Agenda

- **8:45** Start of registration of participants
- **9:15** Start of seminar - welcome
- **9:20 – 10:15** Introduction into Vector Network Analysis
- **10:15 – 10:30** Coffee break
- **10:30 – 11:30** Using network analyzers to measure active devices (e.g. power amplifiers)
- **11:30 – 12:30** Lunch
- **12:30 – 14:30** Modern VNA as complete RF lab (contains various measurement examples like noise figure, deembeding, frequency converter ...)
- **14:30** Discussion, end of seminar

KEYSIGHT

Vector Network Analysis Fundamentals

Andreas Tenggren
Solutions Engineering

Keysight Technologies

Agenda

01 Getting started – Important concepts

- What is a network analyzer?
- Transmission lines and S-parameters

02 Getting to know network analyzers

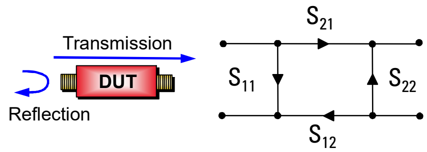
- Block diagram: how do network analyzers work?
- Network analysis measurements
- Difference between network analyzers and spectrum analyzers

03 Calibration and error correction

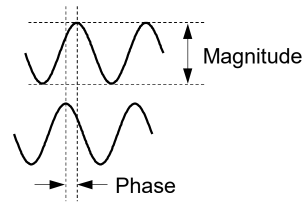
- Why and how?

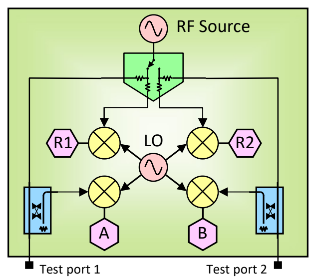
04 Summary

What Are Vector Network Analyzers?



- Are stimulus-response test systems
- Characterize forward and reverse reflection and transmission responses (S-parameters) of RF and microwave components
- Quantify linear magnitude and phase
- Are very fast for swept measurements
- Provide the highest level of measurement accuracy





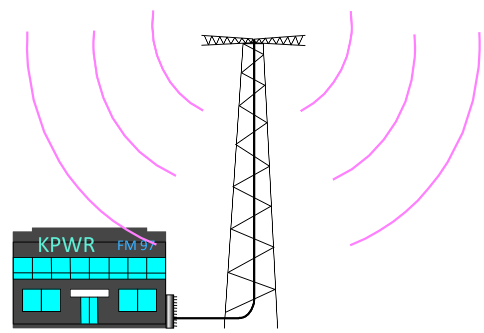
4

A vector network analyzer (VNA) is a precision measuring tool that tests the electrical performance of high frequency components, in the radio frequency (RF), microwave, and millimeter-wave frequency bands (we will use the generic term RF to apply to all of these frequencies). A VNA is a stimulus-response test system, composed of an RF source and multiple measurement receivers. It is specifically designed to measure the forward and reverse reflection and transmission responses, or S-parameters, of RF components. S-parameters have both a magnitude and a phase component, and they characterize the linear performance of the DUT. While VNAs can also be used for characterizing some non-linear behavior like amplifier gain compression or intermodulation distortion, S-parameters are the primary measurement. The network analyzer hardware is optimized for speed, yielding swept measurements that are faster than those obtained from the use of an individual source and an individual receiver like a spectrum analyzer. Through calibration, VNAs provide the highest level of accuracy for measuring RF components.

Why do we need to test components?

- Verify specifications of “building blocks” for more complex RF systems

- Ensure distortionless transmission of communications signals
  - Linear: constant amplitude, linear phase/constant group delay
  - Nonlinear: harmonics, intermodulation, compression, X-parameters
- Ensure good match when absorbing power



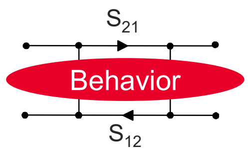
Components are tested for a variety of reasons. Many components are used as "building blocks" in more complicated RF systems. For example, in most transceivers there are amplifiers to boost LO power to mixers, and filters to remove signal harmonics. Often, R&D engineers need to measure these components 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.

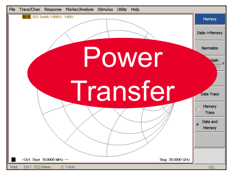
Often it is most important to measure how reflective a component is, to ensure that it absorbs energy efficiently. Measuring antenna match is a good example.

The need for both magnitude and phase

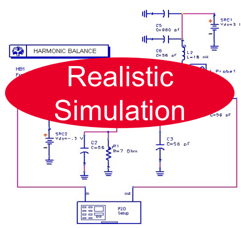
1. Complete characterization of linear networks



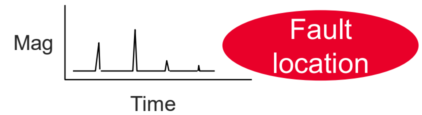
2. Complex impedance needed to design matching circuits



3. Complex values needed for device modeling



4. Time-domain characterization

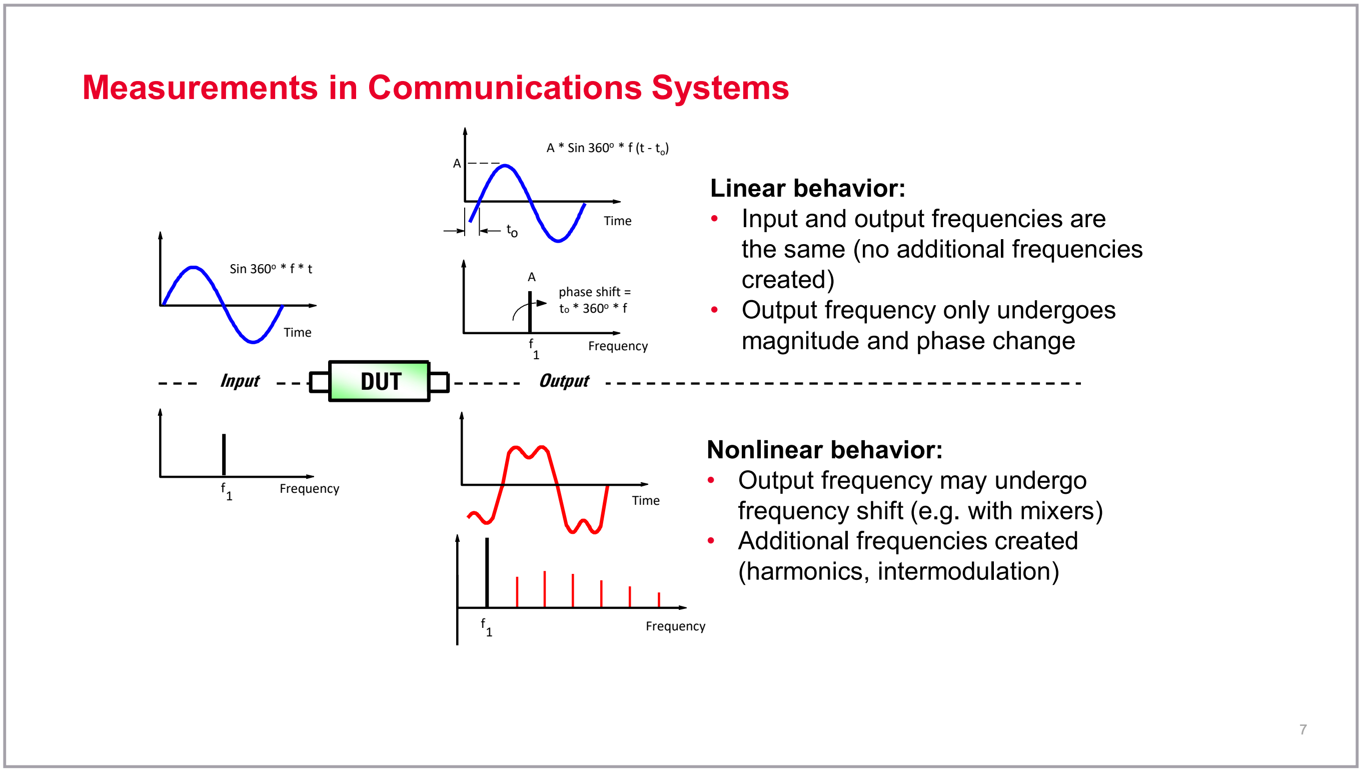


5. Vector-error correction

6. X-parameter (nonlinear) characterization

Pre-distortion

Measurements with both magnitude and phase are required for the best measurement accuracy and best calibration. Complete characterization of devices and networks involves measurement of magnitude and phase. 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.



The possibility to characterize a device in term of magnitude and phase characteristic is very important to understand the behavior of the device itself.

Let's have a look as example of a device that is on use in a communication application, where the aim is to carry a signal from on point to another with the power efficiency and fidelity.

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.

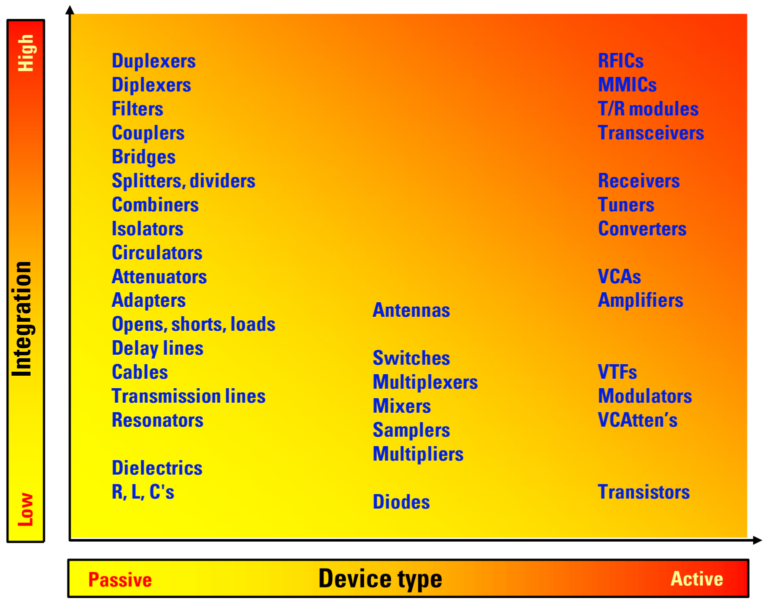
Let's examine the difference between linear and nonlinear behavior more closely. Linear devices impose magnitude and phase changes on input signals (top part of the slide).

Any sinusoid appearing at the input will also appear at the output, and at the same frequency. No new signals are created.

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 (bottom part of the slide).

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.

What Types of Devisces are Tested?



8

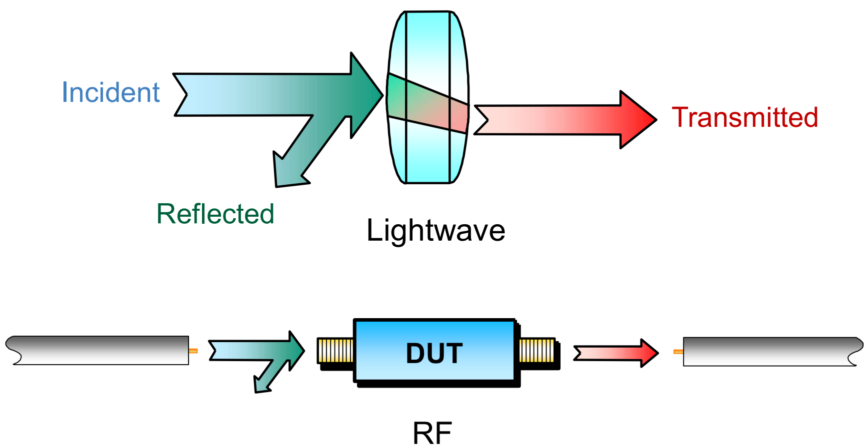
Here are some examples of the types of devices that you can test with network analyzers. They include both passive and active devices (and some that have attributes of both). Many of these devices need to be characterized for both linear and nonlinear behavior. It is not possible to completely characterize all of these devices with just one piece of test equipment.

The next slide shows a model covering the wide range of measurements necessary for complete linear and nonlinear characterization of devices. This model requires a variety of stimulus and response tools. It takes a large range of test equipment to accomplish all of the measurements shown on this chart. Some instruments are optimized for one test only (like bit-error rate), while others, like network analyzers, are much more general-purpose in nature. Network analyzers can measure both linear and nonlinear behavior of devices, although the measurement techniques are different (frequency versus power sweeps for example). This module focuses on swept-frequency and swept-power measurements made with network analyzers

Getting started – Important concepts

- What is a network analyzer
- Transmission lines and S-parameters

The Basis of Incident and Reflected Power

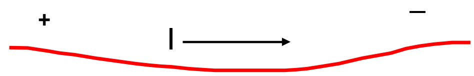


One of the most fundamental concepts of high-frequency network analysis involves incident, reflected and transmitted waves traveling along transmission lines. It is helpful to think of traveling waves along a transmission line in terms of a lightwave analogy. We can imagine incident light striking some 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.

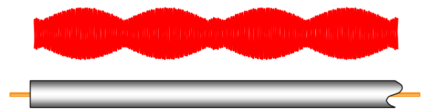
Vector network analyzers accurately measure the incident, reflected, and transmitted energy, e.g., the energy that is launched onto a transmission line, reflected back down the transmission line toward the source (due to impedance mismatch), and successfully transmitted to the terminating device (such as an antenna).

Transmission Line Basics

Low frequencies



- Wavelengths >> wire length
- Current (I) travels down wires easily for efficient power transmission
- Measured voltage and current not dependent on position along wire



High frequencies

- Wavelength » or << length of transmission medium
- Need transmission lines for efficient power transmission
- Matching to characteristic impedance (Zo) is very important for low reflection and maximum power transfer
- Measured envelope voltage dependent on position along line

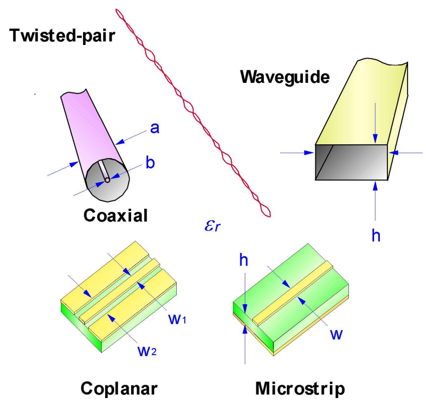
The need for efficient transfer of RF power is one of the main reasons behind the use of transmission lines. At low frequencies where the wavelength of the signals are much larger than the length of the circuit conductors, a simple wire is very useful for carrying power. Current travels down the wire easily, and voltage and current are the same no matter where we measure along the wire.

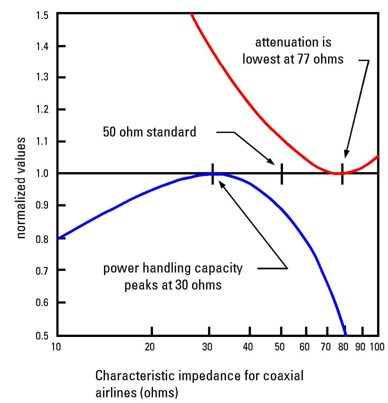
At high frequencies however, the wavelength of signals of interest are comparable to or much smaller than the length of conductors. In this case, power transmission can best be thought of in terms of traveling waves.

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

Transmission line Z_o

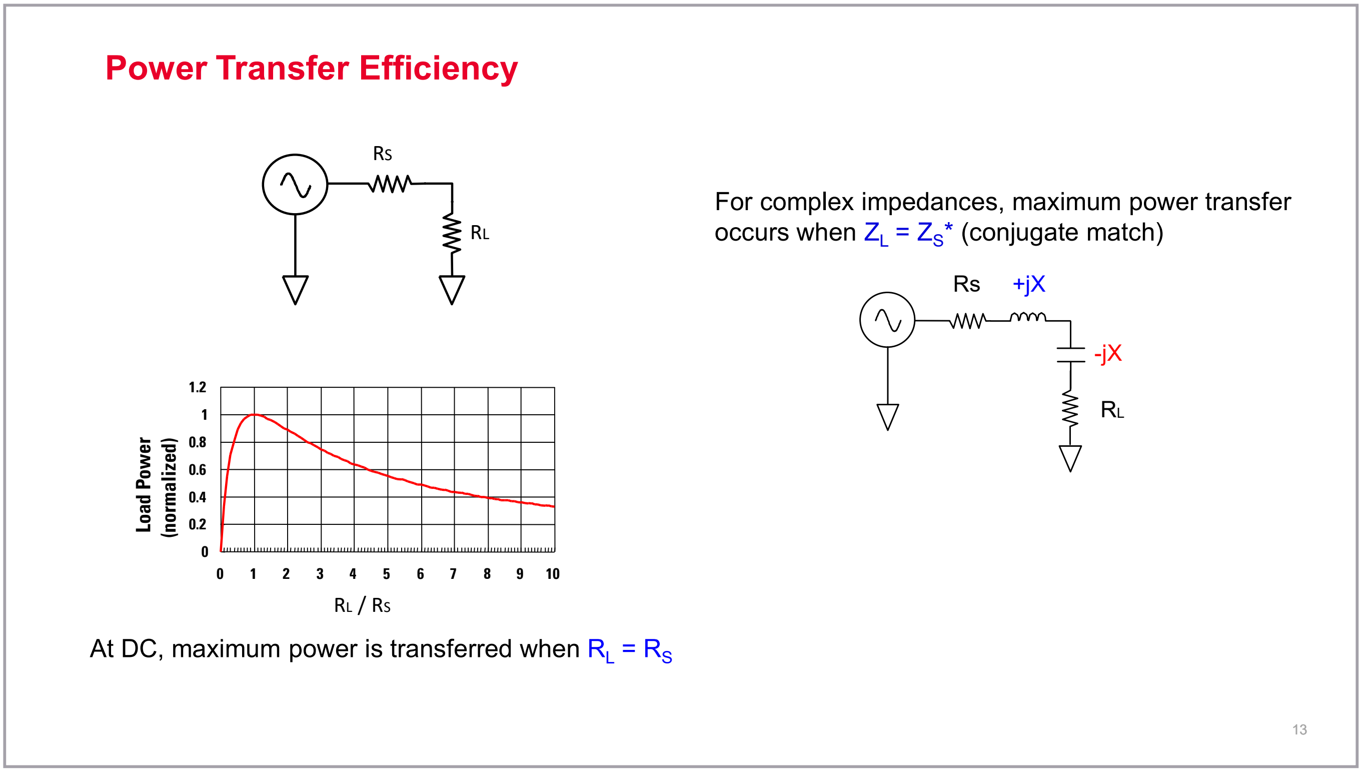
- Z_o determines relationship between voltage and current waves
- Z_o is a function of physical dimensions and ε_r
- Z_o is usually a real impedance (e.g. 50 or 75 ohms)





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 (er) 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.

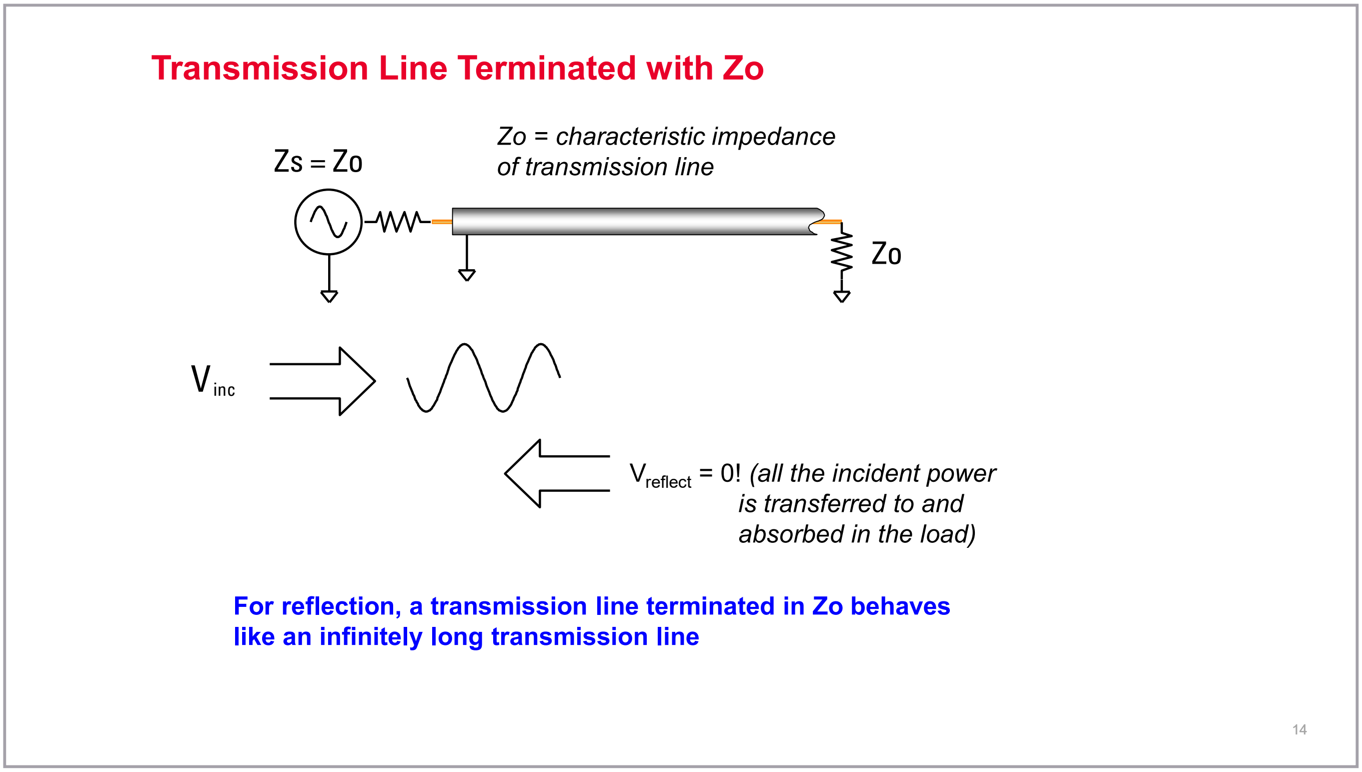


A perfectly matched condition must exist at a connection between two devices for maximum power transfer into a load, given a source resistance of RS and a load resistance of RL. This condition occurs when RL = RS, and is true whether the stimulus is a DC voltage source or a source of RF sine waves.

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 RS = 0.6 + j 0.3, then the complex conjugate is RS* = 0.6 – j 0.3.

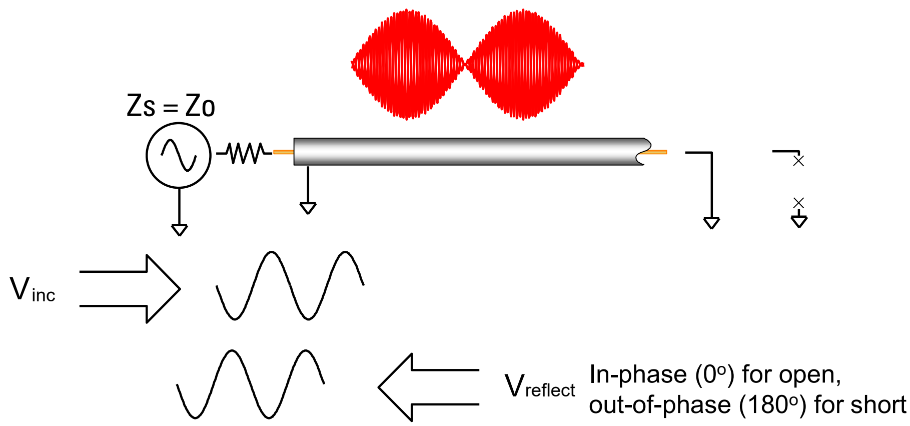
The need for efficient power transfer is one of the main reasons for the use of transmission lines at higher frequencies. At very low frequencies (with much larger wavelengths), a simple wire is adequate for conducting power. The resistance of the wire is relatively low and has little effect on low-frequency 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.



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. On the smith chat we will see a point on the center equivalent to Zo

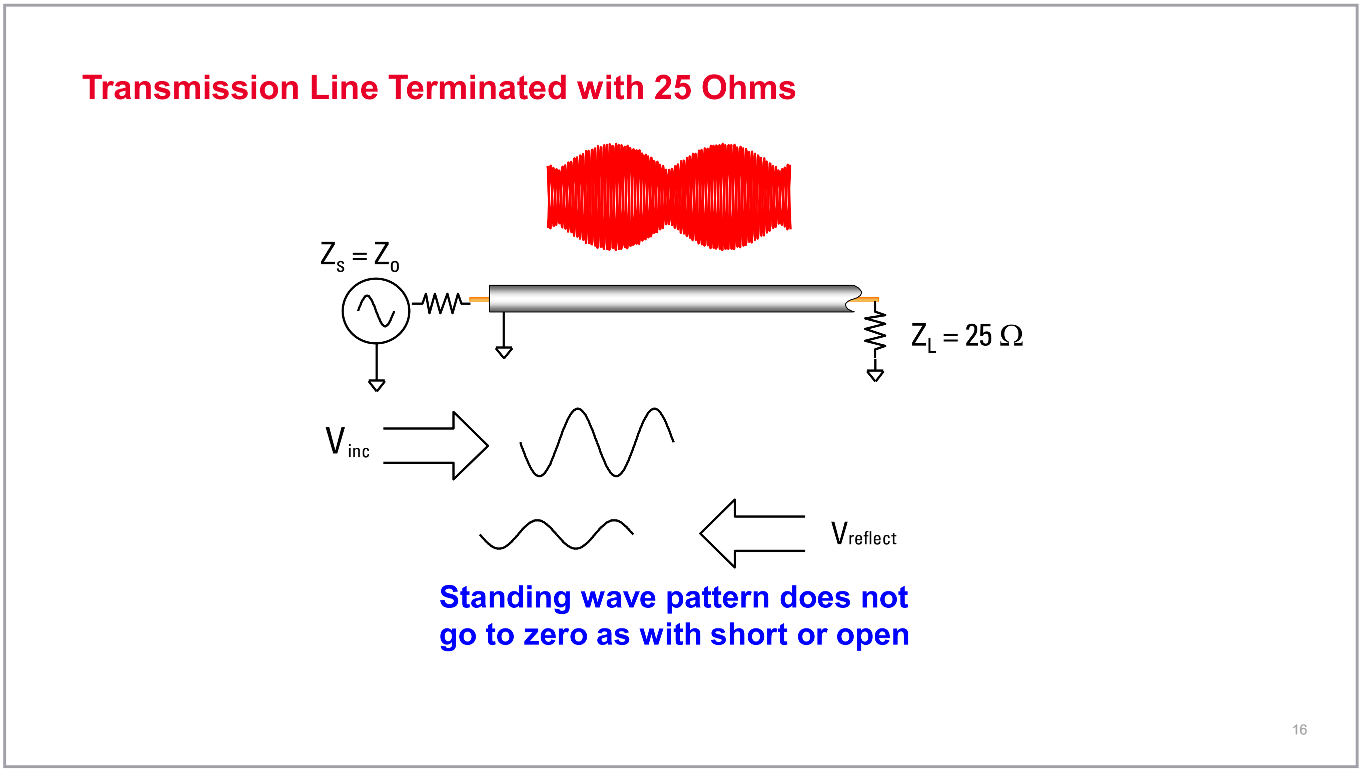
Transmission Line Terminated with Short, Open



For reflection, a transmission line terminated in a short or open reflects all power back to source

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 180o 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.

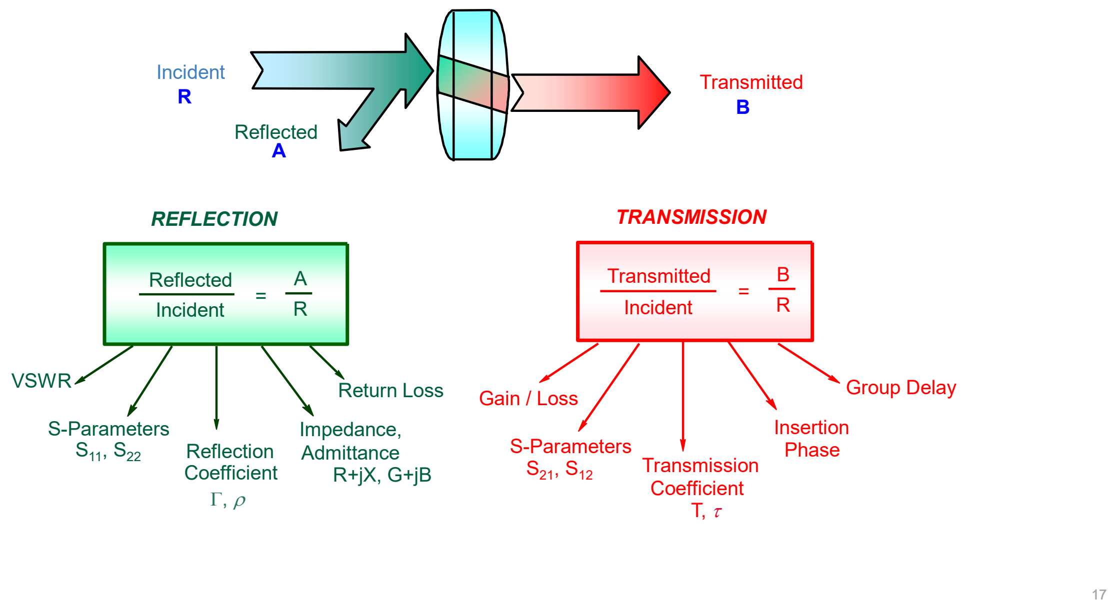
Now let us leave our line open. This time, Ohm's law tells us that the open can support no current. Therefore, our reflected current wave must be 180o out of phase with respect to the incident wave (the voltage wave will be in phase with the incident wave). This guarantees that current at the open will 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.



Finally, let's terminate our line with a 25 ohm resistor (an impedance between the full reflection of an open or short circuit and the perfect termination of a 50 ohm 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 180o 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 180o 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.

Vector Network Analysis Terminology

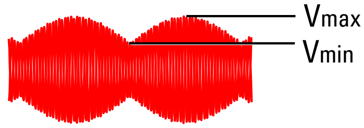


Now that we understand the fundamentals of electromagnetic waves, we must learn the common terms used for measuring them. Vector network analyzer terminology generally denotes measurements of the incident wave with the R or reference channel. The reflected wave is measured with the A channel, and the transmitted wave is measured with the B channel. 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.

Reflection Parameters

Reflection Coefficient $\Gamma = \dfrac{V_{reflected}}{V_{incident}} = \rho \angle \Phi = \dfrac{Z_L - Z_o}{Z_L + Z_o}$

Return loss $= -20 \log(\rho)$ $\qquad \rho = |\Gamma|$



Voltage Standing Wave Ratio

$$\text{VSWR} = \frac{V_{max}}{V_{min}} = \frac{1+\rho}{1-\rho}$$

No reflection ***($Z_L = Z_o$)***

Full reflection ***(Z_L = open, short)***

0	ρ	1
∞ dB	**RL**	0 dB
1	**VSWR**	∞

Γ = Gamma (reflection coefficient) $\quad \rho$ = *Rho (Magnitude)* $\quad \Phi$ = Phi (Phase)

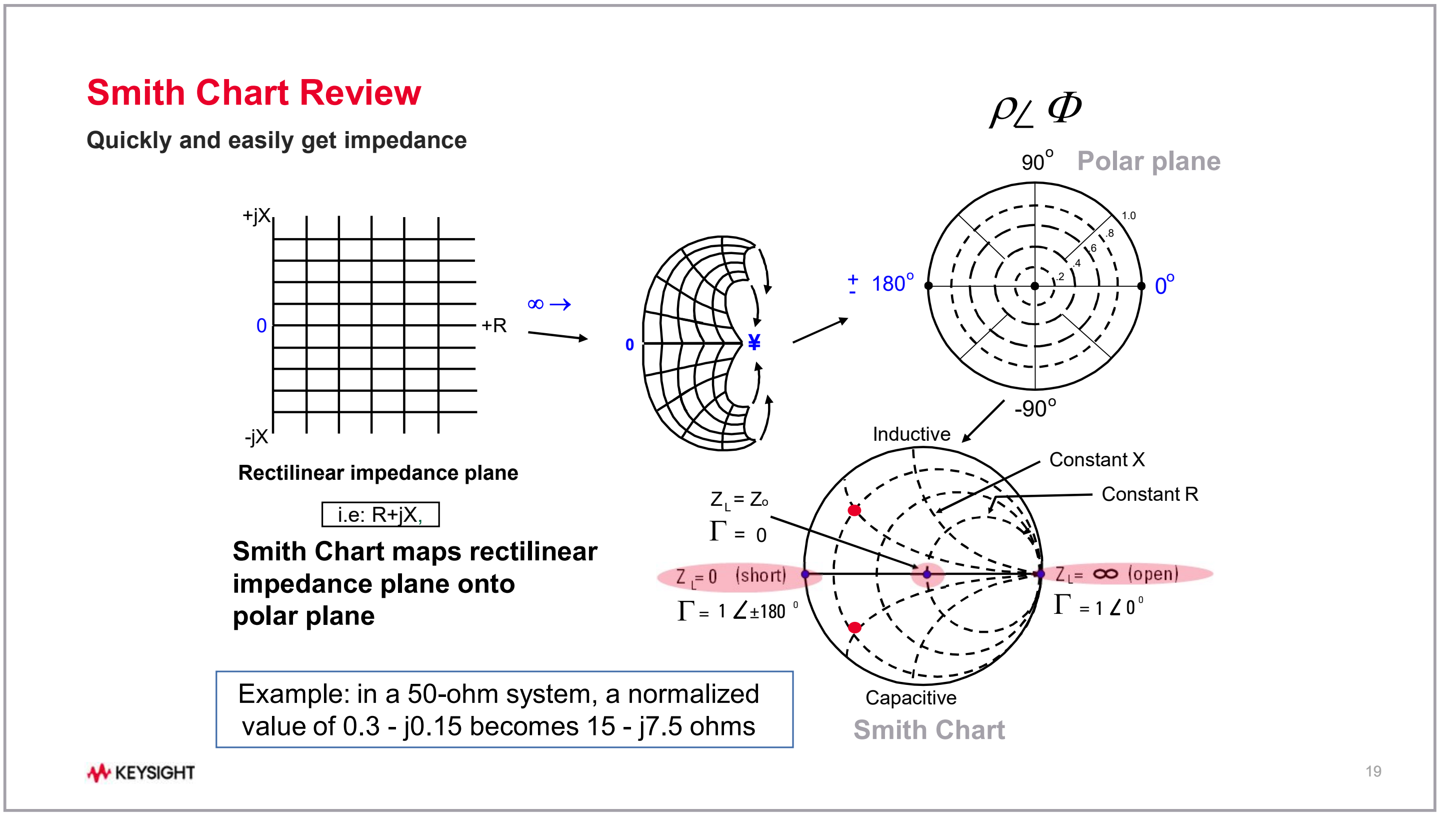
18

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 Vrefl = 0 and ρ = 0. When ZL is not equal to Zo , some energy is reflected and ρ is greater than zero. When ZL is a short or open circuit, all energy is reflected and ρ = 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 on a value between 1 for a Zo impedance and ∞ for an open or short circuit.



Our network analyzer gives us complex reflection coefficient. However, we often want to know the impedance of the DUT. The previous slide shows the relationship between reflection coefficient and impedance, and we could manually perform the complex math to find the impedance. Although programmable calculators and computers take the drudgery out of doing the math, a single number does not always give us the complete picture. In addition, impedance almost certainly changes with frequency, so even if we did all the math, we would end up with a table of numbers that may be difficult to interpret.

A simple, graphical method solves this problem. Let's first plot reflection coefficient using a polar display. For positive resistance, the absolute magnitude of Γ varies from zero (perfect load) to unity (full reflection) at some angle. So we have a unit circle, which marks the boundary of the polar plane shown on the slide. An open would plot at $1 \angle 0^\circ$; a short at $1 \angle 180^\circ$; a perfect load at the center, and so on. How do we get from the polar data to impedance graphically? Since there is a one-to-one correspondence between complex reflection coefficient and impedance, we can map one plane onto the other. If we try to map the polar plane onto the rectilinear impedance plane, we find that we have problems. **First of all, the rectilinear plane does not have values to infinity. Second, circles of constant reflection coefficient are concentric on the polar plane but not on the rectilinear plane, making it difficult to make judgments regarding two different impedances. Finally, phase angles plot as radii on the polar plane but plot as arcs on the rectilinear plane, making it difficult to pinpoint**.

The proper solution was first used in the 1930's, when Phillip H. Smith mapped the impedance plane onto the polar plane, creating the chart that bears his name (the venerable *Smith chart*). **Since unity at zero degrees on the polar plane represents infinite impedance, both plus and minus infinite reactances, as well as infinite resistance can be plotted. On the Smith chart, the vertical lines on the rectilinear plane that indicate values of constant resistance map to circles, and the horizontal lines that indicate values of constant reactance map to arcs. Zo maps to the exact center of the chart**.

In general, **Smith charts are normalized to Zo; that is, the impedance values are divided by Zo**. The chart is then independent of the characteristic impedance of the system in question. Actual impedance values are derived by multiplying the indicated value by Zo. For example, in a 50-ohm system, a normalized value of 0.3 - j0.15 becomes 15 - j7.5 ohms; in a 75-ohm system, 22.5 - j11.25 ohms.

Fortunately, we no longer have to go through the exercise ourselves. **Out network analyzer can display the Smith chart, plot measure data on it, and provide adjustable markers that show the calculated impedance at the marked point in a several marker formats.**

Transmission Parameters

$$\text{Transmission Coefficient} = \mathrm{T} = \frac{V_{Transmitted}}{V_{Incident}} = \tau\angle\phi$$

$$\text{Insertion Loss (dB)} = -20 \text{ Log} \left| \frac{V_{Trans}}{V_{Inc}} \right| = -20 \text{ Log}(\tau)$$

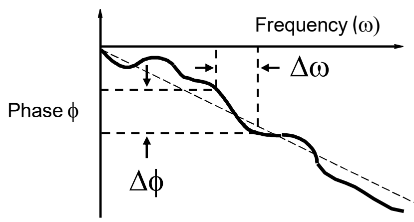
$$\text{Gain (dB)} = 20 \text{ Log} \left| \frac{V_{Trans}}{V_{Inc}} \right| = 20 \text{ Log}(\tau)$$

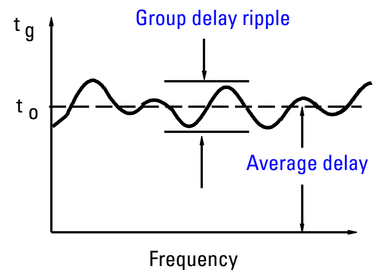
T = Tau (Transmission Coefficient) τ = Small Tau (Magnitude) Φ = Phi (Phase)

Transmission coefficient T is defined as the transmitted voltage divided by the incident voltage. If |Vtrans| > |Vinc|, the DUT has gain, and if |Vtrans| < |Vinc|, 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.

Group Delay





$$\text{Group Delay } (t_g) = \frac{-\Delta\phi}{\Delta\omega} = \frac{-1}{360^{\circ}} * \frac{\Delta\phi}{\Delta f}$$

ϕ in radians

ω in radians/sec

ϕ in degrees

f in Hertz ($\omega = 2\pi f$)

- Group-delay ripple indicates phase distortion, calculated by differentiating insertion-phase response Vs. frequency
- Average delay indicates electrical length of DUT
- Aperture ($\Delta\phi$) of measurement is very important

Derivera / Differentiating

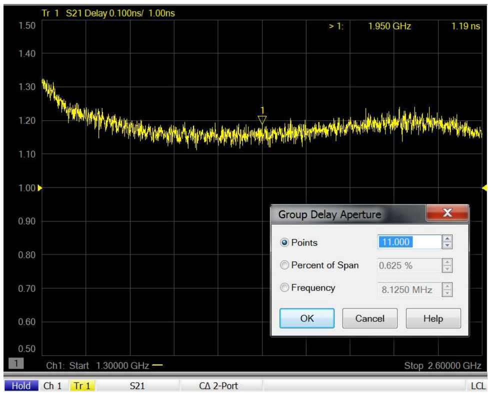
Another useful measure of phase distortion is group delay. Group delay is a measure of the transit time of a signal through the device under test, versus frequency. Group delay is calculated by differentiating the insertion-phase response of the DUT versus frequency. Another way to say this is that group 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.

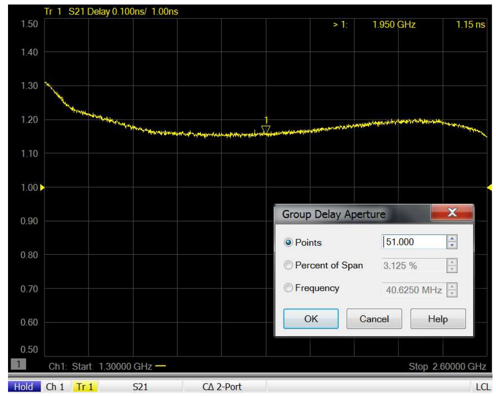
Group Delay

Example measurement on amplifier

Group Delay Aperture 11points (8MHz)



Group Delay Aperture 51points (40MHz)



A larger aperture reduces measurement noise but it comes with the price of reduced resolution over frequency

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).

Trace Format – Selects What Data To Display

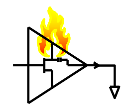
With trace format you select what data to display. For example "Log Mag" for S11/S22 is the same as Return Loss (reflection). For a transmission parameter (S21/S12) "Log Mag" displays gain/loss

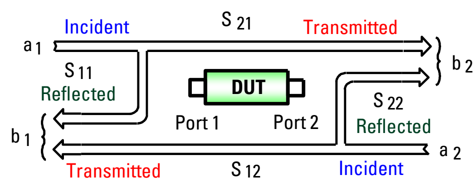
23

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).

S-Parameters (Why use it?)

- Measure voltage traveling waves with a vector network analyzer
- Don't need shorts/opens (can cause active devices to oscillate or self-destruct)
- Relate to **familiar** measurements (gain, loss, reflection coefficient ...)
- Can **cascade** S-parameters of multiple devices to predict system performance
- Can **compute** H-, Y-, or Z-parameters from S-parameters if desired
- Can easily import and use S-parameter files in **electronic-simulation** tools





$$b_1 = S_{11} a_1 + S_{12} a_2$$

$$b_2 = S_{21} a_1 + S_{22} a_2$$

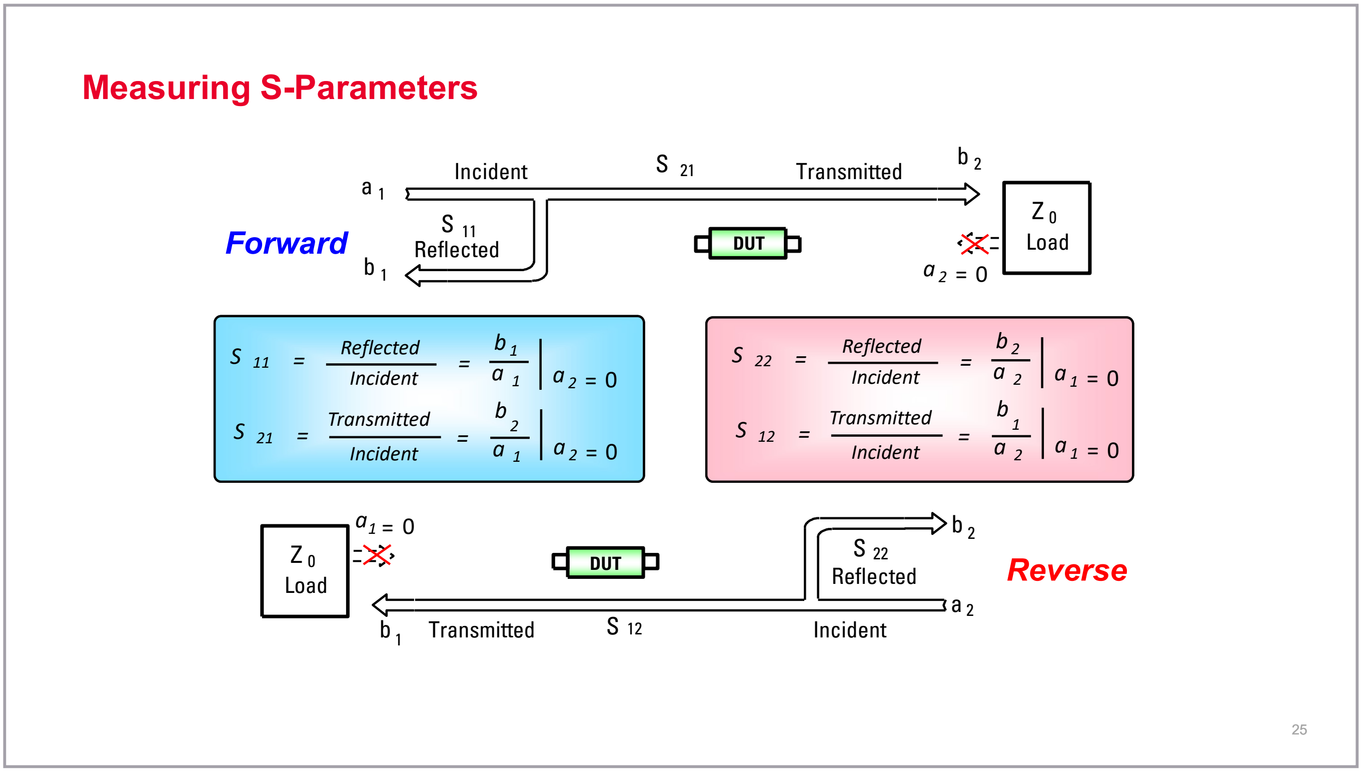
24

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 Keysight's Advanced Design System (ADS). S-parameters are the shared language between simulation and measurement.

An N-port device has N2 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.



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 a2 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 a1 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 a1 or a2 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).

Equating S-parameters with common measurement terms

Remember, S-parameters are inherently complex, linear quantities – however, we often express them in a log-magnitude format

S_{11} = forward reflection coefficient ***(input match)***

S_{22} = reverse reflection coefficient ***(output match)***

S_{21} = forward transmission coefficient ***(gain or loss)***

S_{12} = reverse transmission coefficient ***(isolation)***

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.

Saving S-parameters

S-parameters are saved in a standardized Touchstone format with extension *.SnP where n=number of ports in measurement (ex. *.S1P, *.S2P … …)

```
!Keysight Technologies,N5242A,MY56023156,A.10.49.11
!Date: Monday, February 27, 2017 11:27:08
!Correction: S11(Full 2 Port(1,2))
!S21(Full 2 Port(1,2))
!S12(Full 2 Port(1,2))
!S22(Full 2 Port(1,2))
!S2P File: Measurements: S11, S21, S12, S22:
# Hz S  dB   R 50        S11                    S21                    S12                    S22
1300000000 -14.000709 -84.717308 32.761486 -60.747711 -45.185192 157.18211 -30.362373 -74.943069
1300812500 -14.040403 -84.479553 32.742043 -61.134171 -45.35677 161.31783 -30.100229 -75.812248
1301625000 -14.047897 -85.194618 32.740654 -61.568413 -44.876976 154.43845 -30.224581 -74.018242
1302437500 -14.05684 -85.052895 32.734764 -61.895115 -45.969913 158.91109 -30.147123 -75.628639
1303250000 -14.075928 -85.520554 32.729805 -62.262146 -45.81414 155.65988 -30.140175 -73.774292
.
.
.
```

Frequency   Magnitude   Phase

27

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.

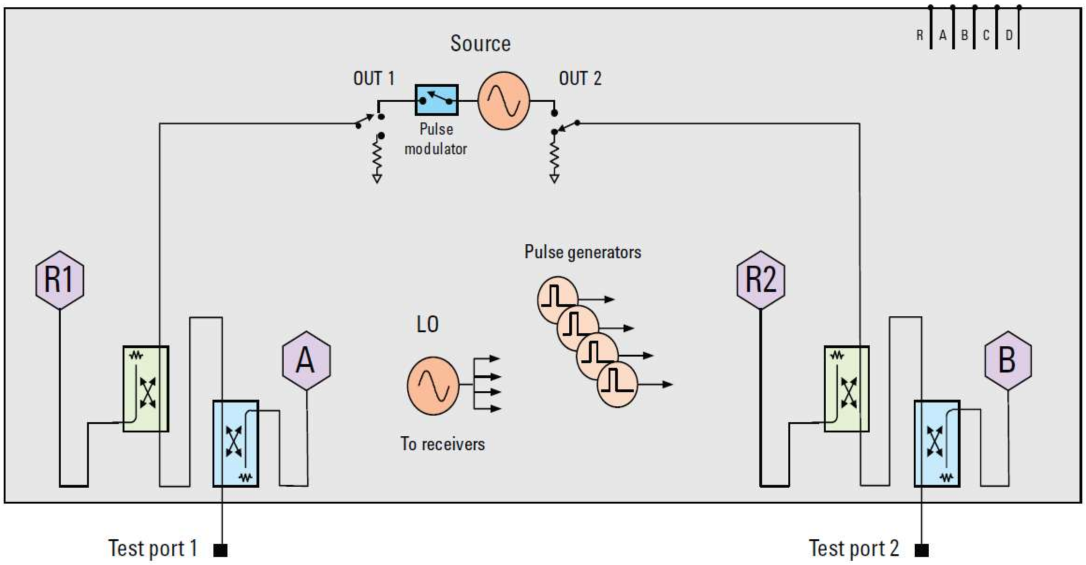
Clearly, some other way of characterizing high-frequency networks is needed that doesn't have these drawbacks. That is why scattering or S-parameters were developed. S-parameters have many advantages over the previously mentioned H, Y or Z-parameters. They relate to familiar measurements such as gain, 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.

Getting to know network analyzers

- Block diagram: how do network analyzers work?

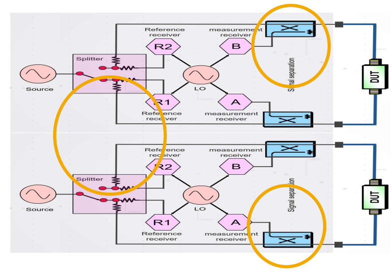
Modern VNA Block Diagram

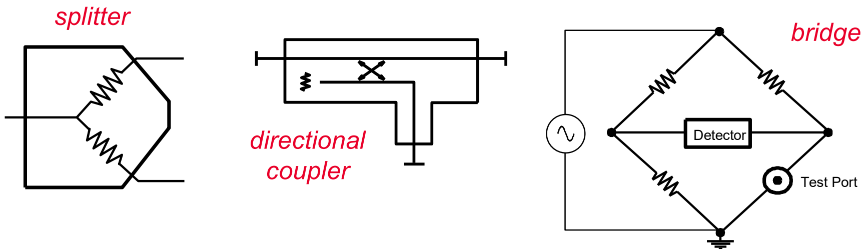


29

Signal separation

- Measure incident signal for reference
- Separate incident and reflected signals





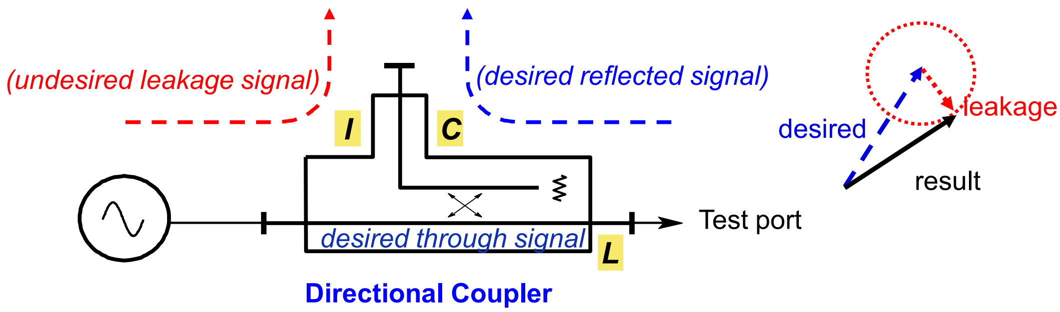
The next major area we will cover is the signal separation block. The hardware used for this function is generally called the "test set". The test set can be a separate box or integrated within the network analyzer. There are two functions that our signal-separation hardware must provide. The first is 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.

Directional coupler & directivity

Directivity is a measure of how well a directional coupler or bridge can separate signals moving in opposite directions

Directivity = Isolation (I) - Fwd Coupling (C) - Main Arm Loss (L)



A directional coupler is a signal separation device. When an input signal is applied in the coupler's forward direction, some amount of that forward power is split or coupled and made available at the coupled arm. This coupled signal power is lower than the input power by the forward coupling factor amount. For example, if using a 10dB coupler and 0dBm input power, the power at the forward coupling arm would be -10dBm.

The power at the straight-through output of the coupler will equal the input power minus any losses in the coupler through arm.

A key specification when using directional couplers in Network Analyzers is directivity. Directivity denotes the difference between the coupler's forward coupling factor and any reverse-coupled leakage at the coupled arm. This indicates the amount of separation between desired coupled signals and undesired leakage signals.

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:

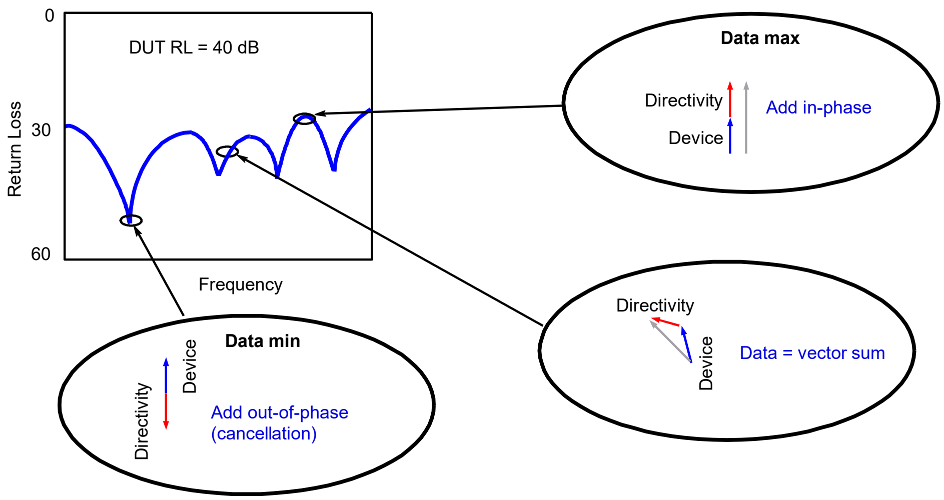
Directivity (dB) = Isolation (dB) - Forward Coupling Factor (dB) - Loss (through-arm) (dB)

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.

Interaction of directivity with the DUT

Without error correction

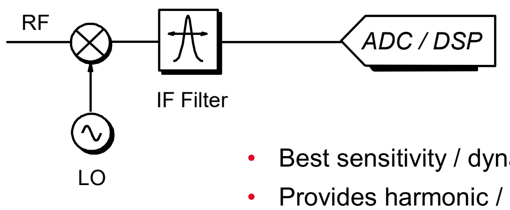


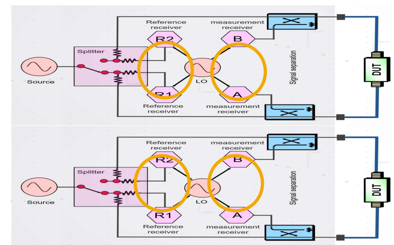
KEYSIGHT

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.

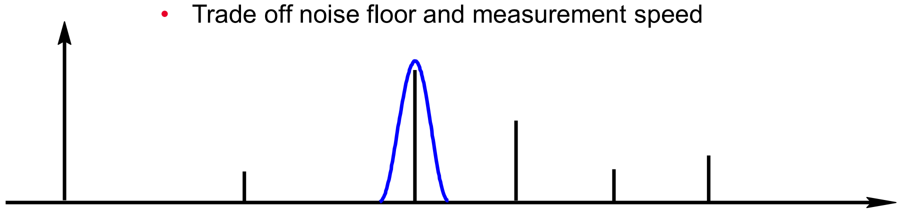
Receivers

Narrowband Detection - Tuned Receiver





- Best sensitivity / dynamic range
- Provides harmonic / spurious signal rejection
- Improve dynamic range by increasing power, decreasing IF bandwidth, or averaging
- Trade off noise floor and measurement speed

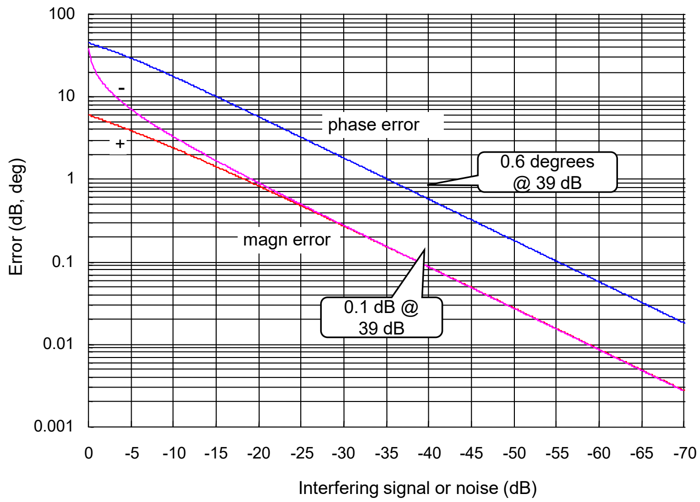


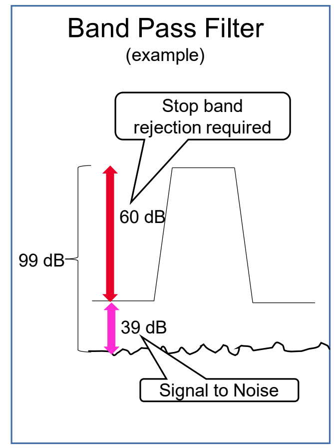
Tuned receivers provide the best sensitivity and dynamic range, and also provide harmonic and spurious-signal rejection. The narrow IF filter produces a considerably lower noise floor, resulting in a significant sensitivity improvement. For example, a microwave vector network analyzer (using a tuned 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.

Dynamic range and accuracy

Error due to interfering signal





Dynamic range for 0.1 dB accuracy = 60 dB (rejection) + 39 dB (SNR) = 99 dB

Dynamic range is very important for measurement accuracy!

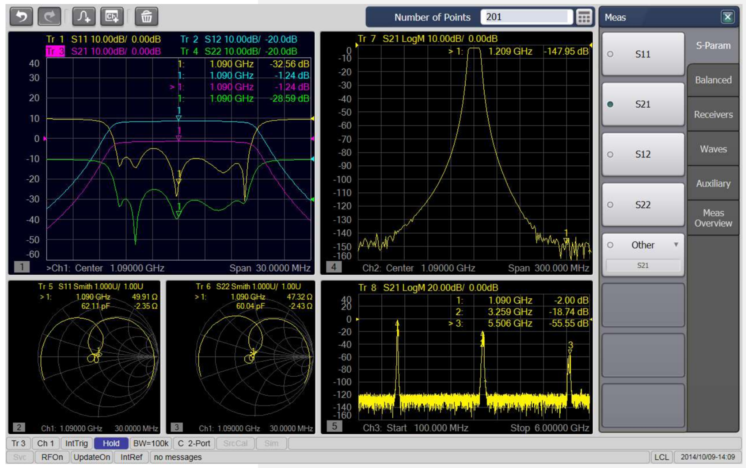
KEYSIGHT

This plot shows the effect that interfering signals (sinusoids or noise) have on measurement accuracy. The magnitude error is calculated as 20*log [1 ± interfering-signal] and the phase error is calculated as arc-tangent [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 60 dB of rejection would require 99 dB of dynamic range. One way to achieve this level is to average test data using a tuned-receiver based network analyzer.

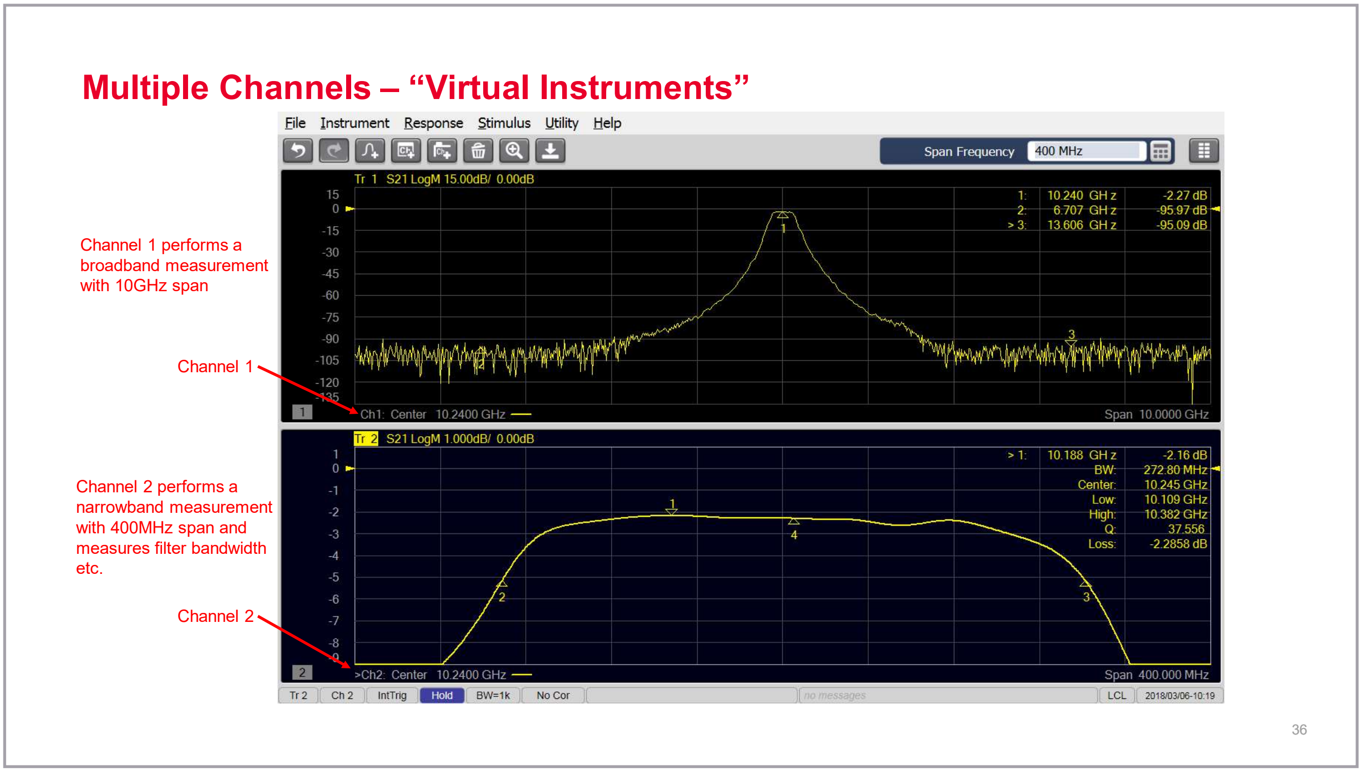
Processor / Display

- Markers
- Limit lines
- Pass/fail indicators
- Linear/log formats
- Grid/polar/Smith charts
- Time-domain transform
- Trace math



KEYSIGHT

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. This is a display example of the E5080A ENA.



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. This is a display example of the E5080A ENA.

Thank you