For EECS142, Lecture presented by Dr. Joel Dunsmore

Slide 1
Welcome to Network Analyzer Basics.

Slide 2
One of the most fundamental concepts of high-frequency network analysis involves incident, reflected and transmitted waves traveling along transmission lines. It is helpful to think of traveling waves along a transmission line in terms of a lightwave analogy. We can imagine incident light striking some optical component like a clear lens. Some of the light is reflected off the surface of the lens, but most of the light continues on through the lens. If the lens were made of some lossy material, then a portion of the light could be absorbed within the lens. If the lens had mirrored surfaces, then most of the light would be reflected and little or none would be transmitted through the lens. This concept is valid for RF signals as well, except the electromagnetic energy is in the RF range instead of the optical range, and our components and circuits are electrical devices and networks instead of lenses and mirrors.
Network analysis is concerned with the accurate measurement of the *ratios* of the reflected signal to the incident signal, and the transmitted signal to the incident signal.

Slide 3
Components are tested for a variety of reasons. Many components are used as "building blocks" in more complicated RF systems. For example, in most transceivers there are amplifiers to boost LO power to mixers, and filters to remove signal harmonics. Often, R&D engineers need to measure these components to verify their simulation models and their actual hardware prototypes. For component production, a manufacturer must measure the performance of their products so they can provide accurate specifications. This is essential so prospective customers will know how a particular component will behave in their application.
When used in communications systems to pass signals, designers want to ensure the component or circuit is not causing excessive signal distortion. This can be in the form of linear distortion where flat magnitude and linear phase shift versus frequency is not maintained over the bandwidth of interest, or in the form of nonlinear effects like intermodulation distortion.
Often it is most important to measure how reflective a component is, to ensure that it absorbs energy efficiently. Measuring antenna match is a good example.

Slide 4
In many situations, magnitude-only data is sufficient for out needs. For example, we may only care about the gain of an amplifier or the stop-band rejection of a filter. However, as we will explore throughout this paper, measuring phase is a critical element of network analysis.
Complete characterization of devices and networks involves measurement of phase as well as magnitude. This is necessary for developing circuit models for simulation and to design matching circuits based on conjugate-matching techniques. Time-domain characterization requires magnitude and phase information to perform the inverse-Fourier transform. Finally, for best measurement accuracy, phase data is required to perform vector error correction.

Slide 5
The need for efficient transfer of RF power is one of the main reasons behind the use of transmission lines. At low frequencies where the wavelength of the signals are much larger than the length of the circuit conductors, a simple wire is very useful for carrying power. Current travels down the wire easily, and voltage and current are the same no matter where we measure along the wire.
At high frequencies however, the wavelength of signals of interest are comparable to or much smaller than the length of conductors. In this case, power transmission can best be thought of in terms of traveling waves.
Of critical importance is that a lossless transmission line takes on a characteristic impedance (Zo). In fact, an infinitely long transmission line appears to be a resistive load! When the transmission line is terminated in its characteristic impedance, maximum power is transferred to the load. When the termination is not Zo, the portion of the signal which is not absorbed by the load is reflected back toward the source. This creates a condition where the envelope voltage along the transmission line varies with position. We will examine the incident and reflected waves on transmission lines with different load conditions in following slides

Slide 6

RF transmission lines can be made in a variety of transmission media. Common examples are coaxial, waveguide, twisted pair, coplanar, stripline and microstrip. RF circuit design on printed-circuit boards (PCB) often use coplanar or microstrip transmission lines. The fundamental parameter of a transmission line is its characteristic impedance Zo. Zo describes the relationship between the voltage and current traveling waves, and is a function of the various dimensions of the transmission line and the dielectric constant (ε_r) of the non-conducting material in the transmission line. For most RF systems, Zo is either 50 or 75 ohms.
For low-power situations (cable TV, for example) coaxial transmission lines are optimized for low loss, which works out to about 75 ohms (for coaxial transmission lines with air dielectric). For RF and microwave communication and radar applications, where high power is often encountered, coaxial transmission lines are designed to have a characteristic impedance of 50 ohms, a compromise between maximum power handling (occurring at 30 ohms) and minimum loss.

Slide 7
Before we begin our discussion about transmission lines, let us look at the condition for maximum power transfer into a load, given a source impedance of Rs. The graph above shows that the matched condition (RL = RS) results in the maximum power dissipated in the load resistor. This condition is true whether the stimulus is a DC voltage source or an RF sinusoid.
For maximum transfer of energy into a transmission line from a source or from a transmission line to a load (the next stage of an amplifier, an antenna, etc.), the impedance of the source and load should match the characteristic impedance of the transmission line. In general, then, Zo is the target for input and output impedances of devices and networks.
When the source impedance is not purely resistive, the 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 RS = 0.6 + j0.3, then the complex conjugate RS* = 0.6 - j0.3.
Sometimes the source impedance is adjusted to be the complex conjugate of the load impedance. For example, when matching to an antenna, the load impedance is determined by the characteristics of the antenna. A designer has to optimize the output match of the RF amplifier over the frequency range of the antenna so that maximum RF power is transmitted through the antenna

Slide8
Let's review what happens when transmission lines are terminated in various impedances, starting with a Zo load. Since a transmission line terminated in its characteristic impedance results in maximum transfer of power to the load, there is no reflected signal. This result is the same as if the transmission line was infinitely long. If we were to look at the envelope of the RF signal versus distance along the transmission line, it would be constant (no standing-wave pattern). This is because there is energy flowing in one direction only.

Slide 9
Next, let's terminate our line in a short circuit. Since purely reactive elements cannot dissipate any power, and there is nowhere else for the energy to go, a reflected wave is launched back down the line toward the source. For Ohm's law to be satisfied (no voltage across the short), this reflected wave must be equal in voltage magnitude to the incident wave, and be 180° out of phase with it. This satisfies the condition that the total voltage must equal zero at the plane of the short circuit. Our reflected and incident voltage (and current) waves will be identical in magnitude but traveling in the opposite direction.
Now let us leave our line open. This time, Ohm's law tells us that the open can support no current. Therefore, our reflected current wave must be 180° out of phase with respect to the incident wave (the voltage wave will be in phase with the incident wave). This guarantees that current at the open will be zero. Again, our reflected and incident current (and voltage) waves will be identical in magnitude, but traveling in the opposite direction. For both the short and open cases, a standing-wave pattern will be set up on the transmission line. The valleys will be at zero and the peaks at twice the incident voltage level. The peaks and valleys of the short and open will be shifted in position along the line with respect to each other, in order to satisfy Ohm's law as described above.

Slide 10
Finally, let's terminate our line with a 25 Ω resistor (an impedance between the full reflection of an open or short circuit and the perfect termination of a 50 Ω load). Some (but not all) of our incident energy will be absorbed in the load, and some will be reflected back towards the source. We will find that our reflected voltage wave will have an amplitude 1/3 that of the incident wave, and that the two waves will be 180° out of phase at the load. The phase relationship between

the incident and reflected waves will change as a function of distance along the transmission line from the load. The valleys of the standing-wave pattern will no longer be zero, and the peak will be less than that of the short/open case. The significance of standing waves should not go unnoticed. Ohm's law tells us the complex relationship between the incident and reflected signals at the load. Assuming a 50-ohm source, the voltage across a 25-ohm load resistor will be two thirds of the voltage across a 50-ohm load. Hence, the voltage of the reflected signal is one third the voltage of the incident signal and is 180^{o} out of phase with it. However, as we move away from the load toward the source, we find that the phase between the incident and reflected signals changes! The vector sum of the two signals therefore also changes along the line, producing the standing wave pattern. The apparent impedance also changes along the line because the relative amplitude and phase of the incident and reflected waves *at any given point* uniquely determine the measured impedance. For example, if we made a measurement one quarter wavelength away from the 25-ohm load, the results would indicate a 100-ohm load. The standing wave pattern repeats every half wavelength, as does the apparent impedance.

Slide 11

Now that we fully understand the relationship of electromagnetic waves, we must also recognize the terms used to describe them. Common network analyzer terminology has the incident wave measured with the R (for reference) receiver. The reflected wave is measured with the A receiver and the transmitted wave is measured with the B receiver. With amplitude and phase information of these three waves, we can quantify the reflection and transmission characteristics of our device under test (DUT). Some of the common measured terms are scalar in nature (the phase part is ignored or not measured), while others are vector (both magnitude and phase are measured). For example, return loss is a scalar measurement of reflection, while impedance results from a vector reflection measurement. Some, like group delay, are purely phase-related measurements.
Ratioed reflection is often shown as A/R and ratioed transmission is often shown as B/R, relating to the measurement receivers used in the network analyzer

Slide 13

Let's now examine reflection measurements. The first term for reflected waves is reflection coefficient gamma (Γ). Reflection coefficient is the ratio of the reflected signal voltage to the incident signal voltage. It can be calculated as shown above by knowing the impedances of the transmission line and the load. The magnitude portion of gamma is called rho (ρ). A transmission line terminated in Zo will have all energy transferred to the load; hence $V_{refl} = 0$ and $\rho = 0$. When Z_L is not equal to Zo , some energy is reflected and ρ is greater than zero. When Z_L is a short or open circuit, all energy is reflected and $\rho = 1$. The range of possible values for ρ is therefore zero to one.
Since it is often very convenient to show reflection on a logarithmic display, the second way to convey reflection is return loss. Return loss is expressed in terms of dB, and is a scalar quantity. The definition for return loss includes a negative sign so that the return loss value is always a positive number (when measuring reflection on a network analyzer with a log magnitude format, ignoring the minus sign gives the results in terms of return loss). Return loss can be thought of as the number of dB that the reflected signal is below the incident signal. Return loss varies between infinity for a Zo impedance and 0 dB for an open or short circuit.
As we have already seen, two waves traveling in opposite directions on the same transmission line cause a "standing wave". This condition can be measured in terms of the voltage-standing-wave ratio (VSWR or SWR for short). VSWR is defined as the maximum value of the RF envelope over the minimum value of the envelope. This value can be computed as $(1+\rho)/(1-\rho)$. VSWR can take

Slide 13

Transmission coefficient T is defined as the transmitted voltage divided by the incident voltage. If $|V_{trans}| > |V_{inc}|$, the DUT has gain, and if $|V_{trans}| < |V_{inc}|$, the DUT exhibits attenuation or insertion loss. When insertion loss is expressed in dB, a negative sign is added in the definition so that the loss value is expressed as a positive number. The phase portion of the transmission coefficient is called insertion phase.
There is more to transmission than simple gain or loss. In communications systems, signals are time varying -- they occupy a given bandwidth and are made up of multiple frequency components. It is important then to know to what extent the DUT alters the makeup of the signal, thereby causing signal distortion. While we often think of distortion as only the result of nonlinear networks, we will see shortly that linear networks can also cause signal distortion.

Slide 14

Before we explore linear signal distortion, lets review the differences between linear and nonlinear behavior. Devices that behave linearly only impose magnitude and phase changes on input signals. Any sinusoid appearing at the input will also appear at the output at the same frequency. No new signals are created. When a single sinusoid is passed through a linear network, we don't consider amplitude and phase changes as distortion. However, when a complex, time-varying signal is passed through a linear network, the amplitude and phase shifts can dramatically distort the time-domain waveform.
Non-linear devices can shift input signals in frequency (a mixer for example) and/or create new signals in the form of harmonics or intermodulation products. Many components that behave linearly under most signal conditions can exhibit nonlinear behavior if driven with a large enough input signal. This is true for both passive devices like filters and even connectors, and active devices like amplifiers

Slide 15
Now lets examine how linear networks can cause signal distortion. There are three criteria that must be satisfied for linear *distortionless* transmission. First, the amplitude (magnitude) response of the device or system must be flat over the bandwidth of interest. This means all frequencies within the bandwidth will be attenuated identically. Second, the phase response must be linear over the bandwidth of interest. And last, the device must exhibit a "minimum-phase response", which means that at 0 Hz (DC), there is 0° phase shift ($0^\circ \pm n*180^\circ$ is okay if we don't mind an inverted signal).
How can magnitude and phase distortion occur? The following two examples will illustrate how both magnitude and phase responses can introduce linear signal distortion.

Slide 16
Here is an example of a square wave (consisting of three sinusoids) applied to a bandpass filter. The filter imposes a non-uniform amplitude change to each frequency component. Even though no phase changes are introduced, the frequency components no longer sum to a square wave at the output. The square wave is now severely distorted, having become more sinusoidal in nature.

Slide 17
Let's apply the same square wave to another filter. Here, the third harmonic undergoes a 180° phase shift, but the other components are not phase shifted. All the amplitudes of the three spectral components remain the same (filters which only affect the phase of signals are called allpass filters). The output is again distorted, appearing very impulsive this time

Slide 18
Now that we know insertion phase versus frequency is a very important characteristic of a component, let's see how we would measure it. Looking at insertion phase directly is usually not very useful. This is because the phase has a negative slope with respect to frequency due to the electrical length of the device (the longer the device, the greater the slope). Since it is only the deviation from linear phase which causes distortion, it is desirable to remove the linear portion of the phase response. This can be accomplished by using the electrical delay feature of the network analyzer to cancel the electrical length of the DUT. This results in a high-resolution display of phase distortion (deviation from linear phase).

Slide 19
Another useful measure of phase distortion is group delay. Group delay is a measure of the transit time of a signal through the device under test, versus frequency. Group delay is calculated by differentiating the insertion-phase response of the DUT versus frequency. Another way to say this is that group delay is a measure of the slope of the transmission phase response. The linear portion of the phase response is converted to a constant value (representing the average signal-transit time) and deviations from linear phase are transformed into deviations from constant group delay. The variations in group delay cause signal distortion, just as deviations from linear phase cause distortion. Group delay is just another way to look at linear phase distortion.
When specifying or measuring group delay, it is important to quantify the aperture in which the measurement is made. The aperture is defined as the frequency delta used in the differentiation process (the denominator in the group-delay formula). As we widen the aperture, trace noise is reduced but less group-delay resolution is available (we are essentially averaging the phase response over a wider window). As we make the aperture more narrow, trace noise increases but we have more measurement resolution.

Slide 20
Why are both deviation from linear phase and group delay commonly measured? Depending on the device, both may be important. Specifying a maximum peak-to-peak value of phase ripple is not sufficient to completely characterize a device since the slope of the phase ripple is dependent on the number of ripples which occur over a frequency range of interest. Group delay takes this into account since it is the differentiated phase response. Group delay is often a more easily interpreted indication of phase distortion.
The plot above shows that the same value of peak-to-peak phase ripple can result in substantially different group delay responses. The response on the right with the larger group-delay variation would cause more signal distortion.

Slide 21
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. For low-frequency characterization of devices, the three most commonly measured parameters are the H, Y and Z-parameters. All of these parameters require measuring the total voltage or current as a function of frequency at the input or output nodes (ports) of the device. Furthermore, we have to apply either open or short circuits as part of the measurement. Extending measurements of these parameters to high frequencies is not very practical.

Slide 22
At high frequencies, it is very hard to measure total voltage and current at the device ports. One cannot simply connect a voltmeter or current probe and get accurate measurements due to the impedance of the probes themselves and the difficulty of placing the probes at the desired positions. In addition, active devices may oscillate or self-destruct with the connection of shorts and opens.
Clearly, some other way of characterizing high-frequency networks is needed that doesn't have these drawbacks. That is why scattering or S-parameters were developed. S-parameters have many advantages over the previously mentioned H, Y or Z-parameters. They relate to familiar measurements such as gain, loss, and reflection coefficient. They are defined in terms of voltage traveling waves, which are relatively easy to measure. S-parameters don't require connection of undesirable loads to the device under test. The measured S-parameters of multiple devices can be cascaded to predict overall system performance. If desired, H, Y, or Z-parameters can be derived from S-parameters. And very important for RF design, S-parameters are easily imported and used for circuit simulations in electronic-design automation (EDA) tools like Agilent's Advanced Design System (ADS). S-parameters are the shared language between simulation and measurement.
An N-port device has N^2 S-parameters. So, 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 where the signal emerges, and the second number is the port where the signal is applied. So, S21 is a measure of the signal coming out port 2 relative to the RF stimulus entering port 1. When the numbers are the same (e.g., S11), it indicates a reflection measurement, as the input and output ports are the same. The incident terms (a1, a2) and output terms (b1, b2) represent voltage traveling waves.

Slide 23
S11 and S21 are determined by measuring the magnitude and phase of the incident, reflected and transmitted voltage signals when the output is terminated in a perfect Zo (a load that equals the characteristic impedance of the test system). This condition guarantees that a_2 is zero, since there is no reflection from an ideal load. S11 is equivalent to the input complex reflection coefficient or impedance of the DUT, and S21 is the forward complex transmission coefficient. Likewise, by placing the source at port 2 and terminating port 1 in a perfect load (making a_1 zero), S22 and S12 measurements can be made. S22 is equivalent to the output complex reflection coefficient or output impedance of the DUT, and S12 is the reverse complex transmission coefficient.
The accuracy of S-parameter measurements depends greatly on how good a termination we apply to the load port (the port not being stimulated). Anything other than a perfect load will result in a_1 or a_2 not being zero (which violates the definition for S-parameters). When the DUT is connected to the test ports of a network analyzer and we don't account for imperfect test-port match, we have not done a very good job satisfying the condition of a perfect termination. For this reason, two-port error correction, which corrects for source and load match, is very important for accurate S-parameter measurements (two-port correction is covered in the calibration section).

Slide 24
S-parameters are essentially the same parameters as some of the terms we have mentioned before, such as input match and insertion loss. It is important to separate the fundamental definition of S-parameters and the format in which

they are often displayed. S-parameters are inherently complex, linear quantities. They are expressed as real-and-imaginary or magnitude-and-phase pairs. However, it isn't always very useful to view them as linear pairs. Often we want to look only at the magnitude of the S-parameter (for example, when looking at insertion loss or input match), and often, a logarithmic display is most useful. A log-magnitude format lets us see far more dynamic range than a linear format.

Slide 25
Here is a generalized block diagram of a network analyzer, showing the major signal-processing sections. In order to measure the incident, reflected and transmitted signal, four sections are required:

- Source for stimulus
- Signal-separation devices
- Receivers that downconvert and detect the signals
- Processor/display for calculating and reviewing the results

We will briefly examine each of these sections. More detailed information about the signal separation devices and receiver section are in the appendix.

Slide 26
The signal source supplies the stimulus for our stimulus-response test system. We can either sweep the frequency of the source or sweep its power level. Traditionally, network analyzers used a separate source. These sources were either based on open-loop voltage-controlled oscillators (VCOs) which were cheaper, or more expensive synthesized sweepers which provided higher performance, especially for measuring narrowband devices. Excessive phase noise on open-loop VCOs degrades measurement accuracy considerably when measuring narrowband components over small frequency spans. Most network analyzers that Agilent sells today have integrated, synthesized sources, providing excellent frequency resolution and stability.

Slide 27
The next major area we will cover is the signal separation block. The hardware used for this function is generally called the "test set". The test set can be a separate box or integrated within the network analyzer. There are two functions that our signal-separation hardware must provide. The first is to measure a portion of the incident signal to provide a reference for ratioing. This can be done with splitters or directional couplers. Splitters are usually resistive. They are non-directional devices (more on directionality later) and can be very broadband. The trade-off is that they usually have 6 dB or more of loss in each arm. Directional couplers have very low insertion loss (through the main arm) and good isolation and directivity. They are generally used in microwave network analyzers, but their inherent high-pass response makes them unusable below 40 MHz or so.
The second function of the signal-splitting hardware is to separate the incident (forward) and reflected (reverse) traveling waves at the input of our DUT. Again, couplers are ideal in that they are directional, have low loss, and high reverse isolation. However, due to the difficulty of making truly broadband couplers, bridges are often used instead. Bridges work down to DC, but have more loss, resulting in less signal power delivered to the DUT. See the appendix for a more complete description of how a directional bridge works.

Slide 28
Here we consider a Wheatstone Bridge. Under the conditions listed, the bridge will be balanced, and no signal will be detected at the Rdet element; the current through Rdet will be zero, and the voltage on each side will be the same. The voltage across the bridge is Vin, and it will depend upon the impedance of the bridge, the value of Rs and the value of the source voltage.

Slide 29

We can change this into a more familiar form by replacing some of the resistive elements with “ports” represented here as transmission lines. Here Rdet is replaced with the Isolated Port

Slide 30
Here we have added the test port. Also, we will assume that the bridge is matched, and that R4 is 50 ohms. This means that R2 must also be 50 ohms.

Slide 31

Here we can see that for the forward flow of power, no power will flow into the isolated port (that's why it's isolated).

Slide 32
We can now set the loss of the bridge by determining the ratio of R1 and R2. Once we have determined that ratio, the value of R3 is also fixed, to keep the bridge balanced

Slide 33
We can redraw the bridge, without removing any elements, to show the reverse direction. Here we're part-way through redrawing the bridge, pulling the ground node down, and the junction of R2, R3, and the input port up

Slide 34
In this way, we can see that the input and the coupled ports are on opposite legs of the re-drawn bridge. Further, the loss to the input port from the test port, by symmetry, must be equal to the earlier loss computed. And we've moved the source over to the test port side, and change the input port to resemble a transmission line.
Slide 35
Here we can now clearly see that the loss from test port to input port is the same as that from input port to test port. Also, it is clear that some of the signal will couple into the Coupled Port (which we have renamed from the isolated port).

Slide 36
Here we can now test what happens when we leave the test port open. The signal that appears at the isolated port is exactly the same as the combination of Vin*loss*coupling.

Slide 37
And here we look at a short. It has the same magnitude of signal at the isolated port, but notice that the sign of the signal will be reversed (negative).

Slide 38
Unfortunately, real signal-separation devices are never perfect. For example, let's take a closer look at the actual performance of a 3-port directional coupler.
Ideally, a signal traveling in the coupler's reverse direction will not appear at all at the coupled port. In reality, however, some energy does leak through to the coupled arm, as a result of finite isolation.
One of the most important parameter for couplers is their directivity. Directivity is a measure of a coupler's ability to separate signals flowing in opposite directions within the coupler. It can be thought of as the dynamic range available for reflection measurements. Directivity can be defined as:
Directivity (dB) = Isolation (dB) - Forward Coupling Factor (dB) - Loss (through-arm) (dB)
The appendix contains a slide showing how adding attenuation to the ports of a coupler can affect the effective directivity of a system (such as a network analyzer) that uses a directional coupler.
As we will see in the next slide, finite directivity adds error to our measured results.

Slide 39
Directivity error is the main reason we see a large ripple pattern in many measurements of return loss. At the peaks of the ripple, directivity is adding in phase with the reflection from the DUT. In some cases, directivity will cancel the DUT's reflection, resulting in a sharp dip in the response.

Slide 40
The two main advantages of diode detectors are that they provide broadband frequency coverage (< 10 MHz on the low end to > 26.5 GHz at the high end) and they are inexpensive compared to a tuned receiver. Diode detectors provide medium sensitivity and dynamic range: they can measure signals to -60 dBm or so and have a dynamic range around 60 to 75 dB, depending on the detector type. Their broadband nature limits their sensitivity and makes them sensitive to source harmonics and other spurious signals. Dynamic range is improved in measurements by increasing input power.

Slide 41

Tuned receivers provide the best sensitivity and dynamic range, and also provide harmonic and spurious-signal rejection. The narrow IF filter produces a considerably lower noise floor, resulting in a significant sensitivity improvement. For example, a microwave vector network analyzer (using a tuned receiver) might have a 3 kHz IF bandwidth, where a scalar analyzer's diode detector noise bandwidth might be 26.5 GHz. Measurement dynamic range is improved with tuned receivers by increasing input power, by decreasing IF bandwidth, or by averaging. The latter two techniques provide a trade off between noise floor and measurement speed. Averaging reduces the noise floor of the network analyzer (as opposed to just reducing the noise excursions as happens when averaging spectrum analyzer data) because we are averaging complex data. Without phase information, averaging does not improve analyzer sensitivity.
The same narrowband nature of tuned receivers that produces increased dynamic range also eliminates harmonic and spurious responses. As was mentioned earlier, the RF signal is downconverted and filtered before it is measured. The harmonics associated with the source are also downconverted, but they appear at frequencies outside the IF bandwidth and are therefore removed by filtering.

Slide 42
This plot shows the effect that interfering signals (sinusoids or noise) have on measurement accuracy. The magnitude error is calculated as 20*log [1 $\pm$ interfering-signal] and the phase error is calculated as arc-sin [interfering-signal], where the interfering signal is expressed in linear terms. Note that a 0 dB interfering signal results in (plus) 6 dB error when it adds in phase with the desired signal, and (negative) infinite error when it cancels the desired signal.
To get low measurement uncertainty, more dynamic range is needed than the device exhibits. For example, to get less than 0.1 dB magnitude error and less than 0.6 degree phase error, our noise floor needs to be more than 39 dB below our measured power levels (note that there are other sources of error besides noise that may limit measurement accuracy). To achieve that level of accuracy while measuring 80 dB of rejection would require 119 dB of dynamic range. One way to achieve this level is to average test data using a tuned-receiver based network analyzer.

Slide 43
The last major block of hardware in the network analyzer is the display/processor section. This is where the reflection and transmission data is formatted in ways that make it easy to interpret the measurement results. Most network analyzers have similar features such as linear and logarithmic sweeps, linear and log formats, polar plots, Smith charts, etc. Other common features are trace markers, limit lines, and pass/fail testing. Many of network analyzers have specialized measurement features tailored to a particular market or application.

Slide 44
Let's look at the three basic sources of measurement error: systematic, random and drift.
Systematic errors are due to imperfections in the analyzer and test setup. They are repeatable (and therefore predictable), and are assumed to be time invariant. Systematic errors are characterized during the calibration process and mathematically removed during measurements.
Random errors are unpredictable since they vary with time in a random fashion. Therefore, they cannot be removed by calibration. The main contributors to random error are instrument noise (source phase noise, sampler noise, IF noise).
Drift errors are due to the instrument or test-system performance changing *after* a calibration has been done. Drift is primarily caused by temperature variation and it can be removed by further calibration(s). The timeframe over which a calibration remains accurate is dependent on the rate of drift that the test system undergoes in the user's test environment. Providing a stable ambient temperature usually goes a long way towards minimizing drift.

Slide 45
Shown here are the major systematic errors associated with network measurements. The errors relating to signal leakage are directivity and crosstalk. Errors related to signal reflections are source and load match. The final class of errors are related to frequency response of the receivers, and are called reflection and transmission tracking. The full two-port error model includes all six of these terms for the forward direction and the same six (with different data) in the reverse direction, for a total of twelve error terms. This is why we often refer to two-port calibration as twelve-term error correction

Slide 46

The two main types of error correction that can be done are response (normalization) corrections and vector corrections. Response calibration is simple to perform, but only corrects for a few of the twelve possible systematic error terms (the tracking terms). Response calibration is essentially a normalized measurement where a reference trace is stored in memory, and subsequent measurement data is divided by this memory trace. A more advanced form of response calibration is open/short averaging for reflection measurements using broadband diode detectors. In this case, two traces are averaged together to derive the reference trace.
Vector-error correction requires an analyzer that can measure both magnitude and phase. It also requires measurements of more calibration standards. Vector-error correction can account for all the major sources of systematic error and can give very accurate measurements.
Note that a response calibration can be performed on a vector network analyzer, in which case we store a complex (vector) reference trace in memory, so that we can display normalized magnitude or phase data. This is not the same as vector-error correction however (and not as accurate), because we are not measuring and removing the individual systematic errors, all of which are complex or vector quantities.

Slide 47
Vector-error correction is the process of characterizing systematic error terms by measuring known calibration standards, and then removing the effects of these errors from subsequent measurements.
One-port calibration is used for reflection measurements and can measure and remove three systematic error terms (directivity, source match, and reflection tracking). Full two-port calibration can be used for both reflection and transmission measurements, and all twelve systematic error terms are measured and removed. Two-port calibration usually requires twelve measurements on four known standards (short-open-load-through or SOLT). Some standards are measured multiple times (e.g., the through standard is usually measured four times). The standards themselves are defined in a cal-kit definition file, which is stored in the network analyzer. Agilent network analyzers contain all of the cal-kit definitions for our standard calibration kits. In order to make accurate measurements, the cal-kit definition MUST MATCH THE ACTUAL CALIBRATION KIT USED! If user-built calibration standards are used (during fixtured measurements for example), then the user must characterize the calibration standards and enter the information into a user cal-kit file. Sources of more information about this topic can be found in the appendix.

Slide 48
Taking the simplest case of a one-port reflection measurement, we have three systematic errors and one equation to solve in order to calculate the actual reflection coefficient from the measured value. In order to do this, we must first calculate the individual error terms contained in this equation. We do this by creating three more equations with three unknowns each, and solving them simultaneously. The three equations come from measuring three known calibration standards, for example, a short, an open, and a Zo load. Solving the equations will yield the systematic error terms and allow us to derive the actual reflection S-parameter of the device from our measurements.
When measuring reflection two-port devices, a one-port calibration assumes a good termination at port two of the device. If this condition is met (by connecting a load calibration standard for example), the one-port calibration is quite accurate. If port two of the device is connected to the network analyzer and the reverse isolation of the DUT is low (for example, filter passbands or cables), the assumption of a good load termination is not valid. In these cases, two-port error correction provides more accurate measurements. An example of a two-port device where load match is not important is an amplifier. The reverse isolation of the amplifier allows one-port calibration to be used effectively. An example of the measurement error that can occur when measuring a two-port filter using a one-port calibration will be shown shortly.

Slide 49
Shown here is a plot of reflection with and without one-port calibration. Without error correction, we see the classic ripple pattern caused by the systematic errors interfering with the measured signal. The error-corrected trace is much smoother and better represents the device's actual reflection performance.

Slide 50
When performing a two-port calibration, the user has the option of omitting the part of the calibration that characterizes crosstalk or isolation. The definition of crosstalk is the signal leakage between test ports when no device is present. Crosstalk can be a problem with high-isolation devices (e.g., switch in open position) and high-dynamic range devices (some filter stopbands). The isolation calibration adds noise to the error model since we usually are measuring near the noise floor of the system. For this reason, one should only perform the isolation calibration if it is really needed. If

the isolation portion of the calibration is done, trace averaging should be used to ensure that the system crosstalk is not obscured by noise. In some network analyzers, crosstalk can be minimized by using the alternate sweep mode instead of the chop mode (the chop mode makes measurements on both the reflection (A) and transmission (B) receivers at each frequency point, whereas the alternate mode turns off the reflection receiver during the transmission measurement).
The best way to perform an isolation calibration is by placing the devices that will be measured on each test port of the network analyzer, with terminations on the other two device ports. Using this technique, the network analyzer sees the same impedance versus frequency during the isolation calibration as it will during subsequent measurements of the DUT. If this method is impractical (in test fixtures, or if only one DUT is available, for example), than placing a terminated DUT on the source port and a termination on the load port of the network analyzer is the next best alternative (the DUT and termination must be swapped for the reverse measurement). If no DUT is available or if the DUT will be tuned (which will change its port matches), then terminations should be placed on each network analyzer test port for the isolation calibration.

Slide 51
A network analyzer can be used for uncorrected measurements, or with any one of a number of calibration choices, including response calibrations and one- or two-port vector calibrations. A summary of these calibrations is shown above. We will explore the measurement uncertainties associated with the various calibration types in this section.

Slide 52
Here is an example of how much measurement uncertainty we might encounter when measuring the input match of a filter after a one-port calibration. In this example, our filter has a return loss of 16 dB, and 1 dB of insertion loss. Let's say the raw (uncorrected) load match of our network analyzer is specified to be 18 dB (generally, typical performance is significantly better than the specified performance). The reflection from the test port connected to the filter's output is attenuated by twice the filter loss, which is only 2 dB total in this case. This value is not adequate to sufficiently suppress the effects of this error signal, which illustrates why low-loss devices are difficult to measure accurately. To determine the measurement uncertainty of this example, it is necessary to convert all dB values into linear values. Next, we must add and subtract the undesired reflection signal resulting from the load match (with a reflection coefficient of 0.100) with the signal reflecting from the DUT (0.158). To be consistent with the next example, we will also include the effect of the directivity error signal (0.010). As a worst case analysis, we will add this signal to the error signal resulting from the load match. The combined error signal is then 0.100 + 0.010 = 0.110. When we add and subtract this error signal from the desired 0.158, we see the measured return loss of the 16-dB filter may appear to be anywhere from 11.4 dB to 26.4 dB, allowing too much margin for error. In production testing, these errors could easily cause filters which met specification to fail, while filters that actually did not meet specification might pass. In tuning applications, filters could be mistuned as operators try to compensate for the measurement error.
When measuring an amplifier with good isolation between output and input (i.e., where the isolation is much greater than the gain), there is much less measurement uncertainty. This is because the reflection caused by the load match is severely attenuated by the product of the amplifier's isolation and gain. To improve measurement uncertainty for a filter, the output of the filter must be disconnected from the analyzer and terminated with a high-quality load, or a high-quality attenuator can be inserted between the filter and port two of the network analyzer. Both techniques improve the analyzer's effective load match.

Slide 53
Let's do an example transmission measurement using only response calibration. Response calibrations offer simplicity, but with some compromise in measurement accuracy. In making a filter transmission measurement using only response calibration, the first step is to make a through connection between the two test port cables (with no DUT in place). For this example, some typical test port specifications will be used. The ripple caused by this amount of mismatch is calculated as ± 0.22 dB, and is now present in the reference data. Since we don't know the relative phase of this error signal once it passes through the DUT, it must be added to the uncertainty when the DUT is measured (see next slide) in order to compute the worst-case overall measurement uncertainty.

Slide 54
Now let's look at the measurement uncertainty when the DUT is inserted. We will use the same loss and mismatch specifications for the DUT and analyzer as before. We have three main error signals due to reflections between the ports of the analyzer and the DUT. Higher-order reflections are present as well, but they don't add any significant error since they are small compared to the three main terms. In this example, we will normalize the error terms to the

desired signal that passes through the DUT once. The desired signal is therefore represented as 1, and error terms only show the additional transmission loss due to traveling more than once through the DUT. One of the signals passes through the DUT two extra times, so it is attenuated by twice the loss of the DUT. The worst case is when all of the reflected error signals add together in-phase (.020 + .020 + .032 = .072). In that case, we get a measurement uncertainty of +0.60 dB, -0.65 dB. The total measurement uncertainty, which must include the 0.22 dB of error incorporated into our calibration measurement, is about ± 0.85 dB.

Slide 55
Now let's look at an example of measuring an amplifier that has port matches of 16 dB. The match of our test ports remains the same as our previous transmission response example. We see that the middle error term is no longer present, due to the reverse isolation of the amplifier. This fact has reduced our measurement uncertainty to about ± 0.45 dB. Our total measurement error now has been reduced to about ± 0.67 dB, versus the ± 0.85 dB we had when measuring the filter.

Slide 56
Two-port error correction is the most accurate form of error correction since it accounts for all of the majorsources of systematic error. The error model for a two-port device is shown above. Shown below are the equations to derive the actual device S-parameters from the measured S-parameters, once the systematic error terms have been characterized. Notice that each actual S-parameter is a function of all four measured S-parameters. The network analyzer must make a forward and reverse sweep to update any one S-parameter. Luckily, you don't need to know these equations to use network analyzers!!!

Slide 57
Let's look at some actual measurements done on a bandpass filter with different levels of error correction. The uncorrected trace shows considerable loss and ripple. In fact, the passband response varies about ± 1 dB around the filter's center frequency. Is the filter really this bad? No. What we are actually measuring is the sum of the filter's response and that of our test system.
Performing a normalization prior to the measurement of the filter removes the frequency response of the system (transmission tracking error) from the measurement. The loss that was removed was most likely caused by the test cables. After normalization, the frequency response of the filter still contains ripple caused by an interaction between the system's source and load match. This ripple even goes above the 0 dB reference line, indicating gain! However, we know that a passive device cannot amplify signals. This apparent anomaly is due to mismatch error.
The measurement shown after a two-port calibration is the most accurate of the three measurements shown. Using vector-error correction, the filter's passband response shows variation of about ± 0.1 dB around its center frequency. This increased level of measurement flatness will ensure minimum amplitude distortion, increase confidence in the filter's design, and ultimately increase manufacturing yields due to lower test-failure rates.

Slide 58
Whenever possible, reflection calibrations should be done with a cal kit that matches the connector type of the DUT. If adapters need to be used to mate the calibrated test system to the DUT, the effect of these adapters on measurement accuracy can be very large. This error is often ignored, which may or may not be acceptable. As the slide shows, the adapter causes an error signal which can add or subtract with the desired reflection signal from the DUT. Worst-case effective directivity (in dB) is now:

$$-20 \log (\text{Corrected-coupler-directivity} + \rho_{\text{adapters}})$$

If the adapter has a SWR of say 1.5 (the less-expensive variety), the effective directivity of the coupler drops to around 14 dB worst case, even if the coupler itself had infinite directivity! In other words, with a perfect Zo load ($\rho_L = 0$) on the output of the adapter; the reflected signal appearing at the coupled port would only be 14 dB less than the reflection from a short or open circuit. Stacking adapters compounds the problem, as is illustrated above. Consequently, it is very important to use quality adapters (or preferably, no adapters at all) in your measurement system, so system directivity is not excessively degraded. While error-correction can mitigate the effect of adapters on the test port, our system is more susceptible to drift with degraded raw (uncorrected) directivity.

Slide 59
Although the previous slides have all shown mechanical calibration standards, Agilent offers a solid-state calibration solution which makes two, three, and four-port calibration fast, easy, and less prone to operator errors. A variety of

calibration modules are available with different connector types and frequency ranges. You can configure a single module with different connector types or choose all the same type. The calibration modules are solid-state devices with programmable, repeatable impedance states. These states are characterized at the Agilent factory using a network analyzer calibrated with coaxial, airline-TRL standards (the best calibration available), making the ECal modules transfer standards (rather than direct standards).
For the microwave calibration modules, the various impedance states are achieved by PIN-diode switches which shunt the transmission line to ground. The number of diodes and their location vary depending upon the module's frequency range. A multitude of reflection coefficients can be generated by applying various combinations of the shunts. With no shunts, the network acts as a low loss transmission line. High isolation between the ports is obtained by driving several of the PIN shunts simultaneously. Four different states are used to compute the error terms at each frequency point. Four states are used because this gives the best trade-off between high accuracy and the time required for the calibration. With four reflection states, we have four equations but only three unknowns. To achieve the best accuracy from this over-determined set of equations, a least-squares-fit algorithm is used. Adding more impedance states at each frequency point would further improve accuracy but at the expense of more calibration time.
The RF module uses the more traditional short, open, and load terminations, and a through transmission line.

Slide 60
When performing a two-port calibration, we have some choices based on the type of calibration standards we want to use. So far, we have only discussed coaxial calibration techniques. Let's briefly look at TRL (through-reflect-line), a calibration technique that is especially useful for microwave, noncoaxial environments such as fixture, wafer probing, or waveguide. It is the second-most common type of two-port calibration, after SOLT. TRL solves for the same 12 error terms as the more common SOLT calibration, but uses a slightly different error model.
The main advantage of TRL is that the calibration standards are relatively easy to make and define at microwave frequencies. This is a big benefit since is difficult to build good, noncoaxial, open and load standards at microwave frequencies. TRL uses a transmission line of known length and impedance as one standard. The only restriction is that the line needs to be significantly longer in electrical length than the through line, which typically is of zero length. TRL also requires a high-reflection standard (usually, a short or open) whose impedance does not have to be well characterized, but it must be electrically the same for both test ports.
For RF applications, the lengths of the transmission lines needed to cover down to low frequencies become impractical (too long). It is also difficult to make good TRL standards on printed-circuit boards, due to dielectric, line-dimension, and board-thickness variations. And, the typical TRL fixture tends to be more complicated and expensive, due do the need to accommodate throughs of two different physical lengths.
There are two variations of TRL. True TRL calibration requires a 4-receiver network analyzer. The version for three-receiver analyzers is called TRL* ("TRL-star"). Other variations of this type of calibration (that share a common error model) are Line-Reflect-Line (LRL), Line-Reflect-Match (LRM), Thru-Reflect-Match (TRM), plus many others.

Slide 61
Take care of your network analyzer
Always use an adpater on the port of the analyzer
Never drive too much power into the Network Analyzer
Watch out for running too much bias current through the NA
Never drive too much power into the Network Analzyer
Don’t hood up DC voltage directly to the NA (use the bias tees)
Touch the case of the NA first before touching the cable ends (discharge your ESD).
Did I say “Don’t drive too much power into the NA”?

Slide 62
Start by setting up the start/stop/number of points for your measurement, under the Stimulus block
Set the IF BW: 1 KHz for precise measurements, 10 kHz for fast.
Set the power if you’re measuring an active device, to avoid over driving the NA
Select the traces: on the ENA select “display traces” to change then number of traces shown.
Hit the Meas key to select what parameter to display
Hit the MARKER key to put one (or more) markers on the screen

Slide 63
Use the FORMAT to change between Log and Linear
Use the Scale key to bring up the scale. Use autoscale or select the scale in dB/div, the reference live value, and reference line position
Use the Data->Memory and Data&Mem to save compare traces (DISPLAY)
Save your data using “Save S2P”
Use the equation editor to change the value of your trace
Use Save/Recall to save your setups