}$, then the power gains reduce to:

$$G_T = |S_{21}|^2\,, \quad G_a = \frac{|S_{21}|^2}{1-|S_{22}|^2}\,, \quad G_p = \frac{|S_{21}|^2}{1-|S_{11}|^2} \tag{13.6.12}$$

A *unilateral* two-port has by definition zero reverse transmission coefficient, that is, $S_{12} = 0$. In this case, the input and output reflection coefficients simplify into:

$$\Gamma_{\text{in}} = S_{11}\,, \quad \Gamma_{\text{out}} = S_{22} \qquad \text{(unilateral two-port)} \tag{13.6.13}$$

The expressions of the power gains simplify somewhat in this case:

$$\begin{aligned} G_{Tu} &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \\ G_{au} &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1}{1-|S_{22}|^2} \qquad \text{(unilateral gains)} \\ G_{pu} &= \frac{1}{1-|S_{11}|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \end{aligned} \tag{13.6.14}$$

For both the bilateral and unilateral cases, the gains G_a, G_p are obtainable from G_T by setting $\Gamma_L = \Gamma_{\text{out}}^*$ and $\Gamma_{\text{in}} = \Gamma_G^*$, respectively, as was the case for P_{avN} and P_{avG}.

The relative power ratios $P_{\text{in}}/P_{\text{avG}}$ and P_L/P_{avN} measure the mismatching between the generator and the two-port and between the load and the two-port. Using the definitions for the power gains, we obtain the input and output *mismatch factors*:

$$M_{\text{in}} = \frac{P_{\text{in}}}{P_{\text{avG}}} = \frac{G_T}{G_p} = \frac{4R_{\text{in}}R_G}{|Z_{\text{in}}+Z_G|^2} = \frac{(1-|\Gamma_{\text{in}}|^2)\,(1-|\Gamma_G|^2)}{|1-\Gamma_{\text{in}}\Gamma_G|^2} \tag{13.6.15}$$

$$M_{\text{out}} = \frac{P_L}{P_{\text{avN}}} = \frac{G_T}{G_a} = \frac{4R_{\text{out}}R_L}{|Z_{\text{out}}+Z_L|^2} = \frac{(1-|\Gamma_{\text{out}}|^2)\,(1-|\Gamma_L|^2)}{|1-\Gamma_{\text{out}}\Gamma_L|^2} \tag{13.6.16}$$

The mismatch factors are always less than or equal to unity (for positive R_{in} and R_{out}.) Clearly, $M_{\text{in}} = 1$ under the conjugate-match condition $Z_{\text{in}} = Z_G^*$ or $\Gamma_{\text{in}} = \Gamma_G^*$, and $M_{\text{out}} = 1$ if $Z_L = Z_{\text{out}}^*$ or $\Gamma_L = \Gamma_{\text{out}}^*$. The mismatch factors can also be written in the following forms, which show more explicitly the mismatch properties:

$$M_{\text{in}} = 1 - \left|\frac{\Gamma_{\text{in}}-\Gamma_G^*}{1-\Gamma_{\text{in}}\Gamma_G}\right|^2\,, \quad M_{\text{out}} = 1 - \left|\frac{\Gamma_{\text{out}}-\Gamma_L^*}{1-\Gamma_{\text{out}}\Gamma_L}\right|^2 \tag{13.6.17}$$

These follow from the identity:

$$|1-\Gamma_1\Gamma_2|^2 - |\Gamma_1-\Gamma_2^*|^2 = (1-|\Gamma_1|^2)\,(1-|\Gamma_2|^2) \tag{13.6.18}$$

The transducer gain is *maximized* when the two-port is *simultaneously* conjugate matched, that is, when $\Gamma_{\text{in}} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\text{out}}^*$. Then, $M_{\text{in}} = M_{\text{out}} = 1$ and the three gains become equal. The common maximum gain achieved by simultaneous matching is called the *maximum available gain* (MAG):

$$G_{T,\max} = G_{a,\max} = G_{p,\max} = G_{\text{MAG}} \tag{13.6.19}$$

Simultaneous matching is discussed in Sec. 13.8. The necessary and sufficient condition for simultaneous matching is $K \geq 1$, where K is the Rollett stability factor. It can be shown that the MAG can be expressed as:

$$\boxed{G_{\text{MAG}} = \frac{|S_{21}|}{|S_{12}|}\left(K - \sqrt{K^2-1}\right)} \qquad \text{(maximum available gain)} \tag{13.6.20}$$

The *maximum stable gain* (MSG) is the maximum value G_{MAG} can have, which is achievable when $K = 1$:

$$\boxed{G_{\text{MSG}} = \frac{|S_{21}|}{|S_{12}|}} \qquad \text{(maximum stable gain)} \tag{13.6.21}$$

In the unilateral case, the MAG is obtained either by setting $\Gamma_G = \Gamma_{\text{in}}^* = S_{11}^*$ and $\Gamma_L = \Gamma_{\text{out}}^* = S_{22}^*$ in Eq. (13.6.14), or by a careful limiting process in Eq. (13.6.20), in which $K \to \infty$ so that both the numerator factor $K - \sqrt{K^2-1}$ and the denominator factor $|S_{12}|$ tend to zero. With either method, we find the unilateral MAG:

$$G_{\text{MAG},u} = \frac{|S_{21}|^2}{(1-|S_{11}|^2)(1-|S_{22}|^2)} = G_1 |S_{21}|^2 G_2 \qquad \text{(unilateral MAG)} \tag{13.6.22}$$

The maximum unilateral input and output gain factors are:

$$G_1 = \frac{1}{1-|S_{11}|^2}\,, \quad G_2 = \frac{1}{1-|S_{22}|^2} \tag{13.6.23}$$

They are the maxima of the input and output gain factors in Eq. (13.6.14) realized with conjugate matching, that is, with $\Gamma_G = S_{11}^*$ and $\Gamma_L = S_{22}^*$. For any other values of the reflection coefficients (such that $|\Gamma_G| < 1$ and $|\Gamma_L| < 1$), we have the following inequalities, which follow from the identity (13.6.18):

$$\frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2} \leq \frac{1}{1-|S_{11}|^2}\,, \qquad \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \leq \frac{1}{1-|S_{22}|^2} \tag{13.6.24}$$

Often two-ports, such as most microwave transistor amplifiers, are approximately unilateral, that is, the measured S-parameters satisfy $|S_{12}| \ll |S_{21}|$. To decide whether the two-port should be treated as unilateral, a figure of merit is used, which is essentially the comparison of the maximum unilateral gain to the transducer gain of the actual device under the same matching conditions, that is, $\Gamma_G = S_{11}^*$ and $\Gamma_L = S_{22}^*$.

For these matched values of Γ_G, Γ_L, the ratio of the bilateral and unilateral transducer gains can be shown to have the form:

$$g_u = \frac{G_T}{G_{Tu}} = \frac{1}{|1-U|^2}\,, \qquad U = \frac{S_{12}S_{21}S_{11}^*S_{22}^*}{(1-|S_{11}|^2)(1-|S_{22}|^2)} \tag{13.6.25}$$

The quantity $|U|$ is known as the *unilateral figure of merit*. If the relative gain ratio g_u is near unity (typically, within 10 percent of unity), the two-port may be treated as unilateral.

The MATLAB function `sgain` computes the transducer, available, and operating power gains, given the S-parameters and the reflection coefficients Γ_G, Γ_L. In addition,

it computes the unilateral gains, the maximum available gain, and the maximum stable gain. It also computes the unilateral figure of merit ratio (13.6.25). It has usage:

```
Gt = sgain(S,gG,gL);      transducer power gain at given ΓG, ΓL
Ga = sgain(S,gG,'a');     available power gain at given ΓG with ΓL = Γout*
Gp = sgain(S,gL,'p');     operating power gain at given ΓL with ΓG = Γin*

Gmag = sgain(S);          maximum available gain (MAG)
Gmsg = sgain(S,'msg');    maximum stable gain (MSG)

Gu = sgain(S,'u');        maximum unilateral gain, Eq. (13.6.22)
G1 = sgain(S,'ui');       maximum unilateral input gain, Eq. (13.6.23)
G2 = sgain(S,'uo');       maximum unilateral output gain, Eq. (13.6.23)
gu = sgain(S,'ufm');      unilateral figure of merit gain ratio, Eq. (13.6.25)
```

The MATLAB functions `gin` and `gout` compute the input and output reflection coefficients from S and Γ_G, Γ_L. They have usage:

```
Gin  = gin(S,gL);         input reflection coefficient, Eq. (13.4.3)
Gout = gout(S,gG);        output reflection coefficient, Eq. (13.4.6)
```

Example 13.6.1: A microwave transistor amplifier uses the Hewlett-Packard AT-41410 NPN bipolar transistor with the following S-parameters at 2 GHz [1355]:

$$S_{11} = 0.61\angle 165^\circ\,, \quad S_{21} = 3.72\angle 59^\circ\,, \quad S_{12} = 0.05\angle 42^\circ\,, \quad S_{22} = 0.45\angle -48^\circ$$

Calculate the input and output reflection coefficients and the various power gains, if the amplifier is connected to a generator and load with impedances $Z_G = 10 - 20j$ and $Z_L = 30 + 40j$ ohm.

Solution: The following MATLAB code will calculate all the required gains:

```
Z0 = 50;                                  % normalization impedance
ZG = 10+20j;  gG = z2g(ZG,Z0);            % ΓG = −0.50 + 0.50j = 0.71∠135°
ZL = 30-40j;  gL = z2g(ZL,Z0);            % ΓL = −0.41 − 0.43j = 0.59∠−133.15°

S = smat([0.61 165 3.72 59 0.05 42 0.45 -48]);      % reshape S into matrix

Gin = gin(S,gL);                          % Γin = 0.54∠162.30°
Gout = gout(S,gG);                        % Γout = 0.45∠−67.46°

Gt = sgain(S,gG,gL);                      % GT = 4.71, or, 6.73 dB
Ga = sgain(S,gG,'a');                     % Ga = 11.44, or, 10.58 dB
Gp = sgain(S,gL,'p');                     % Gp = 10.51, or, 10.22 dB

Gu = sgain(S,'u');                        % Gu = 27.64, or, 14.41 dB
G1 = sgain(S,'ui');                       % G1 = 1.59, or, 2.02 dB
G2 = sgain(S,'uo');                       % G2 = 1.25, or, 0.98 dB
gu = sgain(S,'ufm');                      % gu = 1.23, or, 0.89 dB

Gmag = sgain(S);                          % GMAG = 41.50, or, 16.18 dB
Gmsg = sgain(S,'msg');                    % GMSG = 74.40, or, 18.72 dB
```

The amplifier cannot be considered to be unilateral as the unilateral figure of merit ratio $g_u = 1.23$ is fairly large (larger than 10 percent from unity.)

The amplifier is operating at a gain of $G_T = 6.73$ dB, which is far from the maximum value of $G_{\text{MAG}} = 16.18$ dB. This is because it is mismatched with the given generator and load impedances.

To realize the optimum gain G_{MAG} the amplifier must 'see' certain optimum generator and load impedances or reflection coefficients. These can be calculated by the MATLAB function `smatch` and are found to be:

$$\Gamma_G = 0.82\angle -162.67^\circ \quad \Rightarrow \quad Z_G = \text{g2z}(Z_G, Z_0) = 5.12 - 7.54j\ \Omega$$
$$\Gamma_L = 0.75\angle 52.57^\circ \quad \Rightarrow \quad Z_L = \text{g2z}(Z_L, Z_0) = 33.66 + 91.48j\ \Omega$$

The design of such optimum matching terminations and the function `smatch` are discussed in Sec. 13.8. The functions `g2z` and `z2g` were discussed in Sec. 10.7 . □

13.7 Generalized S-Parameters and Power Waves

The practical usefulness of the S-parameters lies in the fact that the definitions (13.1.4) represent forward and backward traveling waves, which can be measured remotely by connecting a network analyzer to the two-port with transmission lines of characteristic impedance equal to the normalization impedance Z_0. This was depicted in Fig. 13.1.3.

A generalized definition of S-parameters and wave variables can be given by using in Eq. (13.1.4) two different normalization impedances for the input and output ports.

Anticipating that the two-port will be connected to a generator and load of impedances Z_G and Z_L, a particularly convenient choice is to use Z_G for the input normalization impedance and Z_L for the output one, leading to the definition of the *power waves* (as opposed to traveling waves) [982–984,986]:

$$\boxed{\begin{aligned} a'_1 &= \frac{V_1 + Z_G I_1}{2\sqrt{R_G}} \\ b'_1 &= \frac{V_1 - Z_G^* I_1}{2\sqrt{R_G}} \end{aligned}} \qquad \boxed{\begin{aligned} a'_2 &= \frac{V_2 - Z_L I_2}{2\sqrt{R_L}} \\ b'_2 &= \frac{V_2 + Z_L^* I_2}{2\sqrt{R_L}} \end{aligned}} \qquad \text{(power waves)} \tag{13.7.1}$$

We note that the b-waves involve the complex-conjugates of the impedances. The quantities R_G, R_L are the resistive parts of Z_G, Z_L and are assumed to be positive. These definitions reduce to the conventional traveling ones if $Z_G = Z_L = Z_0$.

These "wave" variables can no longer be interpreted as incoming and outgoing waves from the two sides of the two-port. However, as we see below, they have a nice interpretation in terms of power transfer to and from the two-port and simplify the expressions for the power gains. Inverting Eqs. (13.7.1), we have:

$$\boxed{\begin{aligned} V_1 &= \frac{1}{\sqrt{R_G}}\left(Z_G^* a'_1 + Z_G b'_1\right) \\ I_1 &= \frac{1}{\sqrt{R_G}}\left(a'_1 - b'_1\right) \end{aligned}} \qquad \boxed{\begin{aligned} V_2 &= \frac{1}{\sqrt{R_L}}\left(Z_L^* a'_2 + Z_L b'_2\right) \\ I_2 &= \frac{1}{\sqrt{R_L}}\left(b'_2 - a'_2\right) \end{aligned}} \tag{13.7.2}$$

The power waves can be related directly to the traveling waves. For example, expressing Eqs. (13.7.1) and (13.1.5) in matrix form, we have for port-1:

$$\begin{bmatrix} a'_1 \\ b'_1 \end{bmatrix} = \frac{1}{2\sqrt{R_G}} \begin{bmatrix} 1 & Z_G \\ 1 & -Z_G^* \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}, \quad \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \frac{1}{\sqrt{Z_0}} \begin{bmatrix} Z_0 & Z_0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$$

It follows that:

$$\begin{bmatrix} a'_1 \\ b'_1 \end{bmatrix} = \frac{1}{2\sqrt{R_G Z_0}} \begin{bmatrix} 1 & Z_G \\ 1 & -Z_G^* \end{bmatrix} \begin{bmatrix} Z_0 & Z_0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \quad \text{or,}$$

$$\begin{bmatrix} a'_1 \\ b'_1 \end{bmatrix} = \frac{1}{2\sqrt{R_G Z_0}} \begin{bmatrix} Z_0 + Z_G & Z_0 - Z_G \\ Z_0 - Z_G^* & Z_0 + Z_G^* \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \tag{13.7.3}$$

The entries of this matrix can be expressed directly in terms of the reflection coefficient Γ_G. Using the identities of Problem 13.3, we may rewrite Eq. (13.7.3) and its inverse as follows::

$$\begin{aligned} \begin{bmatrix} a'_1 \\ b'_1 \end{bmatrix} &= \frac{1}{\sqrt{1-|\Gamma_G|^2}} \begin{bmatrix} e^{j\phi_G} & -\Gamma_G e^{j\phi_G} \\ -\Gamma_G^* e^{-j\phi_G} & e^{-j\phi_G} \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \\ \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} &= \frac{1}{\sqrt{1-|\Gamma_G|^2}} \begin{bmatrix} e^{-j\phi_G} & \Gamma_G e^{j\phi_G} \\ \Gamma_G^* e^{-j\phi_G} & e^{j\phi_G} \end{bmatrix} \begin{bmatrix} a'_1 \\ b'_1 \end{bmatrix} \end{aligned} \tag{13.7.4}$$

where, noting that the quantity $|1-\Gamma_G|/(1-\Gamma_G)$ is a pure phase factor, we defined:

$$\Gamma_G = \frac{Z_G - Z_0}{Z_G + Z_0}, \quad e^{j\phi_G} = \frac{|1-\Gamma_G|}{1-\Gamma_G} = \frac{1-\Gamma_G^*}{|1-\Gamma_G|} \tag{13.7.5}$$

Similarly, we have for the power and traveling waves at port-2:

$$\begin{aligned} \begin{bmatrix} a'_2 \\ b'_2 \end{bmatrix} &= \frac{1}{\sqrt{1-|\Gamma_L|^2}} \begin{bmatrix} e^{j\phi_L} & -\Gamma_L e^{j\phi_L} \\ -\Gamma_L^* e^{-j\phi_L} & e^{-j\phi_L} \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \\ \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} &= \frac{1}{\sqrt{1-|\Gamma_L|^2}} \begin{bmatrix} e^{-j\phi_L} & \Gamma_L e^{j\phi_L} \\ \Gamma_L^* e^{-j\phi_L} & e^{j\phi_L} \end{bmatrix} \begin{bmatrix} a'_2 \\ b'_2 \end{bmatrix} \end{aligned} \tag{13.7.6}$$

where

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0}, \quad e^{j\phi_L} = \frac{|1-\Gamma_L|}{1-\Gamma_L} = \frac{1-\Gamma_L^*}{|1-\Gamma_L|} \tag{13.7.7}$$

The generalized S-parameters are the scattering parameters with respect to the power wave variables, that is,

$$\begin{bmatrix} b'_1 \\ b'_2 \end{bmatrix} = \begin{bmatrix} S'_{11} & S'_{12} \\ S'_{21} & S'_{22} \end{bmatrix} \begin{bmatrix} a'_1 \\ a'_2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{b}' = S'\mathbf{a}' \tag{13.7.8}$$

To relate S' to the conventional scattering matrix S, we define the following diagonal matrices:

$$\Gamma = \begin{bmatrix} \Gamma_G & 0 \\ 0 & \Gamma_L \end{bmatrix}, \quad F = \begin{bmatrix} \dfrac{e^{j\phi_G}}{\sqrt{1-|\Gamma_G|^2}} & 0 \\ 0 & \dfrac{e^{j\phi_L}}{\sqrt{1-|\Gamma_L|^2}} \end{bmatrix} = \begin{bmatrix} F_G & 0 \\ 0 & F_L \end{bmatrix} \tag{13.7.9}$$

Using these matrices, it follows from Eqs. (13.7.4) and (13.7.6):

$$\begin{aligned} a'_1 &= F_G(a_1 - \Gamma_G b_1) \\ a'_2 &= F_L(a_2 - \Gamma_L b_2) \end{aligned} \quad \Rightarrow \quad \mathbf{a}' = F(\mathbf{a} - \Gamma\mathbf{b}) \tag{13.7.10}$$

$$\begin{aligned} b'_1 &= F_G^*(b_1 - \Gamma_G^* a_1) \\ b'_2 &= F_L^*(b_2 - \Gamma_L^* a_2) \end{aligned} \quad \Rightarrow \quad \mathbf{b}' = F^*(\mathbf{b} - \Gamma^*\mathbf{a}) \tag{13.7.11}$$

Using $\mathbf{b} = S\mathbf{a}$, we find

$$\begin{aligned} \mathbf{a}' &= F(\mathbf{a} - \Gamma\mathbf{b}) = F(I - \Gamma S)\mathbf{a} \quad \Rightarrow \quad \mathbf{a} = (I - \Gamma S)^{-1}F^{-1}\mathbf{a}' \\ \mathbf{b}' &= F^*(S - \Gamma^*)\mathbf{a} = F^*(S - \Gamma^*)(I - \Gamma S)^{-1}F^{-1}\mathbf{a}' = S'\mathbf{a}' \end{aligned}$$

where I is the 2×2 unit matrix. Thus, the generalized S-matrix is:

$$\boxed{S' = F^*(S - \Gamma^*)(I - \Gamma S)^{-1}F^{-1}} \tag{13.7.12}$$

We note that $S' = S$ when $Z_G = Z_L = Z_0$, that is, when $\Gamma_G = \Gamma_L = 0$. The explicit expressions for the matrix elements of S' can be derived as follows:

$$\begin{aligned} S'_{11} &= \frac{(S_{11} - \Gamma_G^*)(1 - S_{22}\Gamma_L) + S_{21}S_{12}\Gamma_L}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-2j\phi_G} \\ S'_{22} &= \frac{(S_{22} - \Gamma_L^*)(1 - S_{11}\Gamma_G) + S_{21}S_{12}\Gamma_G}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-2j\phi_L} \end{aligned} \tag{13.7.13a}$$

$$\begin{aligned} S'_{21} &= \frac{\sqrt{1-|\Gamma_G|^2}\, S_{21}\sqrt{1-|\Gamma_L|^2}}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-j(\phi_G + \phi_L)} \\ S'_{12} &= \frac{\sqrt{1-|\Gamma_L|^2}\, S_{12}\sqrt{1-|\Gamma_G|^2}}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-j(\phi_L + \phi_G)} \end{aligned} \tag{13.7.13b}$$

The S'_{11}, S'_{22} parameters can be rewritten in terms of the input and output reflection coefficients by using Eq. (13.13.2) and the following factorization identities:

$$\begin{aligned} (S_{11} - \Gamma_G^*)(1 - S_{22}\Gamma_L) + S_{21}S_{12}\Gamma_L &= (\Gamma_{\text{in}} - \Gamma_G^*)(1 - S_{22}\Gamma_L) \\ (S_{22} - \Gamma_L^*)(1 - S_{11}\Gamma_G) + S_{21}S_{12}\Gamma_G &= (\Gamma_{\text{out}} - \Gamma_L^*)(1 - S_{11}\Gamma_G) \end{aligned}$$

It then follows from Eq. (13.7.13) that:

$$S'_{11} = \frac{\Gamma_{\text{in}} - \Gamma_G^*}{1 - \Gamma_{\text{in}}\Gamma_G}\, e^{-2j\phi_G}, \quad S'_{22} = \frac{\Gamma_{\text{out}} - \Gamma_L^*}{1 - \Gamma_{\text{out}}\Gamma_L}\, e^{-2j\phi_L} \tag{13.7.14}$$

Therefore, the mismatch factors (13.6.17) are recognized to be:

$$M_G = 1 - |S'_{11}|^2, \quad M_L = 1 - |S'_{22}|^2 \tag{13.7.15}$$

The power flow relations (13.2.1) into and out of the two-port are also valid in terms of the power wave variables. Using Eq. (13.7.2), it can be shown that:

$$\begin{aligned} P_{\text{in}} &= \frac{1}{2}\,\text{Re}[V_1^* I_1] = \frac{1}{2}|a'_1|^2 - \frac{1}{2}|b'_1|^2 \\ P_L &= \frac{1}{2}\,\text{Re}[V_2^* I_2] = \frac{1}{2}|b'_2|^2 - \frac{1}{2}|a'_2|^2 \end{aligned} \tag{13.7.16}$$

In the definitions (13.7.1), the impedances Z_G, Z_L are arbitrary normalization parameters. However, if the two-port is actually connected to a generator V_G with impedance Z_G and a load Z_L, then the power waves take particularly simple forms.

It follows from Fig. 13.1.4 that $V_G = V_1 + Z_G I_1$ and $V_2 = Z_L I_2$. Therefore, definitions Eq. (13.7.1) give:

$$\begin{aligned} a'_1 &= \frac{V_1 + Z_G I_1}{2\sqrt{R_G}} = \frac{V_G}{2\sqrt{R_G}} \\ a'_2 &= \frac{V_2 - Z_L I_2}{2\sqrt{R_L}} = 0 \\ b'_2 &= \frac{V_2 + Z_L^* I_2}{2\sqrt{R_L}} = \frac{Z_L + Z_L^*}{2\sqrt{R_L}}\, I_2 = \frac{2R_L}{2\sqrt{R_L}}\, I_2 = \sqrt{R_L}\, I_2 \end{aligned} \tag{13.7.17}$$

It follows that the available power from the generator and the power delivered to the load are given simply by:

$$\begin{aligned} P_{\text{avG}} &= \frac{|V_G|^2}{8R_G} = \frac{1}{2}|a'_1|^2 \\ P_L &= \frac{1}{2}R_L|I_2|^2 = \frac{1}{2}|b'_2|^2 \end{aligned} \tag{13.7.18}$$

Because $a'_2 = 0$, the generalized scattering matrix gives, $b'_1 = S'_{11}a'_1$ and $b'_2 = S'_{21}a'_1$. The power expressions (13.7.16) then become:

$$\begin{aligned} P_{\text{in}} &= \frac{1}{2}|a'_1|^2 - \frac{1}{2}|b'_1|^2 = (1 - |S'_{11}|^2)\frac{1}{2}|a'_1|^2 = (1 - |S'_{11}|^2)P_{\text{avG}} \\ P_L &= \frac{1}{2}|b'_2|^2 - \frac{1}{2}|a'_2|^2 = \frac{1}{2}|b'_2|^2 = |S'_{21}|^2\frac{1}{2}|a'_1|^2 = |S'_{21}|^2 P_{\text{avG}} \end{aligned} \tag{13.7.19}$$

It follows that the transducer and operating power gains are:

$$G_T = \frac{P_L}{P_{\text{avG}}} = |S'_{21}|^2, \quad G_p = \frac{P_L}{P_{\text{in}}} = \frac{|S'_{21}|^2}{1 - |S'_{11}|^2} \tag{13.7.20}$$

These also follow from the explicit expressions (13.7.13) and Eqs. (13.6.10) and (13.6.11). We can also express the available power gain in terms of the generalized S-parameters, that is, $G_a = |S'_{21}|^2/(1 - |S'_{22}|^2)$. Thus, we summarize:

$$\boxed{G_T = |S'_{21}|^2\,, \quad G_a = \frac{|S'_{21}|^2}{1-|S'_{22}|^2}\,, \quad G_p = \frac{|S'_{21}|^2}{1-|S'_{11}|^2}} \tag{13.7.21}$$

When the load and generator are matched to the network, that is, $\Gamma_{\rm in} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\rm out}^*$, the generalized reflections coefficients vanish, $S'_{11} = S'_{22} = 0$, making all the gains equal to each other.

13.8 Simultaneous Conjugate Matching

We saw that the transducer, available, and operating power gains become equal to the maximum available gain $G_{\rm MAG}$ when both the generator and the load are conjugately matched to the two-port, that is, $\Gamma_{\rm in} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\rm out}^*$. Using Eq. (13.5.8), these conditions read explicitly:

$$\begin{aligned} \Gamma_G^* &= S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-S_{22}\Gamma_L} = \frac{S_{11}-\Delta\Gamma_L}{1-S_{22}\Gamma_L} \\ \Gamma_L^* &= S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1-S_{22}\Gamma_G} = \frac{S_{22}-\Delta\Gamma_G}{1-S_{11}\Gamma_G} \end{aligned} \tag{13.8.1}$$

Assuming a bilateral two-port, Eqs. (13.8.1) can be solved in the two unknowns Γ_G, Γ_L (eliminating one of the unknowns gives a quadratic equation for the other.) The resulting solutions can be expressed in terms of the parameters (13.5.1):

$$\boxed{\begin{aligned} \Gamma_G &= \frac{B_1 \mp \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \\ \Gamma_L &= \frac{B_2 \mp \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \end{aligned}} \quad \text{(simultaneous conjugate match)} \tag{13.8.2}$$

where the minus signs are used when $B_1 > 0$ and $B_2 > 0$, and the plus signs, otherwise.

A necessary and sufficient condition for these solutions to have magnitudes $|\Gamma_G| < 1$ and $|\Gamma_L| < 1$ is that the Rollett stability factor be greater than unity, $K > 1$. This is satisfied when the two-port is unconditionally stable, which implies that $K > 1$ and $B_1 > 0$, $B_2 > 0$.

A conjugate match exists also when the two-port is potentially unstable, but with $K > 1$. Necessarily, this means that $B_1 < 0$, $B_2 < 0$, and also $|\Delta| > 1$. Such cases are rare in practice. For example, most microwave transistors have either $K > 1$ and are stable, or, they are potentially unstable with $K < 1$ and $|\Delta| < 1$.

If the two-port is unilateral, $S_{12} = 0$, then the two equations (13.8.1) decouple, so that the optimum conjugately matched terminations are:

$$\boxed{\Gamma_G = S_{11}^*\,, \quad \Gamma_L = S_{22}^*} \quad \text{(unilateral conjugate match)} \tag{13.8.3}$$

The MATLAB function `smatch` implements Eqs. (13.8.2). It works only if $K > 1$. Its usage is as follows:


```
[gG,gL] = smatch(S);          % conjugate matched terminations Γ_G, Γ_L
```

To realize such optimum conjugately matched terminations, matching networks must be used at the input and output of the two-port as shown in Fig. 13.8.1.

The input matching network can be thought as being effectively connected to the impedance $Z_{\rm in} = Z_G^*$ at its output terminals. It must transform $Z_{\rm in}$ into the actual impedance of the connected generator, typically, $Z_0 = 50$ ohm.

The output matching network must transform the actual load impedance, here Z_0, into the optimum load impedance $Z_L = Z_{\rm out}^*$.

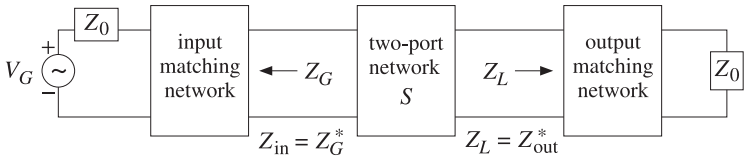


Fig. 13.8.1 Input and output matching networks.

The matching networks may be realized in several possible ways, as discussed in Chap. 12. Stub matching, quarter-wavelength matching, or lumped L-section or Π-section networks may be used. In designing the matching networks, it proves convenient to first design the reverse network as mentioned in Sec. 12.13.

Fig. 13.8.2 shows the procedure for designing the output matching network using a reversed stub matching transformer or a reversed quarter-wave transformer with a parallel stub. In both cases the reversed network is designed to transform the load impedance Z_L^* into Z_0.

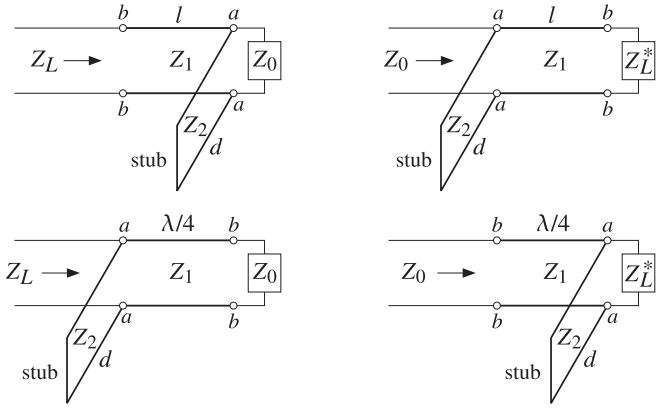


Fig. 13.8.2 Two types of output matching networks and their reversed networks..

Example 13.8.1: A microwave transistor amplifier uses the Hewlett-Packard AT-41410 NPN bipolar transistor having S-parameters at 2 GHz [1355]:

$$S_{11} = 0.61\angle 165^\circ, \quad S_{21} = 3.72\angle 59^\circ, \quad S_{12} = 0.05\angle 42^\circ, \quad S_{22} = 0.45\angle -48^\circ$$

Determine the optimum conjugately matched source and load terminations, and design appropriate input and output matching networks.

Solution: This is the continuation of Example 13.6.1. The transistor is stable with $K = 1.1752$ and $|\Delta| = 0.1086$. The function `smatch` gives:

$$[\Gamma_G, \Gamma_L] = \text{smatch}(S) \quad \Rightarrow \quad \Gamma_G = 0.8179\angle -162.6697^\circ, \quad \Gamma_L = 0.7495\angle 52.5658^\circ$$

The corresponding source, load, input, and output impedances are (with $Z_0 = 50$):

$$Z_G = Z_{\text{in}}^* = 5.1241 - 7.5417j\ \Omega, \quad Z_L = Z_{\text{out}}^* = 33.6758 + 91.4816j\ \Omega$$

The locations of the optimum reflection coefficients on the Smith chart are shown in Fig. 13.8.3. For comparison, the unilateral solutions of Eq. (13.8.3) are also shown.

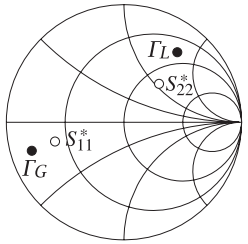


Fig. 13.8.3 Optimum load and source reflection coefficients.

We consider three types of matching networks: (a) microstrip single-stub matching networks with open shunt stubs, shown in Fig. 13.8.4, (b) microstrip quarter-wavelength matching networks with open $\lambda/8$ or $3\lambda/8$ stubs, shown in Fig. 13.8.5, and (c) L-section matching networks, shown in 13.8.6.

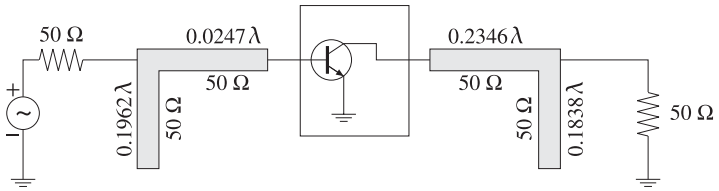


Fig. 13.8.4 Input and output stub matching networks.

In Fig. 13.8.4, the input stub must transform Z_{in} to Z_0. It can be designed with the help of the function `stub1`, which gives the two solutions:

$$dl = \text{stub1}(Z_{\text{in}}/Z_0, \texttt{'po'}) = \begin{bmatrix} 0.3038 & 0.4271 \\ 0.1962 & 0.0247 \end{bmatrix}$$

We choose the lower one, which has the shortest lengths. Thus, the stub length is $d = 0.1962\lambda$ and the segment length $l = 0.0247\lambda$. Both segments can be realized with microstrips of characteristic impedance $Z_0 = 50$ ohm. Similarly, the output matching network can be designed by:

$$dl = \text{stub1}(Z_{\text{out}}/Z_0, \texttt{'po'}) = \begin{bmatrix} 0.3162 & 0.1194 \\ 0.1838 & 0.2346 \end{bmatrix}$$

Again, we choose the lower solutions, $d = 0.1838\lambda$ and $l = 0.2346\lambda$. The solutions using shorted shunt stubs are:

$$\text{stub1}(Z_{\text{in}}/Z_0) = \begin{bmatrix} 0.0538 & 0.4271 \\ 0.4462 & 0.0247 \end{bmatrix}, \quad \text{stub1}(Z_{\text{out}}/Z_0) = \begin{bmatrix} 0.0662 & 0.1194 \\ 0.4338 & 0.2346 \end{bmatrix}$$

Using microstrip lines with alumina substrate ($\epsilon_r = 9.8$), we obtain the following values for the width-to-height ratio, effective permittivity, and wavelength:

$$u = \frac{w}{h} = \text{mstripr}(\epsilon_r, Z_0) = 0.9711$$

$$\epsilon_{\text{eff}} = \text{mstripa}(\epsilon_r, u) = 6.5630$$

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}} = 5.8552 \text{ cm}$$

where $\lambda_0 = 15$ cm is the free-space wavelength at 2 GHz. It follows that the actual segment lengths are $d = 1.1486$ cm, $l = 0.1447$ cm for the input network, and $d = 1.0763$ cm, $l = 1.3734$ cm for the output network.

In the quarter-wavelength method shown in Fig. 13.8.5, we use the function `qwt2` to carry out the design of the required impedances of the microstrip segments. We have for the input and output networks:

$$[Z_1, Z_2] = \text{qwt2}(Z_{\text{in}}, Z_0) = [28.4817, -11.0232]\ \Omega$$

$$[Z_1, Z_2] = \text{qwt2}(Z_{\text{out}}, Z_0) = [118.7832, 103.8782]\ \Omega$$

For the input case, we find $Z_2 = -11.0232\ \Omega$, which means that we should use either a $3\lambda/8$-shorted stub or a $\lambda/8$-opened one. We choose the latter. Similarly, for the output case, we have $Z_2 = 103.8782\ \Omega$, and we choose a $3\lambda/8$-opened stub. The parameters of each microstrip segment are:

$$\begin{array}{llllll}
Z_1 = & 28.4817\ \Omega, & u = 2.5832, & \epsilon_{\text{eff}} = 7.2325, & \lambda = 5.578 \text{ cm}, & \lambda/4 = 1.394 \text{ cm} \\
Z_2 = & 11.0232\ \Omega, & u = 8.9424, & \epsilon_{\text{eff}} = 8.2974, & \lambda = 5.207 \text{ cm}, & \lambda/8 = 0.651 \text{ cm} \\
Z_1 = & 118.7832\ \Omega, & u = 0.0656, & \epsilon_{\text{eff}} = 5.8790, & \lambda = 6.186 \text{ cm}, & \lambda/4 = 1.547 \text{ cm} \\
Z_2 = & 103.8782\ \Omega, & u = 0.1169, & \epsilon_{\text{eff}} = 7.9503, & \lambda = 6.149 \text{ cm}, & 3\lambda/8 = 2.306 \text{ cm}
\end{array}$$

Finally, the designs using L-sections shown in Fig. 13.8.6, can be carried out with the help of the function `lmatch`. We have the dual solutions for the input and output networks:

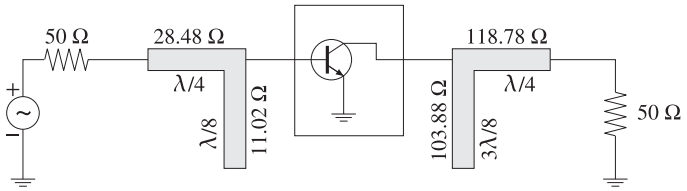


Fig. 13.8.5 Quarter-wavelength matching networks with $\lambda/8$-stubs.

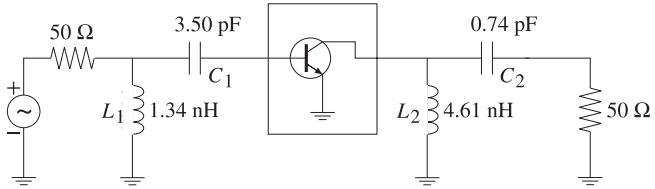


Fig. 13.8.6 Input and output matching with L-sections.

$$[X_1, X_2] = \text{lmatch}(Z_0, Z_{\text{in}}, \texttt{'n'}) = \begin{bmatrix} 16.8955 & -22.7058 \\ -16.8955 & 7.6223 \end{bmatrix}$$

$$[X_1, X_2] = \text{lmatch}(Z_{\text{out}}, Z_0, \texttt{'n'}) = \begin{bmatrix} 57.9268 & -107.7472 \\ 502.4796 & 7.6223 \end{bmatrix}$$

According to the usage of `lmatch`, the output network transforms Z_0 into Z_{out}^*, but that is equal to Z_L as required.

Choosing the first rows as the solutions in both cases, the shunt part X_1 will be inductive and the series part X_2, capacitive. At 2 GHz, we find the element values:

$$L_1 = \frac{X_1}{\omega} = 1.3445 \text{ nH}, \quad C_1 = -\frac{1}{\omega X_2} = 3.5047 \text{ pF}$$
$$L_2 = \frac{X_1}{\omega} = 4.6097 \text{ nH}, \quad C_2 = -\frac{1}{\omega X_2} = 0.7386 \text{ pF}$$

The output network, but not the input one, also admits a reversed L-section solution:

$$[X_1, X_2] = \text{lmatch}(Z_{\text{out}}, Z_0, \texttt{'r'}) = \begin{bmatrix} 71.8148 & 68.0353 \\ -71.8148 & 114.9280 \end{bmatrix}$$

The essential MATLAB code used to generate the above results was as follows:

```
Z0 = 50; f = 2; w=2*pi*f; la0 = 30/f; er = 9.8;      % f in GHz

S = smat([0.61 165 3.72 59 0.05 42 0.45 -48]);       % S-matrix

[gG,gL] = smatch(S);                                 % simultaneous conjugate match
```



```
smith;                                               % draw Fig. 13.8.3
plot(gG, '.'); plot(conj(S(1,1)), 'o');
plot(gL, '.'); plot(conj(S(2,2)), 'o');

ZG = g2z(gG,Z0);  Zin  = conj(ZG);
ZL = g2z(gL,Z0);  Zout = conj(ZL);

dl = stub1(Zin/Z0, 'po');                            % single-stub design
dl = stub1(Zout/Z0, 'po');

u = mstripr(er,Z0);                                  % microstrip w/h ratio
eff = mstripa(er,u);                                 % effective permittivity
la = la0/sqrt(eff);                                  % wavelength within microstrip

[Z1,Z2] = qwt2(Zin, Z0);                             % quarter-wavelength with λ/8 stub
[Z1,Z2] = qwt2(Zout, Z0);

X12 = lmatch(Z0,Zin,'n');  L1 = X12(1,1)/w;  C1 = -1/(w * X12(1,2))*1e3;
X12 = lmatch(Zout,Z0,'n'); L2 = X12(1,1)/w;  C2 = -1/(w * X12(1,2))*1e3;
X12 = lmatch(Zout,Z0,'r');                           % L,C in units of nH and pF
```

One could replace the stubs with balanced stubs, as discussed in Sec. 12.9, or use Π- or T-sections instead of L-sections. □

13.9 Power Gain Circles

For a stable two-port, the maximum transducer gain is achieved at single pair of points Γ_G, Γ_L. When the gain G is required to be less than G_{MAG}, there will be many possible pairs Γ_G, Γ_L at which the gain G is realized. The locus of such points Γ_G and Γ_L on the Γ-plane is typically a circle of the form:

$$|\Gamma - c| = r \tag{13.9.1}$$

where c, r are the center and radius of the circle and depend on the desired value of the gain G.

In practice, several types of such circles are used, such as unilateral, operating, and available power gain circles, as well as constant noise figure circles, constant SWR circles, and others.

The gain circles allow one to select appropriate values for Γ_G, Γ_L that, in addition to providing the desired gain, also satisfy other requirements, such as striking a balance between minimizing the noise figure and maximizing the gain.

The MATLAB function `sgcirc` calculates the stability circles as well as the operating, available, and unilateral gain circles. Its complete usage is:

```
[c,r] = sgcirc(S,'s');          % source stability circle
[c,r] = sgcirc(S,'l');          % load stability circle
[c,r] = sgcirc(S,'p',G);        % operating power gain circle
[c,r] = sgcirc(S,'a',G);        % available power gain circle
[c,r] = sgcirc(S,'ui',G);       % unilateral input gain circle
[c,r] = sgcirc(S,'uo',G);       % unilateral output gain circle
```

where in the last four cases G is the desired gain in dB.

13.10 Unilateral Gain Circles

We consider only the unconditionally stable unilateral case, which has $|S_{11}| < 1$ and $|S_{22}| < 1$. The dependence of the transducer power gain on Γ_G and Γ_L decouples and the value of the gain may be adjusted by separately choosing Γ_G and Γ_L. We have from Eq. (13.6.14):

$$G_T = \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} = G_G\,|S_{21}|^2\,G_L \tag{13.10.1}$$

The input and output gain factors G_G, G_L satisfy the inequalities (13.6.24). Concentrating on the output gain factor, the corresponding gain circle is obtained as the locus of points Γ_L that will lead to a fixed value, say $G_L = G$, which necessarily must be less than the maximum G_2 given in Eq. (13.6.23), that is,

$$\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} = G \le G_2 = \frac{1}{1-|S_{22}|^2} \tag{13.10.2}$$

Normalizing the gain G to its maximum value $g = G/G_2 = G(1-|S_{22}|^2)$, we may rewrite (13.10.2) in the form:

$$\frac{(1-|\Gamma_L|^2)(1-|S_{22}|^2)}{|1-S_{22}\Gamma_L|^2} = g \le 1 \tag{13.10.3}$$

This equation can easily be rearranged into the equation of a circle $|\Gamma_L - c| = r$, with center and radius given by:

$$\boxed{c = \frac{gS_{22}^*}{1-(1-g)|S_{22}|^2}\,,\quad r = \frac{\sqrt{1-g}\,(1-|S_{22}|^2)}{1-(1-g)|S_{22}|^2}} \tag{13.10.4}$$

When $g = 1$ or $G = G_2$, the gain circle collapses onto a single point, that is, the optimum point $\Gamma_L = S_{22}^*$. Similarly, we find for the constant gain circles of the input gain factor:

$$\boxed{c = \frac{gS_{11}^*}{1-(1-g)|S_{11}|^2}\,,\quad r = \frac{\sqrt{1-g}\,(1-|S_{11}|^2)}{1-(1-g)|S_{11}|^2}} \tag{13.10.5}$$

where here, $g = G/G_1 = G(1-|S_{11}|^2)$ and the circles are $|\Gamma_G - c| = r$.

Both sets of c, r satisfy the conditions $|c| < 1$ and $|c| + r < 1$, the latter implying that the circles lie entirely within the unit circle $|\Gamma| < 1$, that is, within the Smith chart.

Example 13.10.1: A unilateral microwave transistor has S-parameters:

$$S_{11} = 0.8\angle 120^\circ,\quad S_{21} = 4\angle 60^\circ,\quad S_{12} = 0,\quad S_{22} = 0.2\angle -30^\circ$$

The unilateral MAG and the maximum input and output gains are obtained as follows:

$$G_{\text{MAG},u} = \text{sgain}(S,\texttt{'u'}) = 16.66\text{ dB}$$

$$G_1 = \text{sgain}(S,\texttt{'ui'}) = 4.44\text{ dB}$$

$$G_2 = \text{sgain}(S,\texttt{'uo'}) = 0.18\text{ dB}$$

Most of the gain is accounted for by the factor $|S_{21}|^2$, which is 12.04 dB. The constant input gain circles for $G_G = 1, 2, 3$ dB are shown in Fig. 13.10.1. Their centers lie along the ray to S_{11}^*. For example, the center and radius of the 3-dB case were computed by

$$[c_3, r_3] = \text{sgcirc}(S,\texttt{'ui'},3) \quad\Rightarrow\quad c_3 = 0.701\angle -120^\circ,\quad r_3 = 0.233$$

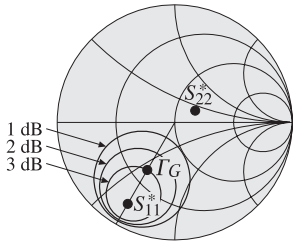


Fig. 13.10.1 Unilateral input gain circles.

Because the output does not provide much gain, we may choose the optimum value $\Gamma_L = S_{22}^* = 0.2\angle 30^\circ$. Then, with any point Γ_G along the 3-dB input gain circle the total transducer gain will be in dB:

$$G_T = G_G + |S_{21}|^2 + G_L = 3 + 12.04 + 0.18 = 15.22\text{ dB}$$

Points along the 3-dB circle are parametrized as $\Gamma_G = c_3 + r_3 e^{j\phi}$, where ϕ is any angle. Choosing $\phi = \arg(S_{11}^*) - \pi$ will correspond to the point on the circle that lies closest to the origin, that is, $\Gamma_G = 0.468\angle -120^\circ$, as shown in Fig. 13.10.1. The corresponding generator and load impedances will be:

$$Z_G = 69.21 + 14.42j\ \Omega,\quad Z_L = 23.15 - 24.02j\ \Omega$$

The MATLAB code used to generate these circles was:

```
S = smat([0.8, 120, 4, 60, 0, 0, 0.2, -30]);

[c1,r1] = sgcirc(S,'ui',1);
[c2,r2] = sgcirc(S,'ui',2);
[c3,r3] = sgcirc(S,'ui',3);

smith;  smithcir(c1,r1);  smithcir(c2,r2);  smithcir(c3,r3);

c = exp(-j*angle(S(1,1))); line([0,real(c)], [0,imag(c)]);

gG = c3 - r3*exp(j*angle(c3));

plot(conj(S(1,1)),'.');  plot(conj(S(2,2)),'.');  plot(gG,'.');
```

The input and output matching networks can be designed using open shunt stubs as in Fig. 13.8.4. The stub lengths are found to be (with $Z_0 = 50\ \Omega$):

$$dl = \text{stub1}(Z_G^*/Z_0, \text{'po'}) = \begin{bmatrix} 0.3704 & 0.3304 \\ 0.1296 & 0.0029 \end{bmatrix}$$

$$dl = \text{stub1}(Z_L^*/Z_0, \text{'po'}) = \begin{bmatrix} 0.4383 & 0.0994 \\ 0.0617 & 0.3173 \end{bmatrix}$$

Choosing the shortest lengths, we have for the input network $d = 0.1296\lambda$, $l = 0.0029\lambda$, and for the output network, $d = 0.0617\lambda$, $l = 0.3173\lambda$. Fig. 13.10.2 depicts the complete matching circuit. □

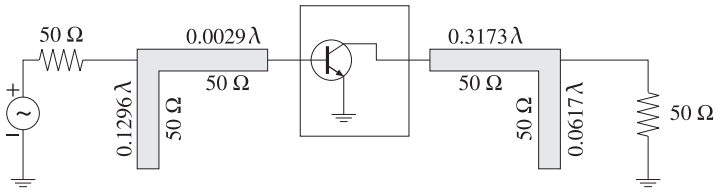


Fig. 13.10.2 Input and output stub matching networks.

13.11 Operating and Available Power Gain Circles

Because the transducer power gain G_T depends on two independent parameters—the source and load reflection coefficients—it is difficult to find the simultaneous locus of points for Γ_G, Γ_L that will result in a given value for the gain.

If the generator is matched, $\Gamma_{\text{in}} = \Gamma_G^*$, then the transducer gain becomes equal to the operating gain $G_T = G_p$ and depends only on the load reflection coefficient Γ_L. The locus of points Γ_L that result in fixed values of G_p are the *operating power gain circles*. Similarly, the *available power gain circles* are obtained by matching the load end, $\Gamma_L = \Gamma_{\text{out}}^*$, and varying Γ_G to achieve fixed values of the available power gain.

Using Eqs. (13.6.11) and (13.5.8), the conditions for achieving a constant value, say G, for the operating or the available power gains are:

$$\begin{aligned} G_p &= \frac{1}{1-|\Gamma_{\text{in}}|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} = G, \qquad \Gamma_G^* = \Gamma_{\text{in}} = \frac{S_{11}-\Delta\Gamma_L}{1-S_{22}\Gamma_L} \\ G_a &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1}{1-|\Gamma_{\text{out}}|^2} = G, \qquad \Gamma_L^* = \Gamma_{\text{out}} = \frac{S_{22}-\Delta\Gamma_G}{1-S_{11}\Gamma_G} \end{aligned} \tag{13.11.1}$$

We consider the operating gain first. Defining the normalized gain $g = G/|S_{21}|^2$, substituting Γ_{in}, and using the definitions (13.5.1), we obtain the condition:

$$\begin{aligned} g &= \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2-|S_{11}-\Delta\Gamma_L|^2} \\ &= \frac{1-|\Gamma_L|^2}{(|S_{22}|^2-|\Delta|^2)|\Gamma_L|^2-(S_{22}-\Delta S_{11}^*)\Gamma_L-(S_{22}^*-\Delta^* S_{11})\Gamma_L^*+1-|S_{11}|^2} \\ &= \frac{1-|\Gamma_L|^2}{D_2|\Gamma_L|^2-C_2\Gamma_L-C_2^*\Gamma_L^*+1-|S_{11}|^2} \end{aligned}$$

This can be rearranged into the form:

$$|\Gamma_L|^2 - \frac{gC_2}{1+gD_2}\Gamma_L - \frac{gC_2^*}{1+gD_2}\Gamma_L^* = \frac{1-g(1-|S_{11}|^2)}{1+gD_2}$$

and then into the circle form:

$$\left|\Gamma_L - \frac{gC_2^*}{1+gD_2}\right|^2 = \frac{g^2|C_2|^2}{(1+gD_2)^2} + \frac{1-g(1-|S_{11}|^2)}{1+gD_2}$$

Using the identities (13.5.2) and $1-|S_{11}|^2 = 2K|S_{12}S_{21}| + D_2$, which follows from (13.5.1), the right-hand side of the above circle form can be written as:

$$\frac{g^2|C_2|^2}{(1+gD_2)^2} + \frac{1-g(1-|S_{11}|^2)}{1+gD_2} = \frac{g^2|S_{12}S_{21}|^2 - 2gK|S_{12}S_{21}|+1}{(1+gD_2)^2} \tag{13.11.2}$$

Thus, the *operating power gain circle* will be $|\Gamma_L - c|^2 = r^2$ with center and radius:

$$\boxed{c = \frac{gC_2^*}{1+gD_2}, \quad r = \frac{\sqrt{g^2|S_{12}S_{21}|^2 - 2gK|S_{12}S_{21}|+1}}{|1+gD_2|}} \tag{13.11.3}$$

The points Γ_L on this circle result into the value $G_p = G$ for the operating gain. Such points can be parametrized as $\Gamma_L = c + re^{j\phi}$, where $0 \le \phi \le 2\pi$. As Γ_L traces this circle, the conjugately matched source coefficient $\Gamma_G = \Gamma_{\text{in}}^*$ will also trace a circle because Γ_{in} is related to Γ_L by the bilinear transformation (13.5.8).

In a similar fashion, we find the *available power gain circles* to be $|\Gamma_G - c|^2 = r^2$, where $g = G/|S_{21}|^2$ and:

$$\boxed{c = \frac{gC_1^*}{1+gD_1}, \quad r = \frac{\sqrt{g^2|S_{12}S_{21}|^2 - 2gK|S_{12}S_{21}|+1}}{|1+gD_1|}} \tag{13.11.4}$$

We recall from Sec. 13.5 that the centers of the load and source stability circles were $c_L = C_2^*/D_2$ and $c_G = C_1^*/D_1$. It follows that the centers of the operating power gain circles are along the same ray as c_L, and the centers of the available gain circles are along the same ray as c_G.

For an unconditionally stable two-port, the gain G must be $0 \le G \le G_{\text{MAG}}$, with G_{MAG} given by Eq. (13.6.20). It can be shown easily that the quantities under the square

roots in the definitions of the radii r in Eqs. (13.11.3) and (13.11.4) are non-negative. The gain circles lie inside the unit circle for all such values of G. The radii r vanish when $G = G_{\text{MAG}}$, that is, the circles collapse into single points corresponding to the simultaneous conjugate matched solutions of Eq. (13.8.2).

The MATLAB function `sgcirc` calculates the center and radii c, r of the operating and available power gain circles. It has usage, where G must be entered in dB:

```
[c,r] = sgcirc(S,'p',G);          operating power gain circle
[c,r] = sgcirc(S,'a',G);          available power gain circle
```

Example 13.11.1: A microwave transistor amplifier uses the Hewlett-Packard AT-41410 NPN bipolar transistor with the following S-parameters at 2 GHz [1355]:

$$S_{11} = 0.61\angle 165^\circ, \quad S_{21} = 3.72\angle 59^\circ, \quad S_{12} = 0.05\angle 42^\circ, \quad S_{22} = 0.45\angle -48^\circ$$

Calculate G_{MAG} and plot the operating and available power gain circles for $G = 13, 14, 15$ dB. Then, design source and load matching circuits for the case $G = 15$ dB by choosing the reflection coefficient that has the smallest magnitude.

Solution: The MAG was calculated in Example 13.6.1, $G_{\text{MAG}} = 16.18$ dB. The gain circles and the corresponding load and source stability circles are shown in Fig. 13.11.1. The operating gain and load stability circles were computed and plotted by the MATLAB statements:

```
[c1,r1] = sgcirc(S,'p',13);          % c1 = 0.4443∠52.56°, r1 = 0.5212
[c2,r2] = sgcirc(S,'p',14);          % c2 = 0.5297∠52.56°, r2 = 0.4205
[c3,r3] = sgcirc(S,'p',15);          % c3 = 0.6253∠52.56°, r3 = 0.2968
[cL,rL] = sgcirc(S,'l');             % cL = 2.0600∠52.56°, rL = 0.9753

smith; smithcir(cL,rL,1.7);          % display portion of circle with |ΓL| ≤ 1.7
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);
```

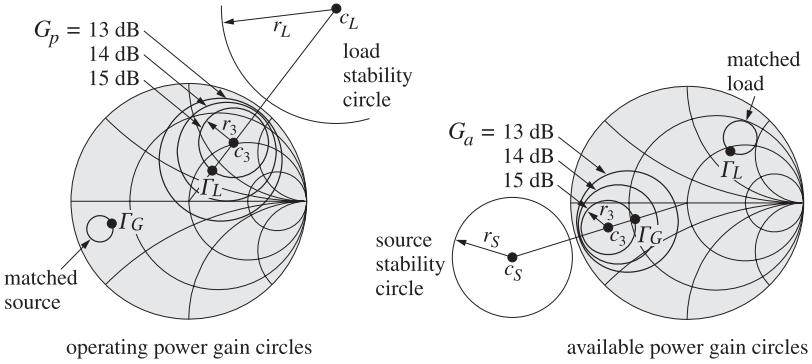


Fig. 13.11.1 Operating and available power gain circles.

The gain circles lie entirely within the unit circle, for example, we have $r_3 + |c_3| = 0.9221 < 1$, and their centers lie along the ray of c_L. As Γ_L traces the 15-dB circle, the corresponding $\Gamma_G = \Gamma_{\text{in}}^*$ traces its own circle, also lying within the unit circle. The following MATLAB code computes and adds that circle to the above Smith chart plots:


```
phi = linspace(0,2*pi,361);          % equally spaced angles at 1° intervals
gammaL = c3 + r3 * exp(j*phi);       % points on 15-dB operating gain circle
gammaG = conj(gin(S,gammaL));        % circle of conjugate matched source points
plot(gammaG);
```

In particular, the point Γ_L on the 15-dB circle that lies closest to the origin is $\Gamma_L = c_3 - r_3 e^{j \arg c_3} = 0.3285\angle 52.56^\circ$. The corresponding matched load will be $\Gamma_G = \Gamma_{\text{in}}^* = 0.6805\angle -163.88^\circ$. These and the corresponding source and load impedances were computed by the MATLAB statements:

```
gL = c3 - r3*exp(j*angle(c3));       zL = g2z(gL);
gG = conj(gin(S,gL));                zG = g2z(gG);
```

The source and load impedances normalized to $Z_0 = 50$ ohm are:

$$z_G = \frac{Z_G}{Z_0} = 0.1938 - 0.1363j, \quad z_L = \frac{Z_L}{Z_0} = 1.2590 + 0.7361j$$

The matching circuits can be designed in a variety of ways as in Example 13.8.1. Using open shunt stubs, we can determine the stub and line segment lengths with the help of the function `stub1`:

$$dl = \text{stub1}(z_G^*, \texttt{'po'}) = \begin{bmatrix} 0.3286 & 0.4122 \\ 0.1714 & 0.0431 \end{bmatrix}$$

$$dl = \text{stub1}(z_L^*, \texttt{'po'}) = \begin{bmatrix} 0.4033 & 0.0786 \\ 0.0967 & 0.2754 \end{bmatrix}$$

In both cases, we may choose the lower solutions as they have shorter total length $d + l$. The available power gain circles can be determined in a similar fashion with the help of the MATLAB statements:

```
[c1,r1] = sgcirc(S,'a',13);          % c1 = 0.5384∠−162.67°, r1 = 0.4373
[c2,r2] = sgcirc(S,'a',14);          % c2 = 0.6227∠−162.67°, r2 = 0.3422
[c3,r3] = sgcirc(S,'a',15);          % c3 = 0.7111∠−162.67°, r3 = 0.2337
[cG,rG] = sgcirc(S,'s');             % cG = 1.5748∠−162.67°, rG = 0.5162

smith; smithcir(cG,rG);              % plot entire source stability circle
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);
```

Again, the circles lie entirely within the unit circle. As Γ_G traces the 15-dB circle, the corresponding matched load $\Gamma_L = \Gamma_{\text{out}}^*$ traces its own circle on the Γ-plane. It can be plotted with:

```
phi = linspace(0,2*pi,361);          % equally spaced angles at 1° intervals
gammaG = c3 + r3 * exp(j*phi);       % points on 15-dB available gain circle
gammaL = conj(gout(S,gammaG));       % circle of conjugate matched loads
plot(gammaL);
```

In particular, the point $\Gamma_G = c_3 - r_3 e^{j\arg c_3} = 0.4774\angle -162.67^\circ$ lies closest to the origin. The corresponding matched load will have $\Gamma_L = \Gamma_{\text{out}}^* = 0.5728\angle 50.76^\circ$. The resulting normalized impedances are:

$$z_G = \frac{Z_G}{Z_0} = 0.3609 - 0.1329j\,, \quad z_L = \frac{Z_L}{Z_0} = 1.1135 + 1.4704j$$

and the corresponding stub matching networks will have lengths:

$$\text{stub1}(z_G^*, \text{'po'}) = \begin{bmatrix} 0.3684 & 0.3905 \\ 0.1316 & 0.0613 \end{bmatrix}, \quad \text{stub1}(z_L^*, \text{'po'}) = \begin{bmatrix} 0.3488 & 0.1030 \\ 0.1512 & 0.2560 \end{bmatrix}$$

The lower solutions have the shortest lengths. For both the operating and available gain cases, the stub matching circuits will be similar to those in Fig. 13.8.4. □

When the two-port is potentially unstable (but with $|S_{11}| < 1$ and $|S_{22}| < 1$,) the stability circles intersect with the unit-circle, as shown in Fig. 13.5.2. In this case, the operating and available power gain circles also intersect the unit-circle and at the *same* points as the stability circles.

We demonstrate this in the specific case of $K < 1$, $|S_{11}| < 1$, $|S_{22}| < 1$, but with $D_2 > 0$, an example of which is shown in Fig. 13.11.2. The intersection of an operating gain circle with the unit-circle is obtained by setting $|\Gamma_L| = 1$ in the circle equation $|\Gamma_L - c| = r$. Writing $\Gamma_L = e^{j\theta_L}$ and $c = |c|e^{j\theta_c}$, we have:

$$r^2 = |\Gamma_L - c|^2 = 1 - 2|c|\cos(\theta_L - \theta_c) + |c|^2 \quad \Rightarrow \quad \cos(\theta_L - \theta_c) = \frac{1 + |c|^2 - r^2}{2|c|}$$

Similarly, the intersection of the load stability circle with the unit-circle leads to the relationship:

$$r_L^2 = |\Gamma_L - c_L|^2 = 1 - 2|c_L|\cos(\theta_L - \theta_{c_L}) + |c_L|^2 \quad \Rightarrow \quad \cos(\theta_L - \theta_{c_L}) = \frac{1 + |c_L|^2 - r_L^2}{2|c_L|}$$

Because $c = gC_2^*/(1 + gD_2)$, $c_L = C_2^*/D_2$, and $D_2 > 0$, it follows that the phase angles of c and c_L will be equal, $\theta_c = \theta_{c_L}$. Therefore, in order for the load stability circle and the gain circle to intersect the unit-circle at the same $\Gamma_L = e^{j\theta_L}$, the following condition must be satisfied:

$$\cos(\theta_L - \theta_c) = \frac{1 + |c|^2 - r^2}{2|c|} = \frac{1 + |c_L|^2 - r_L^2}{2|c_L|} \tag{13.11.5}$$

Using the identities $1 - |S_{11}|^2 = B_2 - D_2$ and $1 - |S_{11}|^2 = (|c_L|^2 - r_L^2)D_2$, which follow from Eqs. (13.5.1) and (13.5.6), we obtain:

$$\frac{1 + |c_L|^2 - r_L^2}{2|c_L|} = \frac{1 + (B_2 - D_2)/D_2}{2|C_2|/|D_2|} = \frac{B_2}{2|C_2|}$$

where we used $D_2 > 0$. Similarly, Eq. (13.11.2) can be written in the form:

$$r^2 = |c|^2 + \frac{1 - g(1 - |S_{11}|^2)}{1 + gD_2} \quad \Rightarrow \quad |c|^2 - r^2 = \frac{g(1 - |S_{11}|^2) - 1}{1 + gD_2} = \frac{g(B_2 - D_2) - 1}{1 + gD_2}$$

Therefore, we have:

$$\frac{1 + |c|^2 - r^2}{2|c|} = \frac{1 + \big(g(B_2 - D_2) - 1\big)/(1 + gD_2)}{2g|C_2|/|1 + gD_2|} = \frac{B_2}{2|C_2|}$$

Thus, Eq. (13.11.5) is satisfied. This condition has two solutions for θ_L that correspond to the two points of intersection with the unit-circle. When $D_2 > 0$, we have $\arg c = \arg C_2^* = -\arg C_2$. Therefore, the two solutions for $\Gamma_L = e^{j\theta_L}$ will be:

$$\Gamma_L = e^{j\theta_L}\,, \quad \theta_L = -\arg(C_2) \pm \operatorname{acos}\left(\frac{B_2}{2|C_2|}\right) \tag{13.11.6}$$

Similarly, the points of intersection of the unit-circle and the available gain circles and source stability circle are:

$$\Gamma_G = e^{j\theta_G}\,, \quad \theta_G = -\arg(C_1) \pm \operatorname{acos}\left(\frac{B_1}{2|C_1|}\right) \tag{13.11.7}$$

Actually, these expressions work also when $D_2 < 0$ or $D_1 < 0$.

Example 13.11.2: The microwave transistor Hewlett-Packard AT-41410 NPN is potentially unstable at 1 GHz with the following S-parameters [1355]:

$$S_{11} = 0.6\angle -163^\circ\,, \quad S_{21} = 7.12\angle 86^\circ\,, \quad S_{12} = 0.039\angle 35^\circ\,, \quad S_{22} = 0.50\angle -38^\circ$$

Calculate G_{MSG} and plot the operating and available power gain circles for $G = 20, 21, 22$ dB. Then, design source and load matching circuits for the 22-dB case by choosing the reflection coefficients that have the smallest magnitudes.

Solution: The MSG computed from Eq. (13.6.21) is $G_{\text{MSG}} = 22.61$ dB. Fig. 13.11.2 depicts the operating and available power gain circles as well as the load and source stability circles. The stability parameters are: $K = 0.7667, \mu_1 = 0.8643, |\Delta| = 0.1893, D_1 = 0.3242, D_2 = 0.2142$. The computations and plots are done with the following MATLAB code:†

```
S = smat([0.60, -163, 7.12, 86, 0.039, 35, 0.50, -38]);      % S-parameters

[K,mu,D,B1,B2,C1,C2,D1,D2] = sparam(S);      % stability parameters

Gmsg = db(sgain(S,'msg'));                   % G_MSG = 22.61 dB
                                             % operating power gain circles:
[c1,r1] = sgcirc(S,'p',20);                  % c1 = 0.6418∠50.80°, r1 = 0.4768
[c2,r2] = sgcirc(S,'p',21);                  % c2 = 0.7502∠50.80°, r2 = 0.4221
[c3,r3] = sgcirc(S,'p',22);                  % c3 = 0.8666∠50.80°, r3 = 0.3893
                                             % load and source stability circles:
[cL,rL] = sgcirc(S,'l');                     % cL = 2.1608∠50.80°, rL = 1.2965
[cG,rG] = sgcirc(S,'s');                     % cG = 1.7456∠171.69°, rG = 0.8566

smith; smithcir(cL,rL,1.5); smithcir(cG,rG,1.5);      % plot Smith charts
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);    % plot gain circles

gL = c3 - r3*exp(j*angle(c3));               % ΓL of smallest magnitude
gG = conj(gin(S,gL));                        % corresponding matched ΓG
plot(gL,'.'); plot(gG,'.');
```

†The function db converts absolute scales to dB. The function ab converts from dB to absolute units.

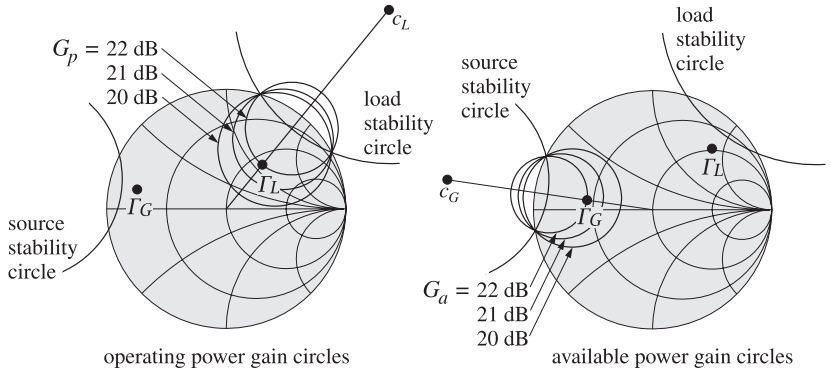


Fig. 13.11.2 Operating and available power gain circles.

```
                                        % available power gain circles:
[c1,r1] = sgcirc(S,'a',20);             % c1 = 0.6809∠171.69°, r1 = 0.4137
[c2,r2] = sgcirc(S,'a',21);             % c2 = 0.7786∠171.69°, r2 = 0.3582
[c3,r3] = sgcirc(S,'a',22);             % c3 = 0.8787∠171.69°, r3 = 0.3228

figure;
smith; smithcir(cL,rL,1.5); smithcir(cG,rG,1.5);
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);

gG = c3 - r3*exp(j*angle(c3));          % ΓG of smallest magnitude
gL = conj(gout(S,gG));                  % corresponding matched ΓL
plot(gL,'.'); plot(gG,'.');
```

Because $D_1 > 0$ and $D_2 > 0$, the stability regions are the portions of the unit-circle that lie outside the source and load stability circles. We note that the operating gain circles intersect the unit-circle at exactly the same points as the load stability circle, and the available gain circles intersect it at the same points as the source stability circle.

The value of Γ_L on the 22-dB operating gain circle that lies closest to the origin is $\Gamma_L = c_3 - r_3 e^{j\arg c_3} = 0.4773\angle 50.80^\circ$ and the corresponding matched source is $\Gamma_G = \Gamma_{\text{in}}^* = 0.7632\angle 167.69^\circ$. We note that both Γ_L and Γ_G lie in their respective stability regions.

For the 22-dB available gain circle (also denoted by c_3, r_3), the closest Γ_G to the origin will be $\Gamma_G = c_3 - r_3 e^{j\arg c_3} = 0.5559\angle 171.69^\circ$ with a corresponding matched load $\Gamma_L = \Gamma_{\text{out}}^* = 0.7147\angle 45.81^\circ$. Again, both Γ_L, Γ_G lie in their stable regions.

Once the Γ_G, Γ_L have been determined, the corresponding matching input and output networks can be designed with the methods of Example 13.8.1. □

13.12 Noise Figure Circles

Every device is a source of internally generated noise. The noise entering the device and the internal noise must be added to obtain the total input system noise. If the device is an amplifier, the total system noise power will be amplified at the output by the gain of the device. If the output load is matched, this gain will be the available gain.

The internally generated noise is quantified in practice either by the *effective noise temperature* T_e, or by the *noise figure* F of the device. The internal noise power is given by $P_n = kT_eB$, where k is the Boltzmann constant and B the bandwidth in Hz. These concepts are discussed further in Sec. 15.8. The relationship between T_e and F is defined in terms of a standard reference temperature $T_0 = 290$ K (degrees Kelvin):

$$F = 1 + \frac{T_e}{T_0} \tag{13.12.1}$$

The noise figure is usually quoted in dB, $F_{\text{dB}} = 10\log_{10} F$. Because the available gain of a two-port depends on the source impedance Z_G, or the source reflection coefficient Γ_G, so will the noise figure.

The optimum source impedance $Z_{G\text{opt}}$ corresponds to the minimum noise figure $F_{\min}$ that can be achieved by the two-port. For other values of Z_G, the noise figure F is greater than $F_{\min}$ and is given by [117–120]:

$$F = F_{\min} + \frac{R_n}{R_G|Z_{G\text{opt}}|^2}\,|Z_G - Z_{G\text{opt}}|^2 \tag{13.12.2}$$

where $R_G = \text{Re}(Z_G)$ and R_n is an equivalent noise resistance. We note that $F = F_{\min}$ when $Z_G = Z_{G\text{opt}}$. Defining the normalized noise resistance $r_n = R_n/Z_0$, where $Z_0 = 50$ ohm, we may write Eq. (13.12.2) in terms of the corresponding source reflection coefficients:

$$\boxed{F = F_{\min} + 4r_n \frac{|\Gamma_G - \Gamma_{G\text{opt}}|^2}{|1 + \Gamma_{G\text{opt}}|^2(1 - |\Gamma_G|^2)}} \tag{13.12.3}$$

The parameters $F_{\min}$, r_n, and $\Gamma_{G\text{opt}}$ characterize the noise properties of the two-port and are usually known.

In designing low-noise microwave amplifiers, one would want to achieve the minimum noise figure and the maximum gain. Unfortunately, the optimum source reflection coefficient $\Gamma_{G\text{opt}}$ does not necessarily correspond to the maximum available gain.

The noise figure circles and the available gain circles are useful tools that allow one to obtain a balance between low noise and high gain designs. The noise figure circles are the locus of points Γ_G that correspond to fixed values of F. They are obtained by rewriting Eq. (13.12.3) as the equation of a circle $|\Gamma_G - c|^2 = r^2$. We write Eq. (13.12.3) in the form:

$$\frac{|\Gamma_G - \Gamma_{G\text{opt}}|^2}{1 - |\Gamma_G|^2} = N, \quad \text{where} \quad N = \frac{(F - F_{\min})|1 + \Gamma_{G\text{opt}}|^2}{4r_n} \tag{13.12.4}$$

which can be rearranged into the circle equation:

$$\left|\Gamma_G - \frac{\Gamma_{G\text{opt}}}{N+1}\right|^2 = \frac{N^2 + N(1 - |\Gamma_{G\text{opt}}|^2)}{(N+1)^2}$$

Thus, the center and radius of the noise figure circle are:

$$\boxed{c = \frac{\Gamma_{G\text{opt}}}{N+1}, \quad r = \frac{\sqrt{N^2 + N(1 - |\Gamma_{G\text{opt}}|^2)}}{N+1}} \tag{13.12.5}$$

The MATLAB function `nfcirc` implements Eq. (13.12.5). Its inputs are the noise parameters $F_{\min}$, r_n, $\Gamma_{G\text{opt}}$, and the desired value of F in dB, and its outputs are c, r:

```
[c,r] = nfcirc(F,Fmin,rn,gGopt);        % noise figure circles
```

The function `nfig` implements Eq. (13.12.3). Its inputs are $F_{\min}$, r_n, $\Gamma_{G\text{opt}}$, and a vector of values of Γ_G, and its output is the corresponding vector of values of F:

```
F = nfig(Fmin, rn, gGopt, gG);          % calculate noise figure F in dB
```

Example 13.12.1: The microwave transistor of Example 13.11.1 has the following noise parameters at 2 GHz [1355]: $F_{\min} = 1.6$ dB, $r_n = 0.16$, and $\Gamma_{G\text{opt}} = 0.26\angle 172^\circ$.

Determine the matched load $\Gamma_{L\text{opt}}$ corresponding to $\Gamma_{G\text{opt}}$ and calculate the available gain. Then, plot the noise figure circles for $F = 1.7, 1.8, 1.9, 2.0$ dB.

For the 1.8-dB noise figure circle, determine Γ_G, Γ_L that correspond to the maximum possible available gain and design appropriate input and output matching networks.

Solution: The conjugate matched load corresponding to $\Gamma_{G\text{opt}}$ is:

$$\Gamma_{L\text{opt}} = \Gamma_{\text{out}}^* = \left[\frac{S_{22} - \Delta\Gamma_{G\text{opt}}}{1 - S_{11}\Gamma_{G\text{opt}}}\right]^* = 0.4927\angle 52.50^\circ$$

The value of the available gain at $\Gamma_{G\text{opt}}$ is $G_{a,\text{opt}} = 13.66$ dB. This is to be compared with the MAG of 16.18 dB determined in Example 13.11.1. To increase the available gain, we must also increase the noise figure. Fig. 13.12.1 shows the locations of the optimum reflection coefficients, as well as several noise figure circles.

The MATLAB code for generating this graph was:[†]

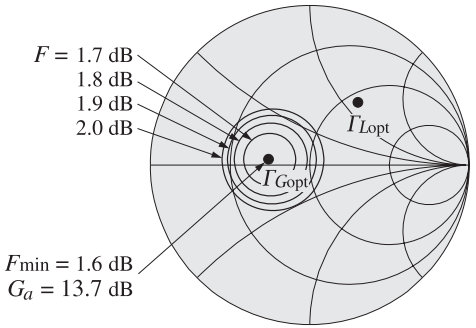


Fig. 13.12.1 Noise figure circles.

```
S = smat([0.61, 165, 3.72, 59, 0.05, 42, 0.45, -48]);

Fmin = 1.6; rn = 0.16; gGopt = p2c(0.26, 172);
```

[†]The function `p2c` converts from phasor form to cartesian complex form, and the function `c2p`, from cartesian to phasor form.


```
Gmag = db(sgain(S,'mag'));                % maximum available gain
Gaopt = db(sgain(S,gGopt,'a'))            % available gain at ΓGopt

gLopt = conj(gout(S,gGopt));              % matched load

[c1,r1] = nfcirc(1.7,Fmin,rn,gGopt);      % noise figure circles
[c2,r2] = nfcirc(1.8,Fmin,rn,gGopt);
[c3,r3] = nfcirc(1.9,Fmin,rn,gGopt);
[c4,r4] = nfcirc(2.0,Fmin,rn,gGopt);

smith; plot([gGopt, gLopt],'.');
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3); smithcir(c4,r4);
```

The larger the noise figure F, the larger the radius of its circle. As F increases, so does the available gain. But as the gain increases, the radius of its circle decreases. Thus, for a fixed value of F, there will be a maximum value of the available gain corresponding to that gain circle that has the smallest radius and is tangent to the noise figure circle.

In the extreme case of the maximum available gain, the available gain circle collapses to a point—the simultaneous conjugate matched point $\Gamma_G = 0.8179\angle -162.67^\circ$— with a corresponding noise figure of $F = 4.28$ dB. These results can be calculated by the MATLAB statements:

```
gG = smatch(S);
F = nfig(Fmin, rn, gopt, gG);
```

Thus, we see that increasing the gain comes at the price of increasing the noise figure. As Γ_G traces the $F = 1.8$ dB circle, the available gain G_a varies as shown in Fig. 13.12.2. Points around this circle can be parametrized as $\Gamma_G = c_2 + r_2 e^{j\phi}$, with $0 \le \phi \le 2\pi$. Fig. 13.12.2 plots G_a versus the angle ϕ. We note that the gain varies between the limits $12.22 \le G_a \le 14.81$ dB.

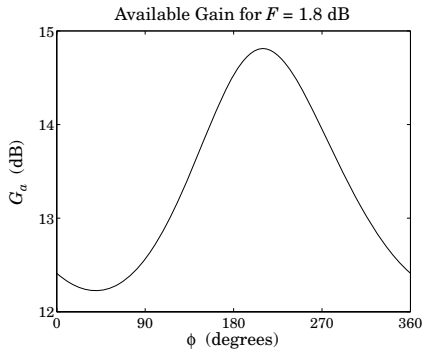


Fig. 13.12.2 Variation of available gain around the noise figure circle $F = 1.8$ dB.

The maximum value, $G_a = 14.81$ dB, is reached when $\Gamma_G = 0.4478\angle -169.73^\circ$, with a resulting matched load $\Gamma_L = \Gamma_{\text{out}}^* = 0.5574\angle 52.50^\circ$. The two points Γ_G, Γ_L, as well as the

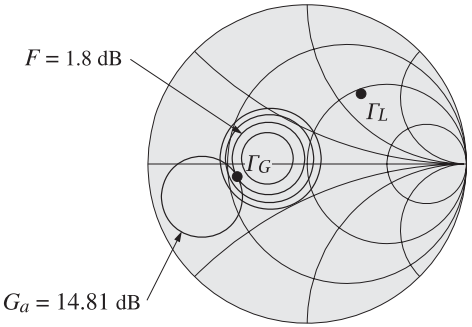


Fig. 13.12.3 Maximum available gain for given noise figure.

$G_a = 14.81$ dB gain circle, which is tangential to the 1.8-dB noise figure circle, are shown in Fig. 13.12.3.

The following MATLAB code performs these calculations and plots:

```
phi = linspace(0,2*pi,721);              % angle in 1/2° increments
gG = c2 + r2*exp(j*phi);                 % Γ_G around the c2,r2 circle
G = db(sgain(S,gG,'a'));                 % available gain in dB
plot(phi*180/pi, G);

[Ga,i] = max(G);                         % maximum available gain

gammaG = gG(i);                          % Γ_G for maximum gain
gammaL = conj(gout(S,gammaG));           % matched load Γ_L

[ca,ra] = sgcirc(S,'a',Ga);              % available gain circle

smith; smithcir(c2,r2); smithcir(ca,ra);
plot([gammaG,gammaL],'.');
```

The maximum gain and the point of tangency with the noise figure circle are determined by direct search, that is, evaluating the gain around the 1.8-dB noise figure circle and finding where it reaches a maximum.

The input and output stub matching networks can be designed with the help of the function `stub1`. The normalized source and load impedances are:

$$z_G = \frac{1+\Gamma_G}{1-\Gamma_G} = 0.3840 - 0.0767j\,, \quad z_L = \frac{1+\Gamma_L}{1-\Gamma_L} = 1.0904 + 1.3993j$$

The stub matching networks have lengths:

$$\text{stub1}(z_G^*, \text{'po'}) = \begin{bmatrix} 0.3749 & 0.3977 \\ 0.1251 & 0.0738 \end{bmatrix}, \quad \text{stub1}(z_L^*, \text{'po'}) = \begin{bmatrix} 0.3519 & 0.0991 \\ 0.1481 & 0.2250 \end{bmatrix}$$

The lower solutions have shorter total lengths $d + l$. The implementation of the matching networks with microstrip lines will be similar to that in Fig. 13.8.4. □

If the two-port is potentially unstable, one must be check that the resulting solutions for Γ_G, Γ_L both lie in their respective stability regions. Problems 13.6 and 13.7 illustrate the design of such potentially unstable low noise microwave amplifiers.

13.13 Problems

13.1 Using the relationships (13.4.3) and (13.4.6), derive the following identities:

$$\begin{aligned}(Z_{11}+Z_G)(Z_{22}+Z_L) - Z_{12}Z_{21} &= \\ (Z_{22}+Z_L)(Z_{\text{in}}+Z_G) &= (Z_{11}+Z_G)(Z_{\text{out}}+Z_L)\end{aligned} \tag{13.13.1}$$

$$\begin{aligned}(1-S_{11}\Gamma_G)(1-S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L &= \\ (1-S_{22}\Gamma_L)(1-\Gamma_{\text{in}}\Gamma_G) &= (1-S_{11}\Gamma_G)(1-\Gamma_{\text{out}}\Gamma_L)\end{aligned} \tag{13.13.2}$$

Using Eqs. (13.4.4) and (13.4.5), show that:

$$\frac{Z_{21}}{Z_{22}+Z_L} = \frac{S_{21}}{1-S_{22}\Gamma_L}\frac{1-\Gamma_L}{1-\Gamma_{\text{in}}}\,, \qquad \frac{Z_{21}}{Z_{11}+Z_G} = \frac{S_{21}}{1-S_{11}\Gamma_G}\frac{1-\Gamma_G}{1-\Gamma_{\text{out}}} \tag{13.13.3}$$

$$\frac{2Z_0}{Z_{\text{in}}+Z_G} = \frac{(1-\Gamma_{\text{in}})(1-\Gamma_G)}{1-\Gamma_{\text{in}}\Gamma_G}\,, \qquad \frac{2Z_0}{Z_{\text{out}}+Z_L} = \frac{(1-\Gamma_{\text{out}})(1-\Gamma_L)}{1-\Gamma_{\text{out}}\Gamma_L} \tag{13.13.4}$$

Finally, for the real part $R_L = \text{Re}(Z_L)$, show that:

$$Z_L = Z_0\frac{1+\Gamma_L}{1-\Gamma_L} \quad \Rightarrow \quad R_L = Z_0\frac{1-|\Gamma_L|^2}{|1-\Gamma_L|^2} \tag{13.13.5}$$

13.2 *Computer Experiment.* The Hewlett-Packard ATF-10136 GaAs FET transistor has the following S-parameters at 4 GHz and 8 GHz [1355]:

$$\begin{aligned} &S_{11} = 0.54\angle{-120^\circ}\,, \quad S_{21} = 3.60\angle{61^\circ}\,, \quad S_{12} = 0.137\angle{31^\circ}\,, \quad S_{22} = 0.22\angle{-49^\circ} \\ &S_{11} = 0.60\angle{87^\circ}\,, \quad S_{21} = 2.09\angle{-32^\circ}\,, \quad S_{12} = 0.21\angle{-36^\circ}\,, \quad S_{22} = 0.32\angle{-48^\circ}\end{aligned}$$

Determine the stability parameters, stability circles, and stability regions at the two frequencies.

13.3 Derive the following relationships, where $R_G = \text{Re}(Z_G)$:

$$\frac{Z_0+Z_G}{2\sqrt{R_G Z_0}} = \frac{1}{\sqrt{1-|\Gamma_G|^2}}\frac{|1-\Gamma_G|}{1-\Gamma_G}\,, \qquad \frac{Z_0-Z_G}{2\sqrt{R_G Z_0}} = -\frac{\Gamma_G}{\sqrt{1-|\Gamma_G|^2}}\frac{|1-\Gamma_G|}{1-\Gamma_G}$$

13.4 Derive Eqs. (13.7.13) relating the generalized S-parameters of power waves to the conventional S-parameters.

13.5 Derive the expression Eq. (13.6.20) for the maximum available gain G_{MAG}, and show that it is the maximum of all three gains, that is, transducer, available, and operating gains.

13.6 *Computer Experiment.* The microwave transistor of Example 13.11.2 has the following noise parameters at a frequency of 1 GHz [1355]: $F_{\min} = 1.3$ dB, $r_n = 0.16$, and $\Gamma_{G\text{opt}} = 0.06\angle{49^\circ}$. Determine the matched load $\Gamma_{L\text{opt}}$ corresponding to $\Gamma_{G\text{opt}}$ and calculate the available gain. Then, plot the noise figure circles for $F = 1.4, 1.5, 1.6$ dB.

For the 1.5-dB noise figure circle, determine the values of Γ_G, Γ_L that correspond to the maximum possible available gain.

Design microstrip stub matching circuits for the computed values of Γ_G, Γ_L.

13.7 *Computer Experiment.* The Hewlett-Packard ATF-36163 pseudomorphic high electron mobility transistor (PHEMT) has the following S- and noise parameters at 6 GHz [1355]:

$$S_{11} = 0.75\angle -131^\circ, \quad S_{21} = 3.95\angle 55^\circ, \quad S_{12} = 0.13\angle -12^\circ, \quad S_{22} = 0.27\angle -116^\circ$$

$$F_{\min} = 0.66 \text{ dB}, \quad r_n = 0.15, \quad \Gamma_{G\text{opt}} = 0.55\angle 88^\circ$$

Plot the $F = 0.7, 0.8, 0.9$ dB noise figure circles. On the 0.7-dB circle, determine the source reflection coefficient Γ_G that corresponds to maximum available gain, and then determine the corresponding matched load coefficient Γ_L.

Design microstrip stub matching circuits for the computed values of Γ_G, Γ_L.

13.8 *Computer Experiment.* In this experiment, you will carry out two low-noise microwave amplifier designs, including the corresponding input and output matching networks. The first design fixes the noise figure and finds the maximum gain that can be used. The second design fixes the desired gain and finds the minimum noise figure that may be achieved.

The Hewlett-Packard Agilent ATF-34143 PHEMT transistor is suitable for low-noise amplifiers in cellular/PCS base stations, low-earth-orbit and multipoint microwave distribution systems, and other low-noise applications.

At 2 GHz, its S-parameters and noise-figure data are as follows, for biasing conditions of $V_{DS} = 4$ V and $I_{DS} = 40$ mA:

$$\begin{aligned} S_{11} &= 0.700\angle -150^\circ, & S_{12} &= 0.081\angle 19^\circ \\ S_{21} &= 6.002\angle 73^\circ, & S_{22} &= 0.210\angle -150^\circ \end{aligned}$$

$$F_{\min} = 0.22 \text{ dB}, \quad r_n = 0.09, \quad \Gamma_{G\text{opt}} = 0.66\angle 67^\circ$$

a. At 2 GHz, the transistor is potentially unstable. Calculate the stability parameters K, μ, Δ, D_1, D_2. Calculate the MSG in dB.

Draw a basic Smith chart and place on it the source and load stability circles (display only a small portion of each circle outside the Smith chart.)

Then, determine the parts of the Smith chart that correspond to the source and load stability regions.

b. For the given optimum reflection coefficient $\Gamma_{G\text{opt}}$, calculate the corresponding load reflection coefficient $\Gamma_{L\text{opt}}$ assuming a matched load.

Place the two points $\Gamma_{G\text{opt}}, \Gamma_{L\text{opt}}$ on the above Smith chart and determine whether they lie in their respective stability regions.

c. Calculate the available gain $G_{a,\text{opt}}$ in dB that corresponds to $\Gamma_{G\text{opt}}$.

Add the corresponding available gain circle to the above Smith chart. (Note that the source stability circle and the available gain circles intersect the Smith chart at the same points.)

d. Add to your Smith chart the noise figure circles corresponding to the noise figure values of $F = 0.25, 0.30, 0.35$ dB.

For the case $F = 0.35$ dB, calculate and plot the available gain G_a in dB as Γ_G traces the noise-figure circle. Determine the maximum value of G_a and the corresponding value of Γ_G.

Place on your Smith chart the available gain circle corresponding to this maximum G_a. Place also the corresponding point Γ_G, which should be the point of tangency between the gain and noise figure circles.

Calculate and place on the Smith chart the corresponding load reflection coefficient $\Gamma_L = \Gamma_{\text{out}}^*$. Verify that the two points Γ_G, Γ_L lie in their respective stability regions.

In addition, for comparison purposes, place on your Smith chart the available gain circles corresponding to the values $G_a = 15$ and 16 dB.

e. The points Γ_G and Γ_L determined in the previous question achieve the maximum gain for the given noise figure of $F = 0.35$ dB.

Design input and output stub matching networks that match the amplifier to a 50-ohm generator and a 50-ohm load. Use "parallel/open" microstrip stubs having 50-ohm characteristic impedance and alumina substrate of relative permittivity of $\epsilon_r = 9.8$.

Determine the stub lengths d, l in units of λ, the wavelength inside the microstrip lines. Choose always the solution with the shortest total length $d + l$.

Determine the effective permittivity ϵ_{eff} of the stubs, the stub wavelength λ in cm, and the width/height ratio, w/h. Then, determine the stub lengths d, l in cm.

Finally, make a schematic of your final design that shows both the input and output matching networks (as in Fig.10.8.3.)

f. The above design sets $F = 0.35$ dB and finds the maximum achievable gain. Carry out an alternative design as follows. Start with a desired available gain of $G_a = 16$ dB and draw the corresponding available gain circle on your Smith chart.

As Γ_G traces the portion of this circle that lies inside the Smith chart, compute the corresponding noise figure F. (Points on the circle can be parametrized by $\Gamma_G = c + re^{j\phi}$, but you must keep only those that have $|\Gamma_G| < 1$.)

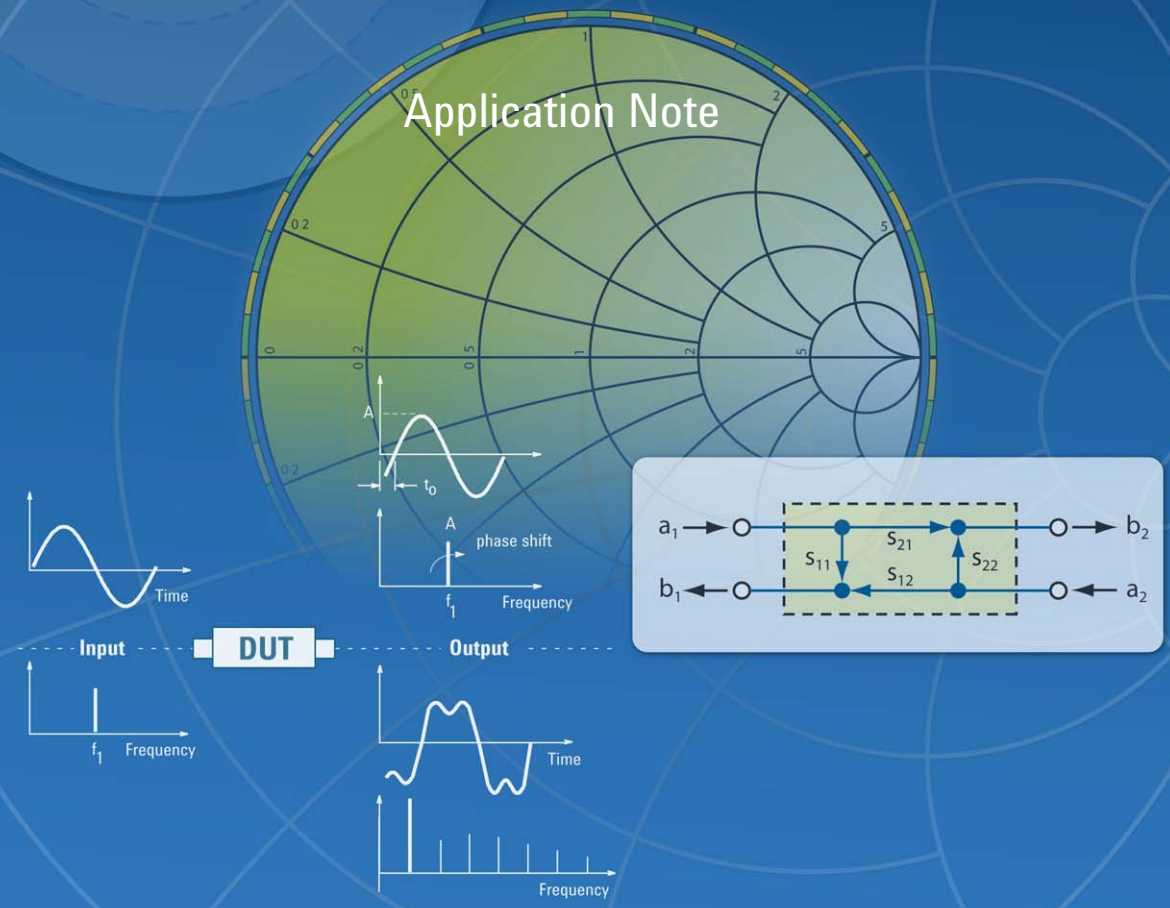
Find the minimum among these values of F in dB and calculate the corresponding value of Γ_G. Calculate the corresponding matched Γ_L.

Add to your Smith chart the corresponding noise figure circle and place on it the points Γ_G and Γ_L.

g. Design the appropriate stub matching networks as in part 13.8.

Understanding the Fundamental Principles of Vector Network Analysis

Application Note



Anticipate ___Accelerate ___Achieve

Agilent Technologies

Table of Contents

Introduction

Network analysis is the process by which designers and manufacturers measure the electrical performance of the components and circuits used in more complex systems. When these systems are conveying signals with information content, we are most concerned with getting the signal from one point to another with maximum efficiency and minimum distortion. Vector network analysis is a method of accurately characterizing such components by measuring their effect on the amplitude and phase of swept-frequency and swept-power test signals.

In this application note, the fundamental principles of vector network analysis will be reviewed. The discussion includes the common parameters that can be measured, including the concept of scattering parameters (S-parameters). RF fundamentals such as transmission lines and the Smith chart will also be reviewed.

Agilent Technologies offers a wide range of portable and benchtop vector network analyzers for characterizing components from DC to 110 GHz. These instruments are available with a wide range of options to simplify testing in the field, laboratory, and production environments.

Measurements in Communications Systems

In any communications system, the effect of signal distortion must be considered. While we generally think of the distortion caused by nonlinear effects (for example, when intermodulation products are produced from desired carrier signals), purely linear systems can also introduce signal distortion. Linear systems can change the time waveform of signals passing through them by altering the amplitude or phase relationships of the spectral components that make up the signal.

Let's examine the difference between linear and nonlinear behavior more closely.

Linear devices impose magnitude and phase changes on input signals (Figure 1). Any sinusoid appearing at the input will also appear at the output, and at the same frequency. No new signals are created. Both active and passive nonlinear devices can shift an input signal in frequency or add other frequency components, such as harmonic and spurious signals. Large input signals can drive normally linear devices into compression or saturation, causing nonlinear operation.

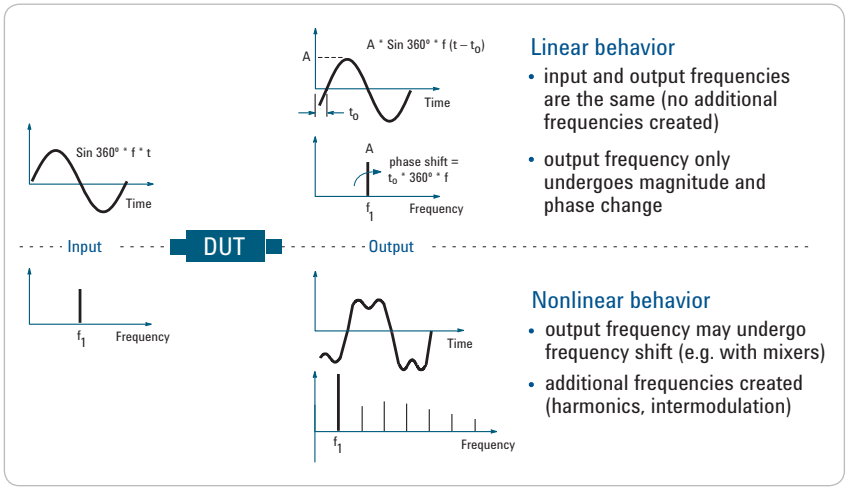


Figure 1. Linear versus nonlinear behavior

For linear distortion-free transmission, the amplitude response of the device under test (DUT) must be flat and the phase response must be linear over the desired bandwidth. As an example, consider a square-wave signal rich in high-frequency components passing through a bandpass filter that passes selected frequencies with little attenuation while attenuating frequencies outside of the passband by varying amounts.

Even if the filter has linear phase performance, the out-of-band components of the square wave will be attenuated, leaving an output signal that, in this example, is more sinusoidal in nature (Figure 2).

If the same square-wave input signal is passed through a filter that only inverts the phase of the third harmonic, but leaves the harmonic amplitudes the same, the output will be more impulse-like in nature (Figure 3). While this is true for the example filter, in general, the output waveform will appear with arbitrary distortion, depending on the amplitude and phase nonlinearities.

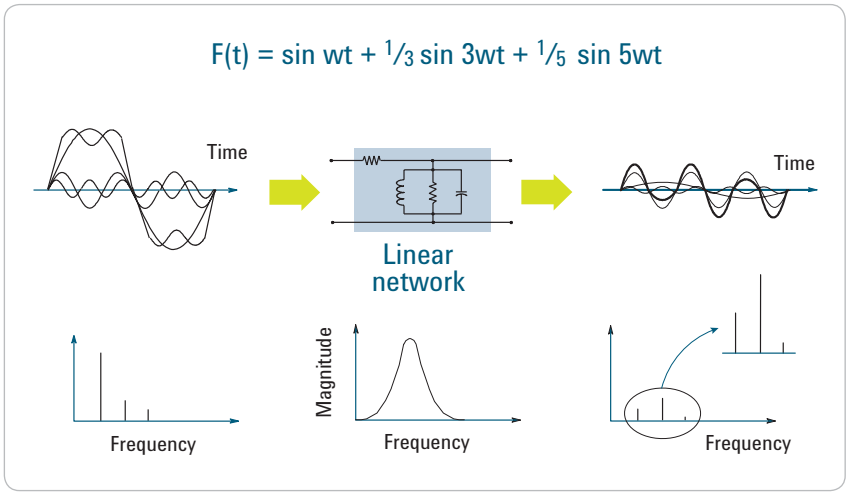


Figure 2. Magnitude variation with frequency

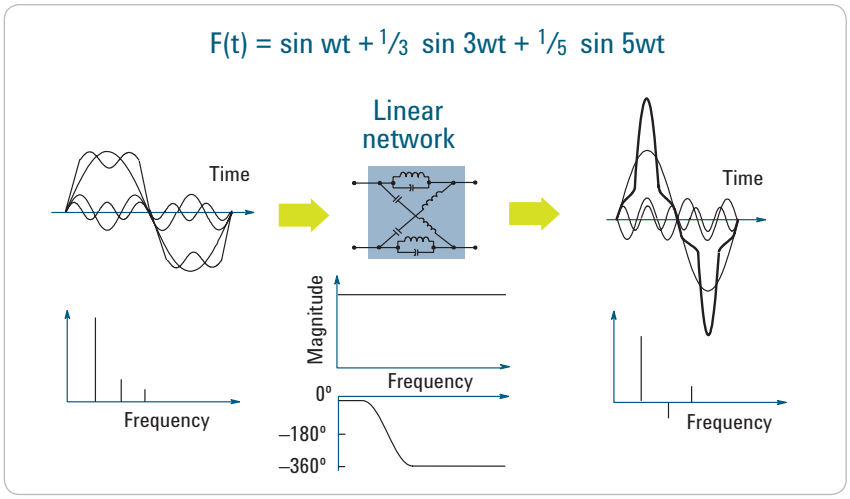


Figure 3. Phase variation with frequency

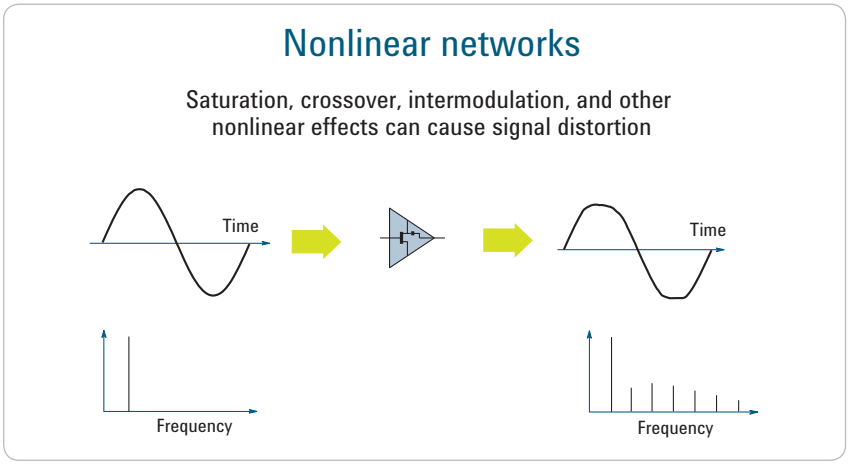


Figure 4. Nonlinear induced distortion

Nonlinear devices also introduce distortion (Figure 4). For example, if an amplifier is overdriven, the output signal clips because the amplifier is saturated. The output signal is no longer a pure sinusoid, and harmonics are present at multiples of the input frequency. Passive devices may also exhibit nonlinear behavior at high power levels, a good example of which is an L-C filter that uses inductors with magnetic cores. Magnetic materials often exhibit hysteresis effects that are highly nonlinear.

Efficient transfer of power is another fundamental concern in communications systems. In order to efficiently convey, transmit or receive RF power, devices such as transmissions lines, antennas and amplifiers must present the proper impedance match to the signal source. Impedance mismatches occur when the real and imaginary parts of input and output impedances are not ideal between two connecting devices.

Importance of Vector Measurements

Measuring both magnitude and phase of components is important for several reasons. First, both measurements are required to fully characterize a linear network and ensure distortion-free transmission. To design efficient matching networks, complex impedance must be measured. Engineers developing models for computer-aided-engineering (CAE) circuit simulation programs require magnitude and phase data for accurate models.

In addition, time-domain characterization requires magnitude and phase information in order to perform an inverse-fourier transform. Vector error correction, which improves measurement accuracy by removing the effects of inherent measurement-system errors, requires both magnitude and phase data to build an effective error model. Phase-measurement capability is very important even for scalar measurements such as return loss, in order to achieve a high level of accuracy (see *Applying Error Correction to Network Analyzer Measurements*, Agilent application note 1287-3).

The Basis of Incident and Reflected Power

In its fundamental form, network analysis involves the measurement of incident, reflected, and transmitted waves that travel along transmission lines. Using optical wavelengths as an analogy, when light strikes a clear lens (the incident energy), some of the light is reflected from the lens surface, but most of it continues through the lens (the transmitted energy) (Figure 5). If the lens has mirrored surfaces, most of the light will be reflected and little or none will pass through it.

While the wavelengths are different for RF and microwave signals, the principle is the same. Network analyzers accurately measure the incident, reflected, and transmitted energy, e.g., the energy that is launched onto a transmission line, reflected back down the transmission line toward the source (due to impedence mismatch), and successfully transmitted to the terminating device (such as an antenna).

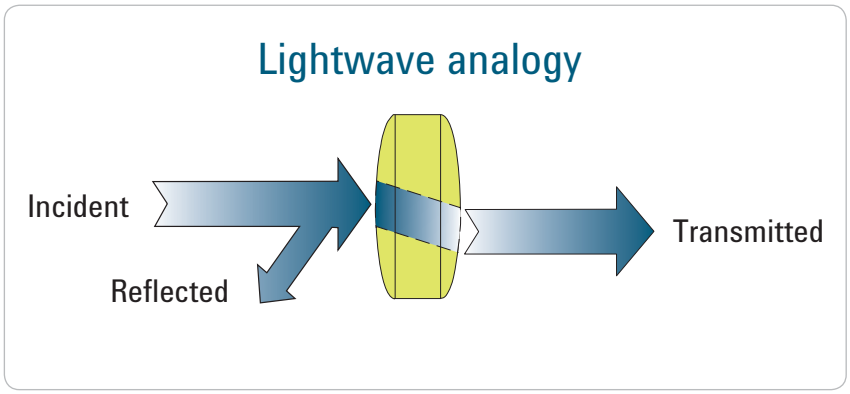


Figure 5. Lightwave analogy to high-frequency device characterization

The Smith Chart

The amount of reflection that occurs when characterizing a device depends on the impedance that the incident signal "sees." Since any impedance can be represented with real and imaginary parts (R + jX or G + jB), they can be plotted on a rectilinear grid known as the complex impedance plane. Unfortunately, an open circuit (a common RF impedence) appears at infinity on the real axis, and therefore cannot be shown.

The polar plot is useful because the entire impedance plane is covered. However, instead of plotting impedance directly, the complex reflection coefficient is displayed in vector form. The magnitude of the vector is the distance from the center of the display, and phase is displayed as the angle of vector referenced to a flat line from the center to the right-most edge. The drawback of polar plots is that impedance values cannot be read directly from the display.

Since there is a one-to-one correspondence between complex impedance and reflection coefficient, the positive real half of the complex impedance plane can be mapped onto the polar display. The result is the Smith chart. All values of reactance and all positive values of resistance from 0 to infinity fall within the outer circle of the Smith chart (Figure 6).

On the Smith chart, loci of constant resistance appear as circles, while loci of constant reactance appear as arcs. Impedances on the Smith chart are always normalized to the characteristic impedance of the component or system of interest, usually 50 ohms for RF and microwave systems and 75 ohms for broadcast and cable-television systems. A perfect termination appears in the center of the Smith chart.

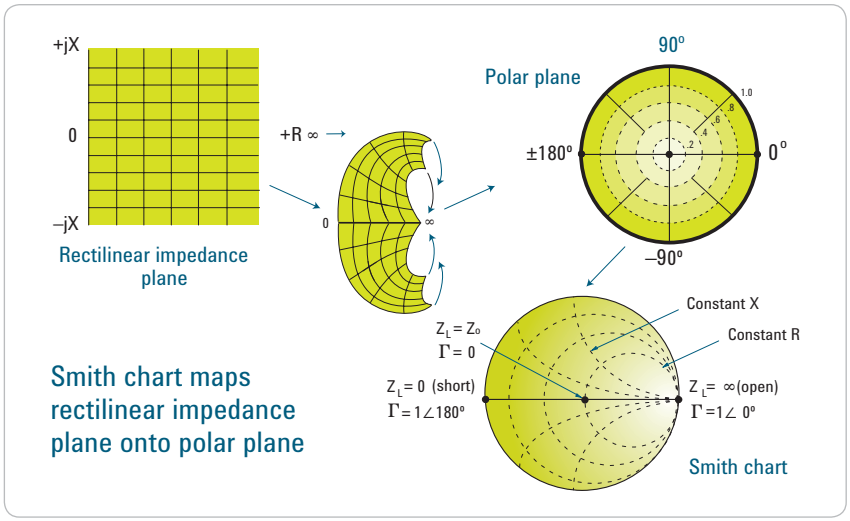


Figure 6. Smith chart review

Power Transfer Conditions

A perfectly matched condition must exist at a connection between two devices for maximum power transfer into a load, given a source resistance of R_S and a load resistance of R_L. This condition occurs when $R_L = R_S$, and is true whether the stimulus is a DC voltage source or a source of RF sine waves (Figure 7).

When the source impedance is not purely resistive, maximum power transfer occurs when the load impedance is equal to the complex conjugate of the source impedance. This condition is met by reversing the sign of the imaginary part of the impedance. For example, if $R_S = 0.6 + j\ 0.3$, then the complex conjugate is $R_S^* = 0.6 - j\ 0.3$.

The need for efficient power transfer is one of the main reasons for the use of transmission lines at higher frequencies. At very low frequencies (with much larger wavelengths), a simple wire is adequate for conducting power. The resistance of the wire is relatively low and has little effect on low-frequency signals. The voltage and current are the same no matter where a measurement is made on the wire.

At higher frequencies, wavelengths are comparable to or smaller than the length of the conductors in a high-frequency circuit, and power transmission can be thought of in terms of traveling waves. When the transmission line is terminated in its characteristic impedance, maximum power is transferred to the load. When the termination is not equal to the characteristic impedance, that part of the signal that is not absorbed by the load is reflected back to the source.

If a transmission line is terminated in its characteristic impedance, no reflected signal occurs since all of the transmitted power is absorbed by the load (Figure 8). Looking at the envelope of the RF signal versus distance along the transmission line shows no standing waves because without reflections, energy flows in only one direction.

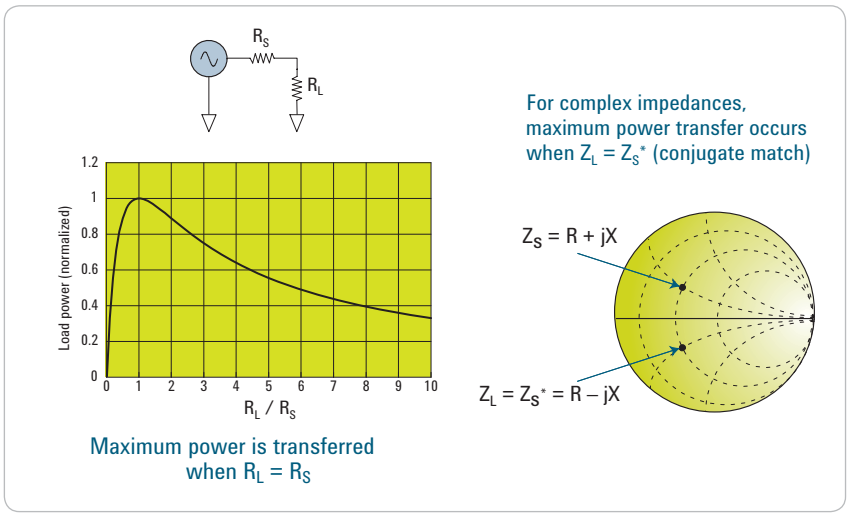


Figure 7. Power transfer

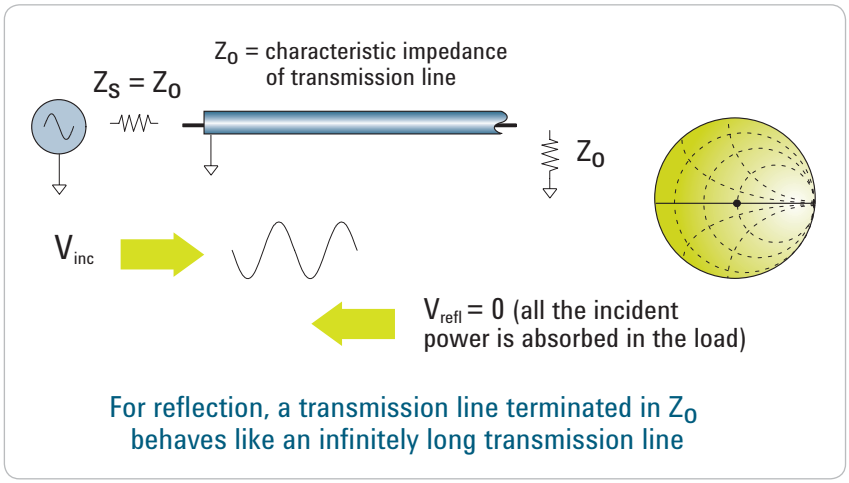


Figure 8. Transmission line terminated with Z_0

When the transmission line is terminated in a short circuit (which can sustain no voltage and therefore dissipates zero power), a reflected wave is launched back along the line toward the source (Figure 9). The reflected voltage wave must be equal in magnitude to the incident voltage wave and be 180 degrees out of phase with it at the plane of the load. The reflected and incident waves are equal in magnitude but traveling in the opposite directions.

If the transmission line is terminated in an open-circuit condition (which can sustain no current), the reflected current wave will be 180 degrees out of phase with the incident current wave, while the reflected voltage wave will be in phase with the incident voltage wave at the plane of the load. This guarantees that the current at the open will be zero. The reflected and incident current waves are equal in magnitude, but traveling in the opposite directions. For both the short and open cases, a standing wave pattern is set up on the transmission line. The voltage valleys will be zero and the voltage peaks will be twice the incident voltage level.

If the transmission line is terminated with say a 25-ohm resistor, resulting in a condition between full absorption and full reflection, part of the incident power is absorbed and part is reflected. The amplitude of the reflected voltage wave will be one-third that of the incident wave, and the two waves will be 180 degrees out of phase at the plane of the load. The valleys of the standing-wave pattern will no longer be zero, and the peaks will be less than those of the short and open cases. The ratio of the peaks to valleys will be 2:1.

The traditional way of determining RF impedance was to measure VSWR using an RF probe/detector, a length of slotted transmission line, and a VSWR meter. As the probe was moved along the transmission line, the relative position and values of the peaks and valleys were noted on the meter. From these measurements, impedance could be derived. The procedure was repeated at different frequencies. Modern network analyzers measure the incident and reflected waves directly during a frequency sweep, and impedance results can be displayed in any number of formats (including VSWR).

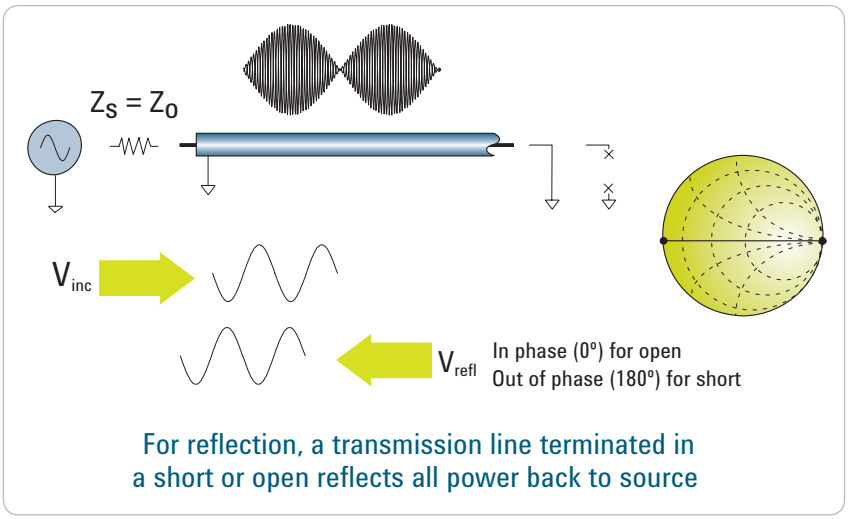


Figure 9. Transmission line terminated with short, open

Network Analysis Terminology

Now that we understand the fundamentals of electromagnetic waves, we must learn the common terms used for measuring them. Network analyzer terminology generally denotes measurements of the incident wave with the R or reference channel. The reflected wave is measured with the A channel, and the transmitted wave is measured with the B channel (Figure 10). With the amplitude and phase information in these waves, it is possible to quantify the reflection and transmission characteristics of a DUT. The reflection and transmission characteristics can be expressed as vector (magnitude and phase), scalar (magnitude only), or phase-only quantities. For example, return loss is a scalar measurement of reflection, while impedance is a vector reflection measurement. Ratioed measurements allow us to make reflection and transmission measurements that are independent of both absolute power and variations in source power versus frequency. Ratioed reflection is often shown as A/R and ratioed transmission as B/R, relating to the measurement channels in the instrument.

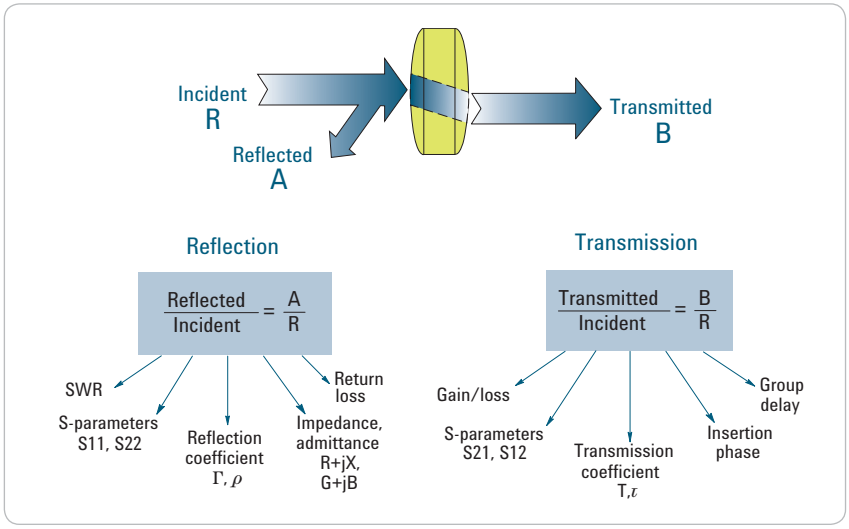


Figure 10. Common terms for high-frequency device characterization

The most general term for ratioed reflection is the complex reflection coefficient, Γ or gamma (Figure 11). The magnitude portion of Γ is called ρ or rho. The reflection coefficient is the ratio of the reflected signal voltage level to the incident signal voltage level. For example, a transmission line terminated in its characteristic impedance Z_o,will have all energy transferred to the load so $V_{refl} = 0$ and $\rho = 0$. When the impedance of the load, Z_L is not equal to the characteristic impedance, energy is reflected and ρ is greater than zero. When the load impedance is equal to a short or open circuit, all energy is reflected and $\rho = 1$. As a result, the range of possible values for ρ is 0 to 1.

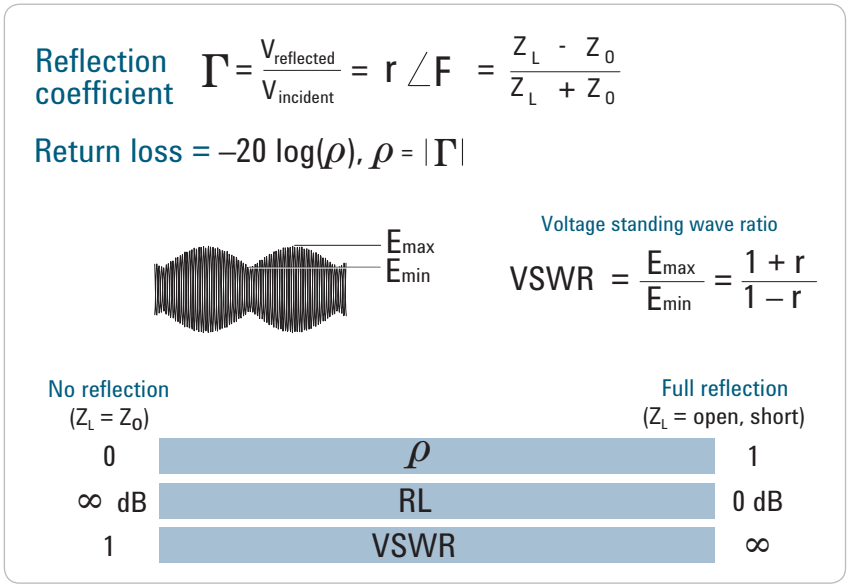


Figure 11. Reflection parameters

Return loss is a way to express the reflection coefficient in logarithmic terms (decibels). Return loss is the number of decibels that the reflected signal is below the incident signal. Return loss is always expressed as a positive number and varies between infinity for a load at the characteristic impedance and 0 dB for an open or short circuit. Another common term used to express reflection is voltage standing wave ratio (VSWR), which is defined as the maximum value of the RF envelope over the minimum value of the RF envelope. It is related to ρ as $(1 + \rho)/(1 - \rho)$. VSWR ranges from 1 (no reflection) to infinity (full reflection).

The transmission coefficient is defined as the transmitted voltage divided by the incident voltage (Figure 12). If the absolute value of the transmitted voltage is greater than the absolute value of the incident voltage, a DUT or system is said to have gain. If the absolute value of the transmitted voltage is less than the absolute value of the incident voltage, the DUT or system is said to have attenuation or insertion loss. The phase portion of the transmission coefficient is called insertion phase.

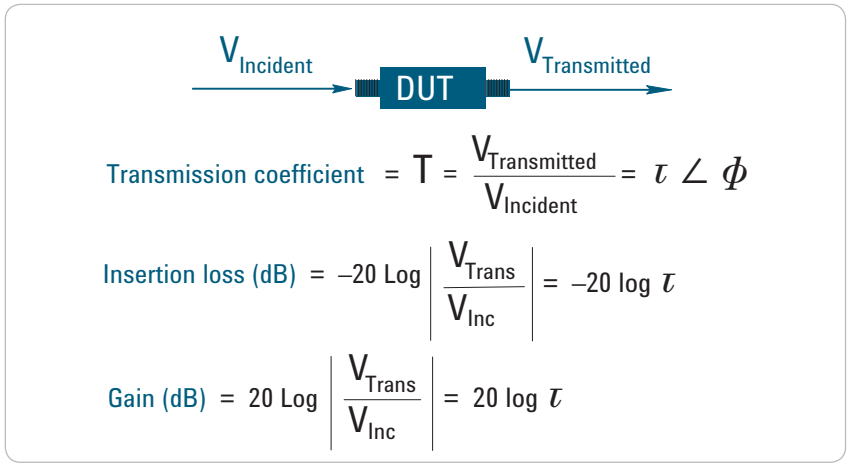


Figure 12. Transmission parameters

Direct examination of insertion phase usually does not provide useful information. This is because the insertion phase has a large (negative) slope with respect to frequency due to the electrical length of the DUT. The slope is proportional to the length of the DUT. Since it is only deviation from linear phase that causes distortion in communications systems, it is desirable to remove the linear portion of the phase response to analyze the remaining nonlinear portion. This can be done by using the electrical delay feature of a network analyzer to mathematically cancel the average electrical length of the DUT. The result is a high-resolution display of phase distortion or deviation from linear phase (Figure 13).

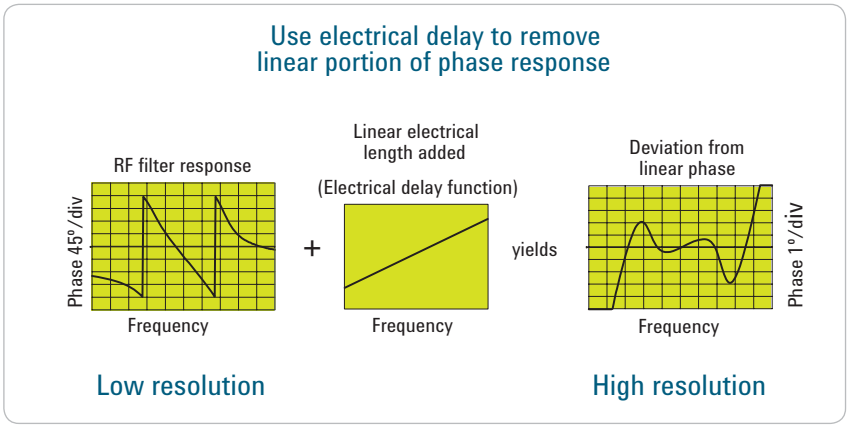


Figure 13. Deviation from linear phase

Measuring Group Delay

Another useful measure of phase distortion is group delay (Figure 14). This parameter is a measure of the transit time of a signal through a DUT versus frequency. Group delay can be calculated by differentiating the DUT's phase response versus frequency. It reduces the linear portion of the phase response to a constant value, and transforms the deviations from linear phase into deviations from constant group delay, (which causes phase distortion in communications systems). The average delay represents the average signal transit time through a DUT.

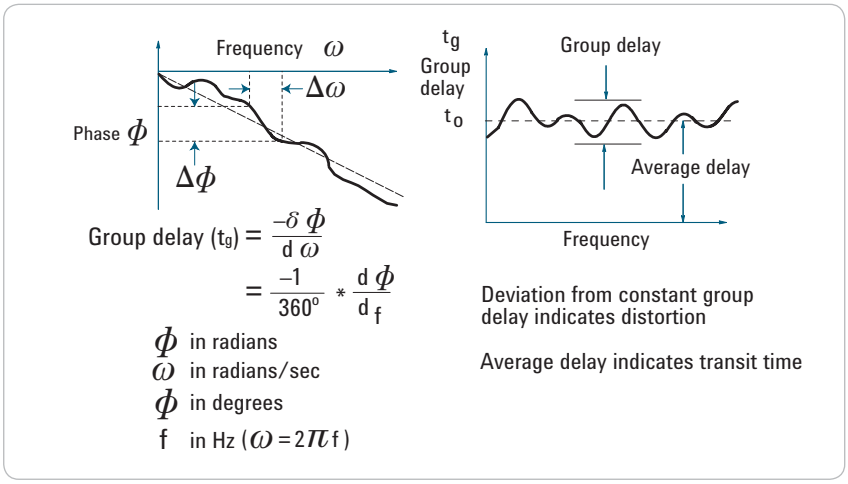


Figure 14. What is group delay?

Depending on the device, both deviation from linear phase and group delay may be measured, since both can be important. Specifying a maximum peak-to-peak phase ripple in a device may not be sufficient to completely characterize it, since the slope of the phase ripple depends on the number of ripples that occur per unit of frequency. Group delay takes this into account because it is the differentiated phase response. Group delay is often a more easily interpreted indication of phase distortion (Figure 15).

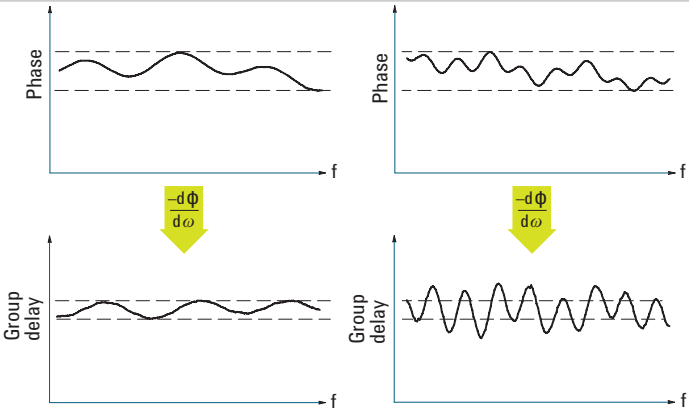


Figure 15. Why measure group delay?

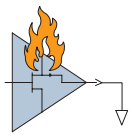
Network Characterization

In order to completely characterize an unknown linear two-port device, we must make measurements under various conditions and compute a set of parameters. These parameters can be used to completely describe the electrical behavior of our device (or network), even under source and load conditions other than when we made our measurements. Low-frequency device or network characterization is usually based on measurement of H, Y, and Z parameters. To do this, the total voltage and current at the input or output ports of a device or nodes of a network must be measured. Furthermore, measurements must be made with open-circuit and short-circuit conditions.

Since it is difficult to measure total current or voltage at higher frequencies, S-parameters are generally measured instead (Figure 16). These parameters relate to familiar measurements such as gain, loss, and reflection coefficient. They are relatively simple to measure, and do not require connection of undesirable loads to the DUT. The measured S-parameters of multiple devices can be cascaded to predict overall system performance. S-parameters are readily used in both linear and nonlinear CAE circuit simulation tools, and H, Y, and Z parameters can be derived from S-parameters when necessary.

The number of S-parameters for a given device is equal to the square of the number of ports. For example, a two-port device has four S-parameters. The numbering convention for S-parameters is that the first number following the S is the port at which energy emerges, and the second number is the port at which energy enters. So S_{21} is a measure of power emerging from Port 2 as a result of applying an RF stimulus to Port 1. When the numbers are the same (e.g. S_{11}), a reflection measurement is indicated.

H,Y, and Z parameters



- Hard to measure total voltage and current at device ports at high frequencies
- Active devices may oscillate or self-destruct with shorts or opens

S-parameters

- Relate to familiar measurements (gain, loss, reflection coefficient, etc.)
- Relatively easy to measure
- Can cascade S-parameters of multiple devices to predict system performance
- Analytically convenient
  - CAD programs
  - Flow-graph analysis
- Can compute H, Y, or Z parameters from S-parameters if desired

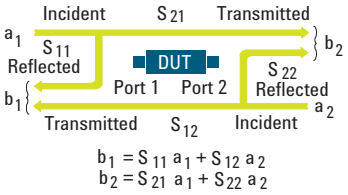


Figure 16. Limitations of H, Y, and Z parameters (Why use S-parameters?)

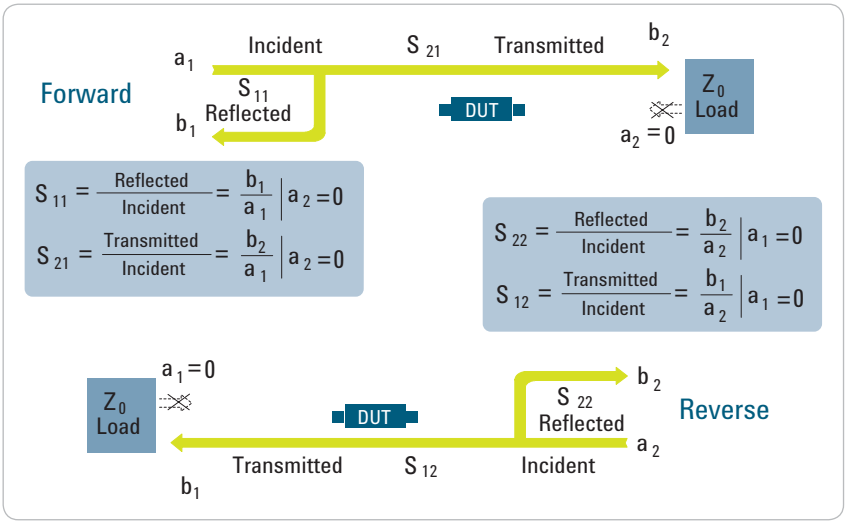


Figure 17. Measuring S-parameters

Forward S-parameters are determined by measuring the magnitude and phase of the incident, reflected, and transmitted signals when the output is terminated in a load that is precisely equal to the characteristic impedance of the test system. In the case of a simple two-port network, S_{11} is equivalent to the input complex reflection coefficient or impedance of the DUT, while S_{21} is the forward complex transmission coefficient. By placing the source at the output port of the DUT and terminating the input port in a perfect load, it is possible to measure the other two (reverse) S-parameters. Parameter S_{22} is equivalent to the output complex reflection coefficient or output impedance of the DUT while S_{12} is the reverse complex transmission coefficient (Figure 17).

www.agilent.com
www.agilent.com/find/NA

Related Literature

Exploring the Architectures of Network Analyzers, Application Note 1287-2, Literature number 5965-7708E

Applying Error Correction to Network Analyzer Measurements, Application Note 1287-3, Literature number 5965-7709E

Network Analyzer Measurements: Filter and Amplifier Examples, Application Note 1287-4, Literature number 5965-7710E

Agilent Advantage Services

Agilent Advantage Services is committed to your success throughout your equipment's lifetime. To keep you competitive, we continually invest in tools and processes that speed up calibration and repair and reduce your cost of ownership. You can also use Infoline Web Services to manage equipment and services more effectively. By sharing our measurement and service expertise, we help you create the products that change our world.

www.agilent.com/find/advantageservices

Agilent Email Updates

www.agilent.com/find/emailupdates
Get the latest information on the products and applications you select.

Agilent Electronic Measurement Group
DEKRA Certified
ISO 9001:2008
Quality Management System

www.agilent.com/quality

For more information on Agilent Technologies' products, applications or services, please contact your local Agilent office. The complete list is available at:

www.agilent.com/find/contactus

Americas

Canada	(877) 894 4414
Brazil	(11) 4197 3600
Mexico	01800 5064 800
United States	(800) 829 4444

Asia Pacific

Australia	1 800 629 485
China	800 810 0189
Hong Kong	800 938 693
India	1 800 112 929
Japan	0120 (421) 345
Korea	080 769 0800
Malaysia	1 800 888 848
Singapore	1 800 375 8100
Taiwan	0800 047 866
Other AP Countries	(65) 375 8100

Europe & Middle East

Belgium	32 (0) 2 404 93 40
Denmark	45 45 80 12 15
Finland	358 (0) 10 855 2100
France	0825 010 700*
	*0.125 €/minute
Germany	49 (0) 7031 464 6333
Ireland	1890 924 204
Israel	972-3-9288-504/544
Italy	39 02 92 60 8484
Netherlands	31 (0) 20 547 2111
Spain	34 (91) 631 3300
Sweden	0200-88 22 55
United Kingdom	44 (0) 118 927 6201

For other unlisted countries:
www.agilent.com/find/contactus

Revised: January 6, 2012

Product specifications and descriptions in this document subject to change without notice.

Published in USA, September 7, 2012
5965-7707E

Agilent Technologies

An Introduction to RF Measurements With a Vector Network Analyzer

Instrument manufacturers' application notes provide essential lessons on measurement procedures

As seen in RF Technology International

This month's review of RF fundamentals is about the vector network analyzer (VNA). Descriptions of VNA operation, calibration, measurement methods and data presentation are adapted from several excellent application notes available from VNA manufacturers. They are the primary source of reliable information on these valuable RF/microwave test instruments.

VNA Operation

The vector network analyzer includes a swept signal source and a tuned receiver that is synchronized to the frequency of the test signal. The receiver also monitors the test signal source, establishing a fixed point of reference for phase and amplitude, the two basic measurements it performs.

The amplitude and phase information received from the device under test (DUT) is compared to the reference to determine the relative difference. Various parameters can be computed from the measurements as illustrated in Figure 1. Some of these measurements are amplitude-only (scalar), some are phase-only (time delay), but most are a combination that take advantage of the VNA's ability to do both.

The signal source is continuous (CW), with the only variations being the choice of amplitude and the frequency sweep range for a particular measurement. In the receiver, the signal from the DUT is converted into the passband of the IF, where it is amplified and detected for measurement.

To measure the phase of this signal after it passes through, or is reflected by the DUT, we must have a point of reference. The phase reference is obtained by splitting off a portion of the source signal before the measurement. The phase of the microwave signal after it has passed through the DUT is then compared with the reference signal. A network analyzer test set automatically samples the reference signal, so no external hardware is needed.

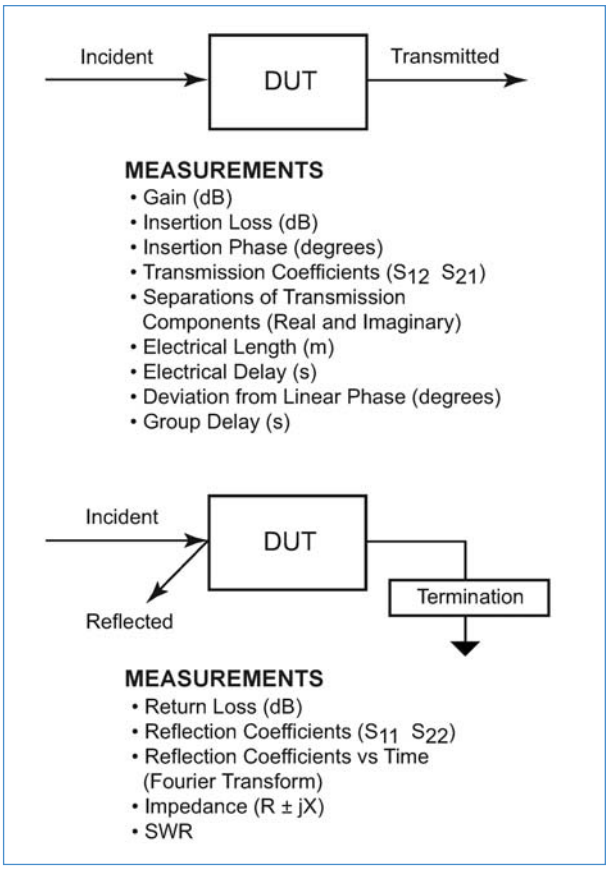


FIGURE 1
Transmission (top) and Reflection (bottom) measurements performed by a VNA. (After [1])

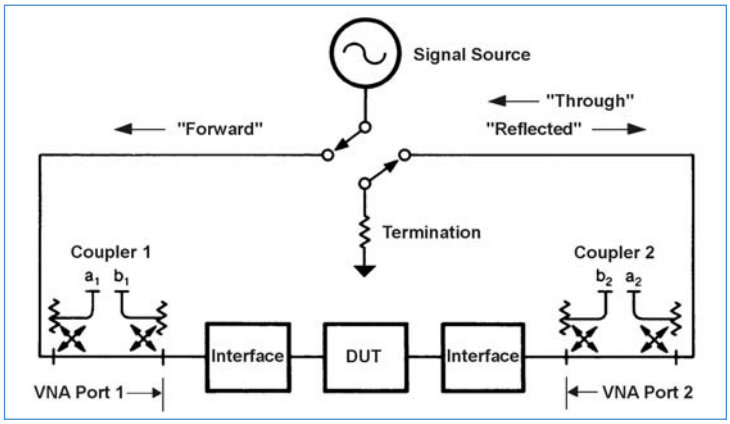


FIGURE 2
Block diagram of basic VNA operation. The sampling ports on the couplers are the inputs to the VNA's receiver. For calibration, the DUT is replaced with precision devices having known characteristics: a transmission line through section, a short, an open, and a termination. (Adapted from [3])

If the phase of a signal is 90 degrees different from the reference signal, the vector network analyzer would read this as –90 degrees, since the test signal is delayed with respect to the reference signal. At a specific frequency, this 90 degree phase shift will correspond to a time delay equal to the propagation of a signal at the speed of light, over a distance of 1/4 wavelength (360 degrees = 1 wavelength). This is a function of frequency, so such time measurements must be calculated after the phase is measured.

Similarly, amplitude is compared to the test signal to determine gain or loss in the DUT.

Figure 2 shows a simplified diagram of the signal paths during VNA measurements. A signal source provides the stimulus to the DUT port under test, with the opposite port terminated in the VNA. Directional couplers in each path sample the forward and reflected signals, which are amplified and detected in the receiver, followed by computation of the desired measurement parameters.

The blocks marked "Interface" include interconnecting cables and connectors external to the VNA unit. To measure an individual component, the "Interface" will also include the additional connectors and circuit board traces up to the terminals of the device itself. For consistent measurement results, test fixtures are available or common components, or may be custom fabricated as needed.

For an amplifier, filter, or other transmission measurement, the excitation signal is applied to the DUT input, with the output terminated. The input and output amplitude and phase are compared in the VNA to obtain results for gain/loss, and

phase shift/delay (S_{12}). The reflected port on the input coupler provides a sample for determing input impedance (S_{11}) or VSWR. When the source and termination are reversed, the VNA can make measurements of output impedance (S_{22}) and reverse path characteristics (S_{21}).

One-port measurements may be needed for devices with other means of termination, or for antenna measurements. In this case, the test signal is applied in the forward direction and the VNA simply does not use the output coupler data.

Multiport measurements may also be desired, for devices like power divider/combiners, couplers, and balanced circuits. Although important, a description of the process is beyond the scope of a "first tutorial" like this one. Interested readers are directed to the many available references available in technical papers and application notes.

The Importance of Calibration

Both phase and amplitude measurements must be calibrated to eliminate the effects of different lengths and variations in gain or loss of the signal paths within the instrument. Futher, the effects of test cables and, as noted above, perhaps the mounting fixture, must also be included in the calibration process to assure that the performance of the DUT is the sole result of the test.

Remember, RF/microwave measurements are made over a range of frequencies. Frequency-dependent effects are always present—phase shift due to changing wavelength, increasing loss with frequency, as well as the frequency response of devices and circuits. These effects must be part of the calibration process, so it is easy to see that the ideal VNA calibration, covering its entire operating frequency range, would require a large amount of data to be measured, stored and applied to the desired measurements.

The most common method of calibration is the through-reflection-line (TRL) technique [3]. The TRL calibration technique relies only on the characteristic impedance of a short transmission line. From two sets of 2-port measurements that differ by this short length of transmission line and two reflection measurements, the full 12-term error model can be determined. Due to the simplicity of the calibration standards, TRL can be applied in dispersive transmission media such as microstrip, stripline and waveguide. Using precision coaxial transmission lines for calibration, TRL currently provides the highest accuracy in coaxial measurements available today.

Many different names have been given to this overall approach: Self Calibration, Thru-Short-Delay, Thru-Reflect-Line, Thru-Reflect-Match, Line-Reflect-

Line, Line-Reflect-Match, and others. These techniques are all variations on the same basic approach.

TRL is an approach to 2-port calibration that relies on transmission lines rather than a set of discrete impedance standards. The technique results in the a 12-term error correction model that assures accurate measurements after calibration.

There are three key advantages gained when using transmission lines as reference standards [3]:

1. Transmission lines are among the simplest elements to realize, especially in non-coaxial media such as microstrip or stripline.
2. The impedance of transmission lines can be accurately determined from physical dimensions and materials characteristics.
3. Transmission lines have traditionally been used as standards and are well understood.

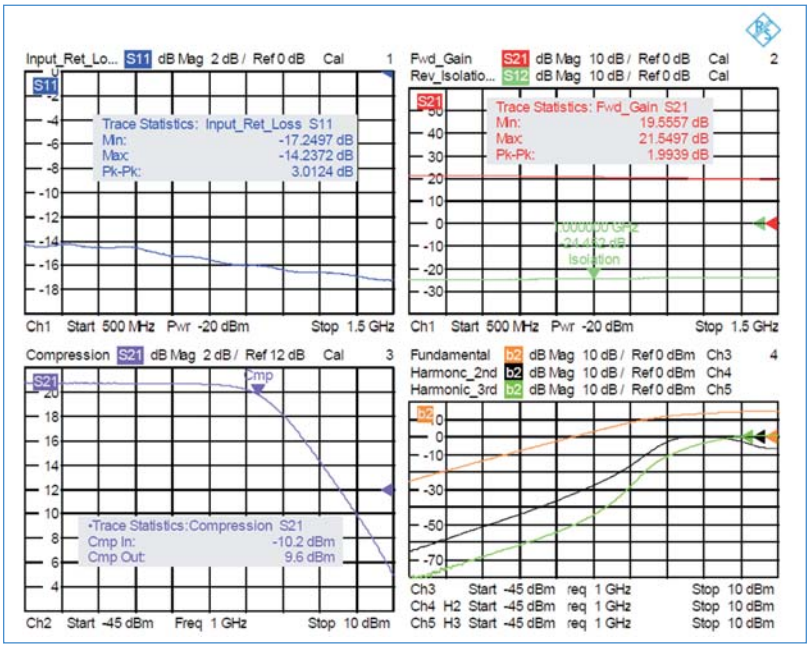


FIGURE 3
A typical set of amplifier results displayed on a Rohde & Schwarz model ZVB VNA [4].

"TRL" refers to the three basic steps in the calibration process:

Through—connection of port 1 and port 2, directly or with a short length of transmission line
Reflection—connect identical one-port high reflection coefficient devices to each port. Typically, these will be short-circuit terminations.
Line—insert a short length of transmission line between ports 1 and 2. *Through* and *Line* must have different line lengths.

VNA manufacturers and other vendors offer precision calibration kits for TRL and other methods of calibration.

Viewing the Results

After obtaining the various measurement results, the VNA user can choose the format for display of the results—parameter, frequency range, and display type. Basic parameters measured include:

S parameters (S_{11}, S_{12}, S_{21}, S_{22})
Voltage Standing Wave Ratio (VSWR)
Reflection Coefficient (Γ, ρ)
Return Loss
Impedance (R ±jX)
Admittance (G ±jB
Gain or Loss (Linear or dB)
Isolation
Insertion Phase
Group Delay

More complex nonlinear measurements are available if the VNA has adjustable power levels, dual receivers, and other features:

Harmonic Levels
Mixer/Multiplier measurements
Compression Point (P_{1dB})
DUT Output Power (dBm)
Power-Added Efficiency (PAE)
Stability Factor (*K*-factor)
AM-PM Conversion
Custom setup with user-defined equations

A set of VNA measurements that includes nonlinear parameters is shown in Figure 3.

The above measurements may be displayed in various formats: linear scale, log scale, vector (Mag/Angle) or Smith chart (Figure 4). The choice of display is selectable by the user, who can choose which parameters to include, along with the appropriate frequency range (horizontal) and scaling (vertical).

All current VNA products allow both internal and

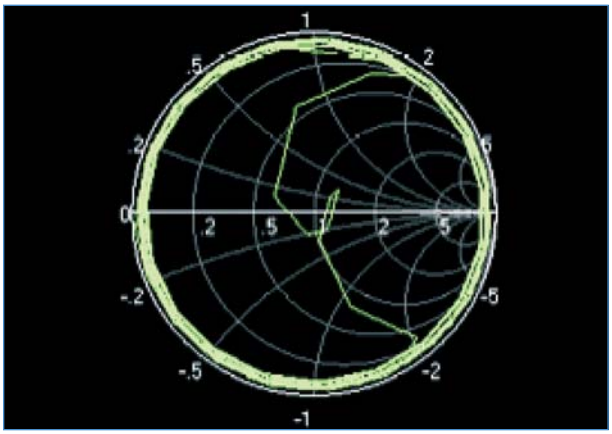


FIGURE 4
VNA measurements may be presented in the form of a Smith chart, like this group delay plot from [1].

external control by a PC, including measurement setup and storage of the results for archiving or post-processing.

A Final Note on New Low-Cost VNAs

Over the past 10 years, a number of analyzers have become available that perform many of the same measurements as the traditional VNAs that are the basis of this tutorial. These units take advantage of recent developments in signal source technology and the power of modern personal computers, enabling vector measurements at very low cost.

These inexpensive units do not have the noise figure, dynamic range, accuracy, stability or features of high-end instruments, and rarely operate above 1 GHz. However, they are proving valuable as alternatives, particularly where their very low cost allows less-critical or preliminary vector measurements to be made at any engineer's (or even a serious hobbyist's) workbench.

References

1. "Vector Network Analyzer Primer," Application Note, Anritsu Company: www.us.anritsu.com

2. "Understanding the Fundamental Principles of Vector Network Analysis," Application Note AN 1287-1, Agilent Technologies: www.agilent. com

3. "Applying the 8510 TRL Calibration for Non-Coaxial Measurements," Application Note 8510-8A, Agilent Technologies: www.agilent.com

4. "Performing Amplifier Measurements with the Vector Network Analyzer ZVB," Application Note 1EZ54, Rohde & Schwarz: www.rohde-schwarz.com