S-Parameters and Related Quantities

Sam Wetterlin

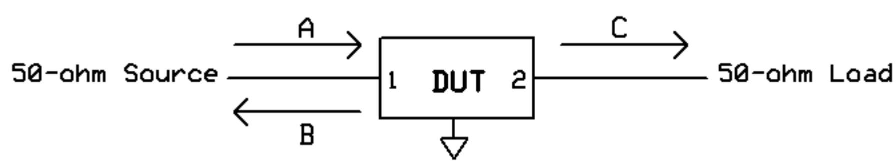
10/20/09

Basic Concept of S-Parameters

S-Parameters are a type of network parameter, based on the concept of "scattering". The more familiar network parameters are based on impedance or admittance. All are ways of measuring the same thing. S-Parameters happen to be the most convenient measurements at RF frequencies. Vector Network Analyzers (VNAs) directly measure S-Parameters. If desired, those S-Parameters can be used to calculate other quantities.

In physics, light interacts with matter by scattering. Even light that is transmitted is considered "scattered"; it just gets scattered in the forward direction. So scattering is just another name for what we would more commonly call reflection and transmission.

Consider the following diagram, in which a device under test ("DUT) is attached to a 50-ohm signal source and a 50-ohm load.

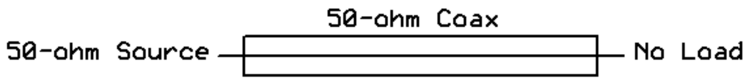


The signal A reaching the DUT at terminal 1 is called the incident signal. Part of that signal is reflected as B, and part of it is transmitted as C out of terminal 2. The ratio of the reflected signal, B, to the incident signal, A, is the reflection coefficient. The ratio of the transmitted signal, C, to the incident signal, A, is the transmission coefficient. Since the reflection is a signal leaving terminal 1 as a result of a signal reaching terminal 1, it is called S_{11}. The transmission is a signal leaving terminal 2 as a result of the signal reaching terminal 1 and is thus called S_{21}. We can also imagine a signal incident at terminal 2 causing a reflection at terminal 2 and a transmission out of terminal 1, which lets us define two more coefficients S_{22} and S_{12}. But we will focus for the moment on the effects of a signal incident on terminal 1.

Technically, the signals A, B and C are calculated in a strange way to derive S-Parameters, so they are not measured directly as voltage or current. They might best be described as the square-root of available power. But as a practical matter we can view them as voltages, so that S-Parameters are the ratio of voltages.

With a source and load of 50 ohms, the reflection coefficient S11 indicates a type of loss, and the transmission coefficient S21 represents voltage gain. With other source and load impedances, we have to consider that the original reflection may reflect off the source, or the transmitted signal may reflect off the load. Therefore, determining how much of the original reflection ends up ultimately being lost, or how much of the original transmission actually represents gain, can become more complicated. Once the S-Parameters are known for the DUT, the source and the load, however, there are straightforward formulas that can be applied. For now we are concerned only with the simple case with the source and load both being 50 ohms.

More on the Reflection Coefficient (S_{11})
The idea of a signal being transmitted is straightforward, but the idea of a signal being reflected is not entirely intuitive. It becomes more intuitive if you imagine the incident signal traveling down a transmission line to reach the DUT. Consider the following diagram, showing such a signal reaching a DUT of infinite impedance—i.e. an open circuit.



When the signal first leaves the source, it sees the coax as the load, and a signal will propagate down the load with a certain voltage V, and a current equal to V/50, because that is what it means for a coax to have an impedance (technically, a "characteristic" impedance) of 50 ohms. When the signal reaches the end of the coax, the current has nowhere to go except to reverse direction. That reverse current (and its associated voltage, based on the 50 ohms of the coax) is the reflection.

In general, a reflection can be looked at as the simple application of Ohms' Law. If the DUT had a 50-ohm impedance, the incident signal would create a certain voltage and current at the DUT, based on Ohm's Law. That voltage/current combination is called the incident signal. If the actual DUT impedance is other than 50 ohms, then some other voltage and current result at the DUT, also based on Ohms' Law. The difference between what the voltage and current would be with 50 ohms, and what it is with the actual DUT impedance, is called the reflection. When the signal arrives at the DUT through a coax cable, the mathematical reflection corresponds to an actual physical reflection. But from the mathematical point of view it doesn't matter whether we can identify a physical reflection. To make the reflection concept more intuitive, it helps to view every DUT as connected to the source through a very short coax cable, just to give a physical reflection a place to occur.

We could write a formula for the voltage and current for those two situations (the source feeding 50 ohms and the source feeding the actual DUT impedance) and subtract to get a formula for the reflection. We could then divide that by the incident signal to get the reflection coefficient. Skipping all the algebra, this is what we would come up with:

Reflection Coefficient = Γ = (Z-50)/(Z+50)
where Z is the DUT input impedance

The symbol "Γ" is the upper case Greek letter gamma, which was probably chosen because it sort of looks like an "r". Gamma is actually the equivalent of a "G", which is sometimes used for the reflection coefficient. The actual Greek equivalent of an "r" is rho (ρ). With a complex impedance Z, the reflection coefficient Γ is also complex and therefore can be represented as a magnitude and angle. The magnitude is commonly labeled ρ and the angle is commonly labeled θ (theta):

$$\Gamma = \rho \angle \theta$$

We have been using 50-ohm sources and loads, and calculated the reflection coefficient by comparing the actual voltage and current at the DUT to that which would exist at a 50-ohm resistor. The 50-ohm resistance is called the Reference Impedance. We can also use other reference impedances. The reference impedance can even include a reactance. However, reference impedances are almost always pure resistances, and generally equal 50 ohms.

The Nature of the Reflected Signal

Once a signal is reflected at the end of the transmission line, the line will have two signals traveling in opposite directions. At any point in the line, the voltage will be the sum of the voltages of the two signals, and the current will the sum of the currents of the two signals (if you treat current in the reverse direction as negative). The voltage of each signal generates the current of the signal, and does so only in the direction of propagation.

However, the situation at the point of reflection is a little different. Supposing a positive reflection from a high impedance termination, the sum of the incident and reflected voltages at that point will be higher than the voltage of the incident signal. That extra voltage effectively generates current in both directions. That is, it generates the current of the reflected wave, and it generates additional current in the load beyond that generated by the incident voltage.

This special situation at the point of reflection is required to conserve energy. For example, if a 150-ohm termination causes a reflection of 50% of the incident voltage and current, the forward current into the load has to be the result of 150% of the incident voltage (that is, the sum of the incident and reflected voltages) acting on the 150-ohm load. Otherwise, if you compute the original power arriving at the termination (voltage squared over 50 ohms), and compute the sum of the reflected power (reflected voltage squared over 50 ohms) and the power into the load, the powers won't match.

There is some temptation to think that if 50% of the voltage is reflected, that leaves 50% to continue in the forward direction. However, it is energy that is conserved at the point of reflection—the amount of energy leaving must equal the amount of energy arriving—not voltage.

S-Parameters in Decibels

It is common in electronics to express ratios in decibels. Typically, decibels are calculated as 10 times the logarithm (base 10) of some base power ratio.

$$\text{Power Ratio in decibels} = 10*\text{Log}(P_1/P_2),$$
$$P_1 \text{ and } P_2 \text{ are power values}$$

We can do the same thing with S-Parameters. But recall that S-Parameters are effectively ratios of voltages. Power is based on the square of the voltage, so power ratios are the square of the corresponding voltage ratio. To convert a voltage ratio to decibels, we could square the ratio and apply the above formula. This has the effect of doubling the dB value (because 10*log(x^2) = 20*log(x)). So, rather than squaring the ratio we just use a factor of 20 rather than 10:

$$\text{Power Ratio in decibels} = 20 \cdot \text{Log}(V_1/V_2),$$
$$V_1 \text{ and } V_2 \text{ are voltage values}$$

So when we calculate voltage gain in decibels, we are actually calculating the decibel value of the power gain. ***Decibels are always based on power ratios, even when we start with a voltage ratio (such as voltage gain).*** (Implicit in this calculation is the assumption that the impedances faced by the two voltages are the same. If they are not, this is a somewhat hypothetical power gain. In the world of reflections, the incident and reflected power are both measured at 50 ohms, so this is not an issue.)

Applying this to the reflection coefficient (S_{11}) we get

$$\text{Magnitude of Reflection Coefficient (in decibels)} = 20 \cdot \text{Log}(|\Gamma|),$$

The absolute value symbols applied to Γ represent the magnitude of Γ. When we express that magnitude in dB, we retain the same phase as the original reflection coefficient, so we still have a complex quantity.

When we use the decibel value of the reflection coefficient, we are looking at the ratio of reflected power to incident power. The dB value of the reflection coefficient will be 0 for a full reflection and negative for anything else. For zero reflection, the reflection coefficient is negative infinity.

The transmission coefficient, S_{21}, can be expressed in dB in the same way as the reflection coefficient.

Return Loss

Closely related to the decibel value of the reflection coefficient is a value known as return loss, which is that same dB value, but with a positive sign. Therefore,

$$\text{Return Loss} = -20 \cdot \text{Log}(|\Gamma|)$$

Technically, return loss is a scalar quantity with no phase, but as a practical matter we often consider it to have the same phase as the reflection coefficient.

The fact that the return loss is a positive number means that the larger its value, the smaller the reflection coefficient, and therefore the smaller the loss due to reflection. This is very counter-intuitive. The only way to make sense of it is to realize that the name “return loss” does not mean loss due to reflection (“return”), but rather means loss of reflection. If we looked at reflection as a desirable thing, then a small reflection indicates a “loss” of some of our potential reflection, and the bigger the positive return loss value, the more we have lost that potential reflection.

However, we are rarely concerned with the loss of potential reflection, but rather we look at the reflection itself as a loss of forward signal. So it is probably more useful to think of return loss as just a way of expressing the reflection coefficient without the inconvenience of the negative sign. To confuse things even more the term “return loss” is frequently used for the magnitude of the reflection coefficient in dB, retaining the negative sign. It is usually best to think of one return loss being “better” or “worse” than another, rather than “bigger” or “smaller” to avoid the confusion that results from retaining or not retaining the negative sign.

SWR

Another quantitative measure of reflections is the voltage standing wave ratio (VSWR), more simply known as the standing wave ratio (SWR). This value is derived from the reflection coefficient as follows:

$$SWR = \frac{1+|\Gamma|}{1-|\Gamma|}$$

The use of the absolute value signs causes SWR to be calculated from the magnitude of the reflection coefficient, so SWR is a scalar quantity without any phase information. Because the magnitude of the reflection coefficient is between 0 and 1, inclusive, SWR ranges from 1 to infinity; the higher it is, the higher the reflection. SWR is often described in the form of a ratio (1:1, 2.3:1, etc.). However, when this is done the second value in the ratio is always 1, so there is no purpose to the ratio format.

SWR does not indicate anything more or less than what return loss indicates, but it is a handy numerical value for two reasons. First, for pure resistances, SWR has the advantage that an SWR of N is produced by a resistance of 50*N or a resistance of 50/N. Thus, for example, if a device has an SWR of 2 and you want an idea of what sort of impedance you are dealing with, you know immediately that resistances of 25 or 100 would create an SWR of 2 (25=50/2; 100=50*2). In addition, there are other impedances with resistance components between 25 and 100 that produce the same SWR.

The second handy use of the numerical value of SWR is for transmission line reflections. If the line is long enough for a standing wave to be generated by a reflection, SWR fairly directly indicates the potential size of the standing wave, which accounts for the name “standing wave ratio”. We won't discuss standing waves here.

As is the case for return loss, the lack of phase information generally makes it pointless to determine SWR with great precision. SWR is used as a general measure of the quality of an impedance match to 50 ohms, and we will generally have some rough threshold value in mind to determine whether an SWR is “good” or “bad”.

Other Quantities
Once we have determined the S-Parameters, we can calculate other useful quantities besides return loss and SWR.

Since S_{11} can be calculated directly from the input impedance of the DUT (if we know that impedance), we can reverse the calculation to determine the input impedance, once we determine S_{11} (Γ):

$$\textbf{Input Impedance} = 50 * (1+\Gamma)/(1-\Gamma)$$

Note that if the reflection coefficient is expressed in dB, we must first convert its magnitude out of dB. Then we must convert the reflection coefficient into rectangular coordinates (a+j*b) to perform the calculation. The resulting impedance will be in rectangular form, consisting of resistance and reactance (R+j*X).

Once we have the input impedance, we can calculate an equivalent RLC circuit that would produce that impedance at a given frequency. For example, a given impedance might be produced at the specified frequency by a 25 ohm resistor in parallel with a 100 pf capacitor. Of course, that does not mean that the DUT is identical to such an RC combination; if we change the frequency it might behave like a 30 ohm resistor in parallel with a 10 pf capacitor. For the equivalent circuit calculation to be useful, the DUT must act consistently like a certain RLC combination over the frequency range of interest.

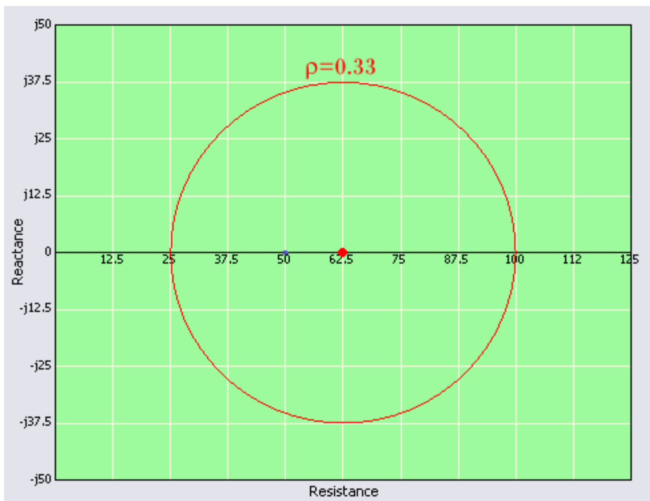
Relating the Quantities to Each Other
Return loss can be translated into Reflection Coefficient, SWR, or into the percent of power reflected by the DUT. Table 1 shows those values for various levels of return loss, as well as two resistance values that would generate that return loss.

Return Loss	Reflect Coef. Mag.	SWR	% Power Reflected	R>50 ohms	R<50 ohms
0	1.000	INF	100.00	INF	0.0
1	0.891	17.4	79.43	869.5	2.9
2	0.794	8.72	63.10	436.2	5.7
3	0.708	5.85	50.12	292.4	8.5
4	0.631	4.42	39.81	221.0	11.3
5	0.562	3.57	31.62	178.5	14.0
6	0.501	3.01	25.12	150.5	16.6
7	0.447	2.61	19.95	130.7	19.1
8	0.398	2.32	15.85	116.1	21.5
9	0.355	2.10	12.59	105.0	23.8
10	0.316	1.92	10.00	96.2	26.0
15	0.178	1.43	3.16	71.6	34.9
20	0.100	1.22	1.00	61.1	40.9
25	0.056	1.12	0.32	56.0	44.7
30	0.032	1.07	0.10	53.3	46.9
35	0.018	1.04	0.03	51.8	48.3
40	0.010	1.02	0.01	51.0	49.0
50	0.003	1.01	0.00	50.3	49.7
60	0.001	1.00	0.00	50.1	49.9

Table 1—Translating return loss to reflection coefficient, SWR,

percentage of power reflected, and equivalent resistance, one greater than and one less than 50 ohms.

The correspondence of return loss and impedance deserves a little more attention. Any given return loss can be generated by a variety of impedances; each such impedance generates the same magnitude of reflection coefficient, but different phases. Since return loss is a scalar quantity, it does not include the phase, so all those impedances generate the same return loss. The following diagram shows the circle of impedances that generate a reflection coefficient whose magnitude (denoted by ρ [rho]) is 0.33, corresponding to a return loss of 9.6 dB.



Impedances that generate a reflection coefficient with magnitude of 0.33 (return loss of 9.6 dB).

As the reflection coefficient decreases (return loss increases), the circle gets smaller and smaller and its center moves closer and closer to 50 ohms. Therefore, return loss can be thought of as a measure of how close the impedance is to 50 ohms. There are two pure resistance values which create any finite return loss. If the lower one is 50/N, then the higher one is 50*N. Thus, in the diagram, 25 ohms is 50/2 and so generates the same return loss as 50*2=100 ohms. Note that the circle of impedances is not centered at 50 ohms, but rather at the average of the two pure resistances that generate it.