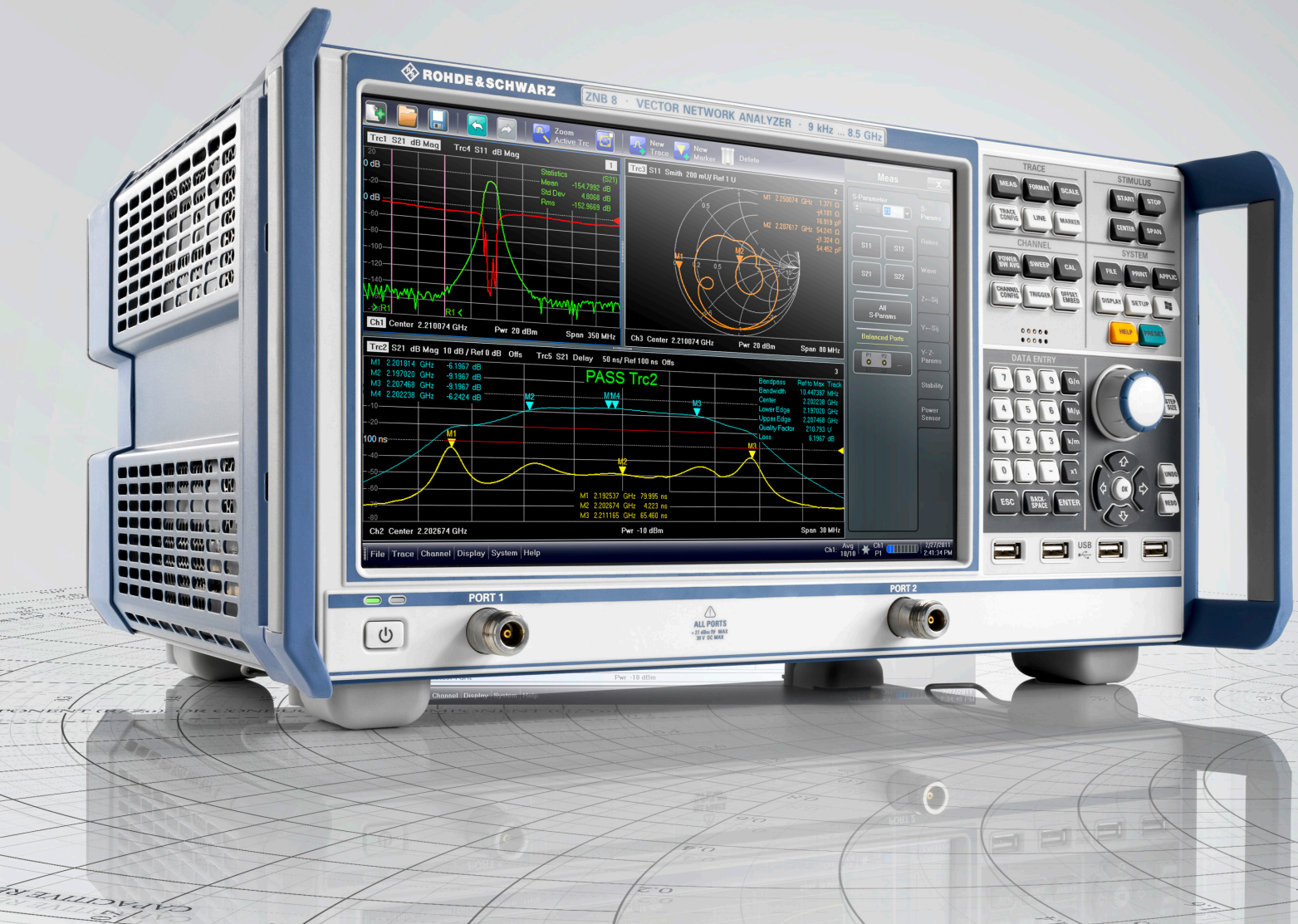
Fundamentals of Vector Network Analysis Primer

ROHDE&SCHWARZ



Fundamentals of Vector Network Analysis
Version 1.1
Published by Rohde & Schwarz USA, Inc.
6821 Benjamin Franklin Drive, Columbia, MD 21046

Contents

Introduction

What is a Network Analyzer?

One of the most common measuring tasks in RF engineering involves analysis of circuits (networks). A network analyzer is an instrument that is designed to handle this job with great precision and efficiency. Circuits that can be analyzed using network analyzers range from simple devices such as filters and amplifiers to complex modules used in communications satellites.

A network analyzer is the most complex and versatile piece of test equipment in the field of RF engineering. It is used in applications in research and development and also for test purposes in production. When combined with one or more antennas, it becomes a radar system. Systems of this type can be used to detect invisible material defects without resorting to X-ray technology. Using data recorded with a network analyzer, imaging techniques were used to produce the following figure which shows a typical material defect. (Figure 1.1.2)

A similar system can be used to verify the radar visibility which forms the basis for a dependable flight control system. For such purpose the radar cross section (RCS) of an aircraft is an important quantity. It is typically measured on a model of the aircraft like the following result. (Figure 1.1.3)

For measurements with less demanding technical requirements such as measurement of a fill level without physical contact or determination of the thickness of layers of varnish, simpler approaches are generally used.

Wave Quantities and S-Parameters

The so-called wave quantities are preferred for use in characterizing RF circuits. We distinguish between the incident wave *a* and the reflected wave *b*. The incident wave propagates from the analyzer to the device under test (DUT). The reflected wave travels in the opposite direction from the DUT back to the analyzer. In the following figures, the incident wave is shown in green and the reflected wave in orange. Fig. 1.2.1 shows a one-port device with its wave quantities.

The true power traveling to the one-port device is given by $|a|^2$ and the true power it reflects by $|b|^2$. The reflection coefficient Γ represents the ratio of the incident wave to the reflected wave.

$$\Gamma = b/a \qquad (1.2\text{-}1)$$

It is generally a complex quantity and can be calculated from the complex impedance Z. With a reference impedance of typically $Z_0 = 50\ \Omega$[1], the normalized impedance

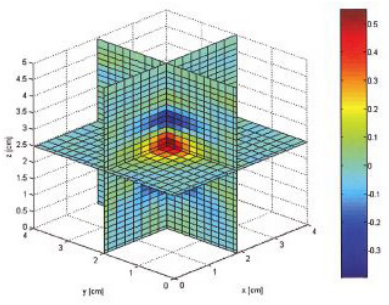


Fig. 1.1.2 Material defect

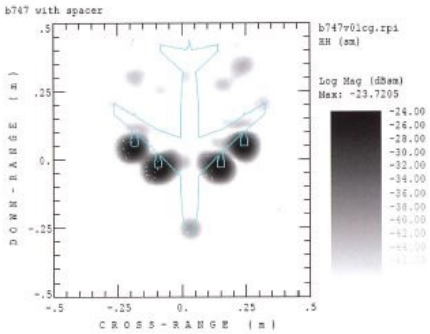


Fig. 1.1.3 ISAR image of a Boeing 747 model

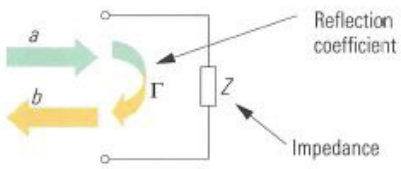


Fig. 1.2.1 One-port device with incident and reflected waves.

1) In RF engineering and RF measurement a reference impedance of 50 Ω is used. In broadcasting systems a reference impedance of 75Ω is preferred. The impedance of 50 Ω offers a compromise which is closely related to coaxial transmission lines. By varying the inner and out conductor diameter of a coaxial transmission line we achieve its minimum attenuation at a characteristic impedance of 77 Ω and its maximum power handling capacity at 30 Ω.

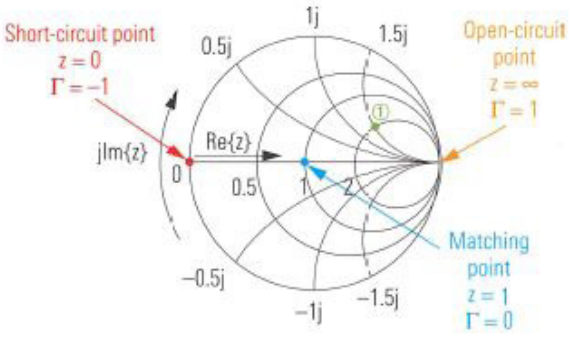


Fig. 1.2.2 Smith chart with sample points.

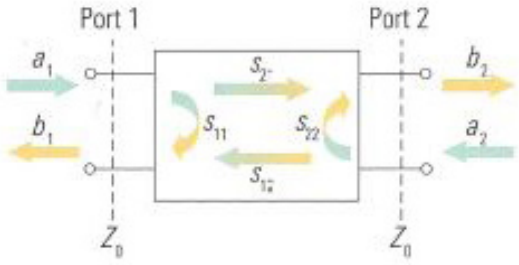


Fig. 1.2.3 Two-port device with its wave quantities.

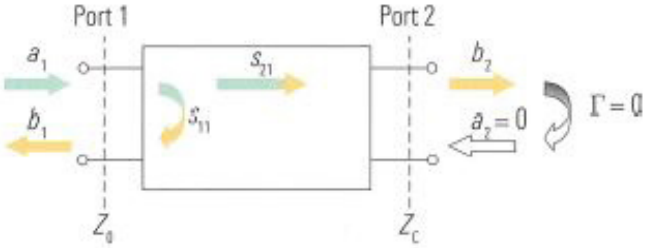


Fig. 1.2.4 Two-port device during forward measurement.

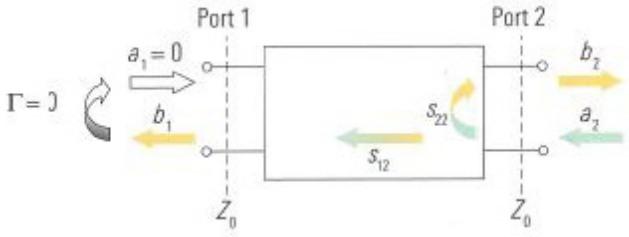


Fig. 1.2.5 Two-port device during reverse measurement.

$z = Z/Z_0$ is defined and used to determine the reflection coefficient.

$\Gamma = z\text{-}1/z+1$ (1.2-2)

The reflection coefficient Γ can be represented in the complex reflection coefficient plane. To draw the normalized impedance z = 2 + 1.5j as point "1" in this plane, we take advantage of the auxiliary coordinate system shown in Fig. 1.2.2 which is known as the Smith chart. The short-circuit point, open-circuit point and matching point are drawn in as examples. (Figure 1.2.2)

In a two-port device, besides the reflection at the two ports, there is also the possibility of transmission in the forward and reverse directions. (Figure 1.2.3)

In comparison to the reflection coefficient, the scattering parameters (S-parameters) s_{11}, s_{12}, s_{21} and s_{22} are defined as the ratios of the respective wave quantities. For the forward measurement, a reflectionfree termination $\Gamma = 0$ (match) is used on port 2. This means that $a_2 = 0$. Port 1 is stimulated by the incident wave $a_1 \neq 0$. (Figure 1.2.4)

Under these operating conditions, we measure the input reflection coefficient s_{11} on port 1 and the forward transmission coefficient s_{21} between port 1 and port 2.

$$s_{11} = \frac{b_1}{a_1}\bigg|_{a_2 = 0}; \; s_{21} = \frac{b_2}{a_1}\bigg|_{a_2 = 0} \qquad (1.2\text{-}3)$$

For the reverse measurement, a match $\Gamma = 0$ is used on port 1 ($a_1 = 0$). Port 2 is stimulated by the incident wave $a_2 \neq 0$. (Figure 1.2.5)

Under these operating conditions, we measure the output reflection coefficient s_{22} on port 2 and the reverse transmission coefficient s_{12} between port 2 and port 1.

$$s_{12} = \frac{b_1}{a_2}\bigg|_{a_1 = 0}; \; s_{22} = \frac{b_2}{a_2}\bigg|_{a_1 = 0} \qquad (1.2\text{-}4)$$

In general both incident waves can be non-zero ($a_1 \neq 0$ and $a_2 \neq 0$). This case can be considered as a superposition of the two measurement situations $a_1 = 0$ and $a_2 \neq 0$ with $a_1 \neq 0$ and $a_2 = O$. This results in the following:

$$b_1 = s_{11}\, a_1 + s_{12} a_2$$
$$b_2 = s_{21}\, a_1 + s_{22} a_2 \qquad (1.2\text{-}5)$$

We can also group together the scattering parameters s_{11}, s_{12}, s_{21} and s_{22} to obtain the S-parameter matrix (S-matrix) and the wave quantities to obtain the vectors *a* and *b*. This results in the following more compact notation:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad (1.2\text{-}6)$$

$$B = Sa \qquad (1.2\text{-}7)$$

Many standard components can be represented as one- or two-port networks. However, as integration increases, DUTs with more than two ports are becoming more commonplace so that the term N-port has been introduced. For example, a three-port network *(N = 3)* is characterized by the following equations:

$$b_1 = s_{11}a_1 + s_{12}a_2 + s_{13}a_3$$
$$b_2 = s_{21}a_1 + s_{22}a_2 + s_{23}a_3 \qquad (1.2\text{-}8)$$
$$b_3 = s_{31}a_1 + s_{32}a_2 + s_{33}a_3$$

The shorter notation (1.2-7) is also valid for a three-port network. In formula (1.2-6), all that is required is expansion of the vectors *a* and *b* to three elements. The associated S-matrix has 3 x 3 elements. The diagonal elements s_{11}, s_{22} and s_{33} correspond to the reflection coefficients for ports 1 to 3 which can be measured in case of reflection-free termination on all ports with $\Gamma = 0$. For the same operating case, the remaining elements characterize the six possible transmissions. The characterizations can be extended in a similar manner for N > 3.

Why Vector Network Analysis?

A network analyzer generates a sinusoidal test signal that is applied to the DUT as a stimulus (e.g. a_1). Considering the DUT to be linear, the analyzer measures the response of the DUT (e.g. b_2) which is also sinusoidal. Fig. 1.3.1 shows an example involving wave quantities a_1 and b_2. They will generally have different values for the amplitude and phase. In this example, the quantity s_{21} represents these differences.

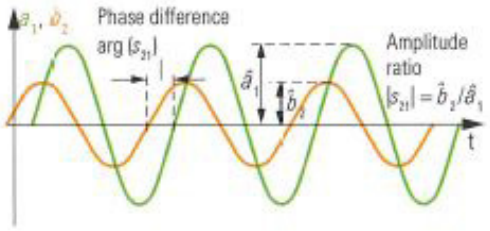


Fig. 1.3.1 Signals a_1 and b_2.

A scalar network analyzer only measures the amplitude difference between the wave quantities. A vector network analyzer (VNA) requires a significantly more complex implementation. It measures the amplitude and phase of the wave quantities and uses these values to calculate a complex S-parameter. The magnitude of the S-parameter (e.g. $|s_{21}|$) corresponds to the amplitude ratio of the wave quantities (e. g. b_2 and $\hat{a}_1$). The phase of the S-parameter (e.g. $\arg(s_{21})$) corresponds to the phase difference between the wave quantities. This primer only considers vector network analysis due to the following benefits it offers:

- Only a vector network analyzer can perform full system error correction. This type of correction compensates the systematic measurement errors of the test instrument with the greatest possible precision.
- Only vectorial measurement data can be unambiguously transformed into the time domain. This opens up many opportunities for interpretation and further processing of the data.

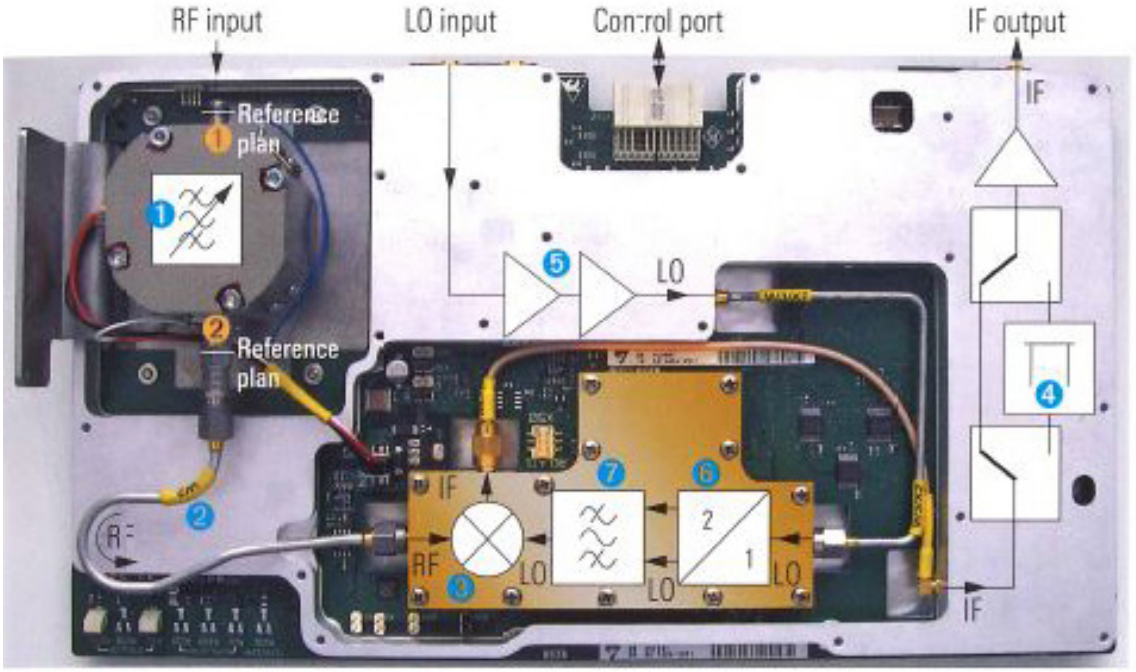


Fig. 1.4.1
A frequency converter module

- Deembedding and embedding are special processing techniques that enable computational compensation of a test fixture or computational embedding into a network that is not physically present. Both of these techniques require vectorial measurement data.

- For presentation in Smith charts, it is necessary to know the reflection coefficient vectorially.

There are two common approaches for building vector network analyzers.

- Network analyzers based on the homodyne principle only have a single oscillator. This oscillator provides the stimulus signal and is also used to process the response. Most analyzers based on this principle are relatively economical. However, due to their various technical limitations, they are suited only for simple applications, e.g. for measuring fill levels based on the radar principle.

- Precise investigation of circuits requires network analyzers that are based on the heterodyne principle. The network analyzer family described in this primer is based on this principle which is discussed in greater detail in section 2.6.

A Circuit Example

Fig. 1.4.1 shows a circuit that is commonly used in RF engineering: a frequency converter. This module converts a frequency f_{RF} in the range from 3 GHz to 7 GHz to a fixed intermediate frequency f_{IF} = 404.4 MHz. To ensure unambiguity of the received frequency, a tunable bandpass filter (1) at the start of the signal processing chain is used. The filtered signal is fed via a semi-rigid cable (2) to a mixer (3) which converts the signal from frequency fRF to frequency ~F' A switchable attenuator (4) is used to set the IF level. The LO signal required by the mixer is prepared using several amplifiers (5), a frequency doubler (6) and a bandpass filter (7). The module itself is controlled via a 48-pin interface and is commonly used in test receivers and spectrum analyzers.

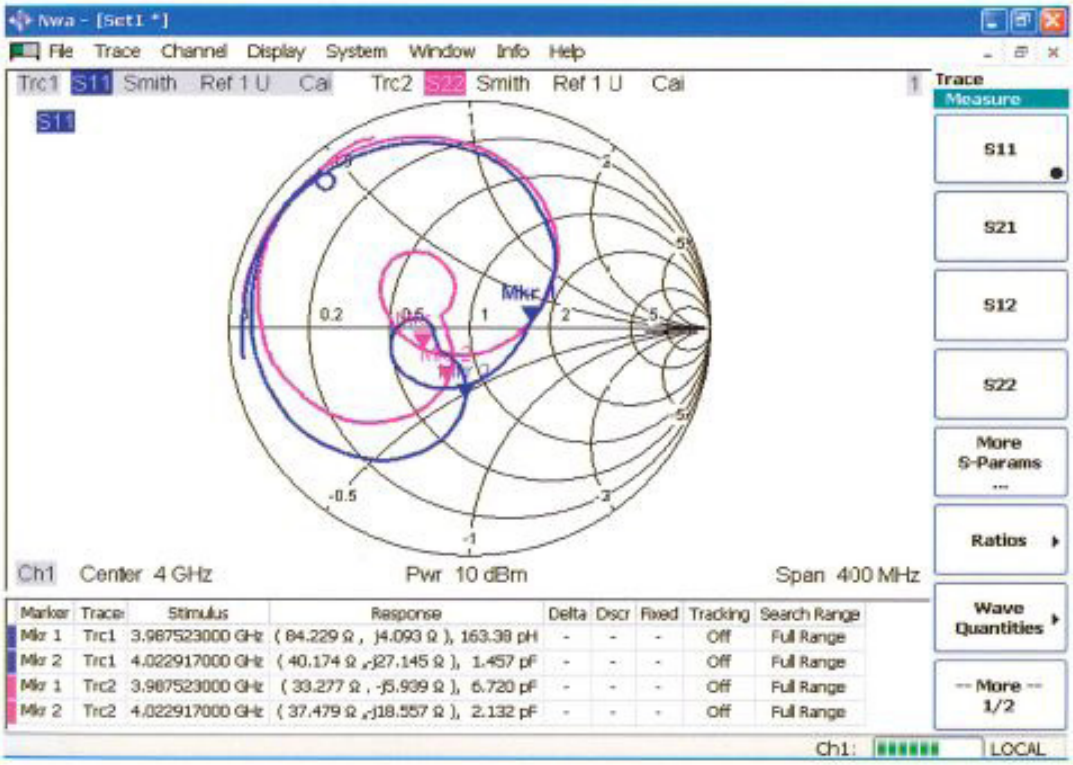


Fig. 1.4.2 Input and output reflection coefficients of the tunable bandpass filter.

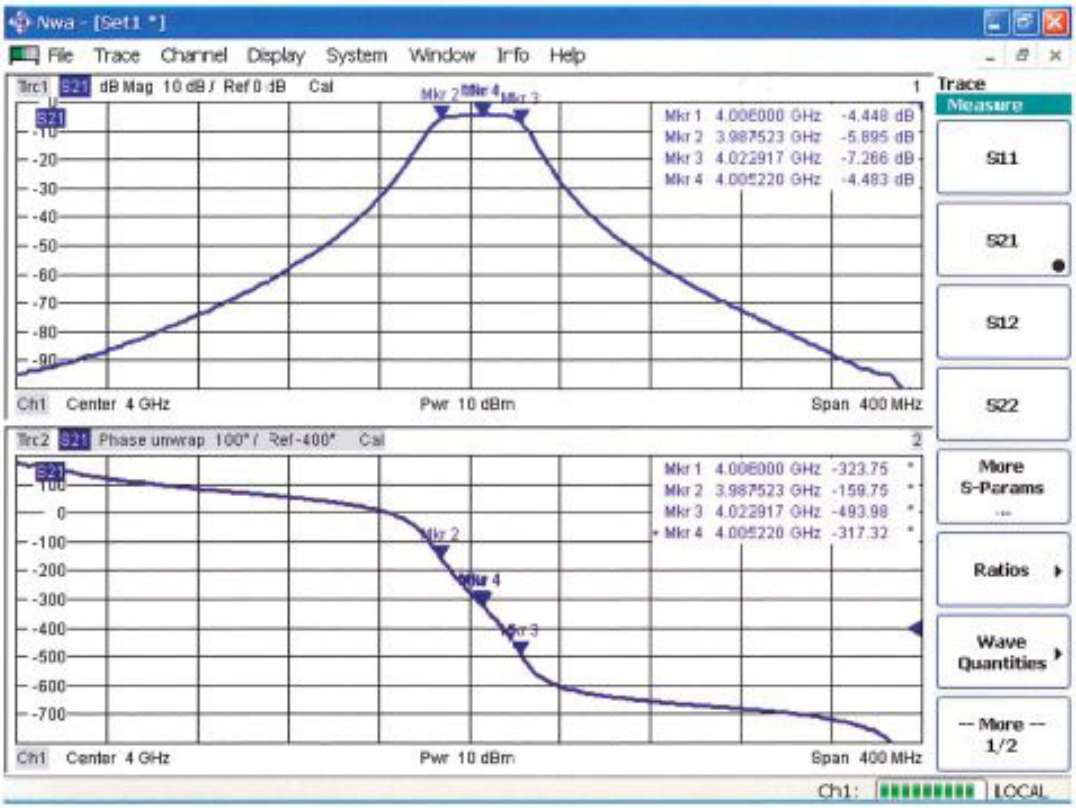


Fig. 1.4.3 Forward transmission and phase of transmission coefficient s_{21} for the bandpass filter.

A network analyzer is useful, for example, for investigating the tunable bandpass filter (1). The test ports of the network analyzer were connected to ports (1) and (2) of the filter. Fig. 1.4.2 shows the input and output reflection coefficients of the filter.

The results were measured in the range from 3.8 GHz to 4.2 GHz: Markers were used to precisely read off selected measurement results. The term bandpass filter is derived from the forward transmission shown in Figure 1.4.3.

Besides these examples, there are many other measurements that we can conceivably make on the module in Fig. 1.4.1. The following table provides a brief summary of some of the possibilities. (Table 1.4.1)

Examples from Fig. 1.4.1 →	Filters ❶	Lines ❷	Mixers ❸	Switches, switchable attenuators ❹	Amplifiers ❺	Frequency multipliers ❻	Balanced filters ❼
Linear measurements							
Reflection coefficient	✓	✓	✓	✓	✓	✓	✓
Transmission coefficient	✓	✓		✓	✓	✓	✓
Group delay	✓	✓	✓	✓	✓		✓
Phase delay	✓	✓		✓	✓		✓
Measurement of balanced DUTs						✓	✓
Stability					✓		
Switching time				✓			
Nonlinear measurements							
Harmonics			✓		✓	✓	
Intercept points			✓		✓	✓	
1 dB compression point			✓		✓	✓	
Hot S-parameters			✓		✓	✓	
Time-domain measurement							
Distance to fault on cables		✓					
Mixer measurements							
LO feedthrough			✓				
Isolation			✓				
Conversion loss			✓				

Table 1.4.1 Usage of different measurements with typical DUTs in Fig 1.4.1.

Design of a Heterodyne N—Port Network Analyzer

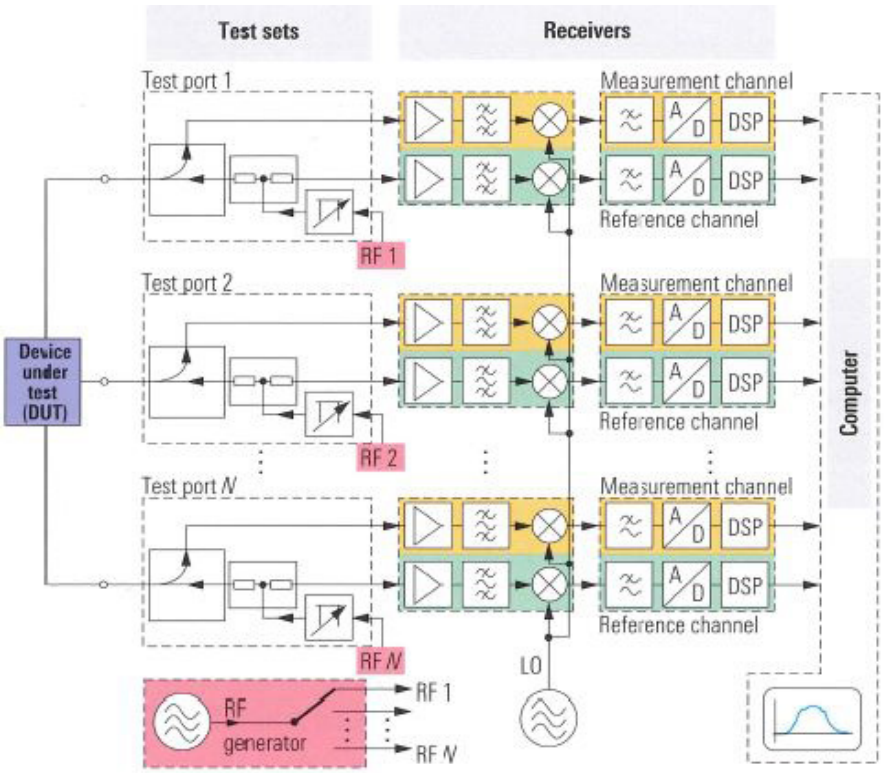


Fig. 2.1.1 Standard block diagram of an N-port vector network analyzer.

The information contained in this primer is based on a heterodyne vector network analyzer. Due to the increasing importance of N-port DUTs, we will assume there are an arbitrary number of N ports.

Block Diagram

The block diagram in Fig. 2.1.1 has four main components:

- The test set separates the incident and reflected waves at the test port. The waves are fed to the reference channel or to the measurement channel. Electronic attenuators are used to vary the test port output power. Any generator step attenuators that might be present extend the lower limit of the output power range.
- The generator provides the RF signal which we refer to as the stimulus. The source switch which is used with the generator passes the stimulus signal to one of the test ports which is then operated as an active test port.
- Each test set is combined with two separate receivers for the measurement channel and the reference channel. They are referred to as the measurement receiver and the reference receiver[1], They consist of an RF signal section (heterodyne principle) and a digital signal processing stage. At the end of the stage, we have raw measurement data in the form of complex numerical values.

1) In place of “measurement channel” the term “test channel” is also common. Furthermore the “measurement receiver” is also called “test receiver”.

- A computer is used to do the system error correction and to display the measurement data. It also provides the user interface and the remote control interfaces. The preinstalled software is known as the firmware.

In the rest of this section, we will examine the individual components starting with the test set and continuing through the instrument and ending at the display of the measured data.

Design of the Test Set

Measuring the reflection coefficient Γ_{DUT} requires separation of the incident and reflected waves traveling to and from the DUT. A directional element is required for this purpose. In the following discussion, it is described as a three-port device.

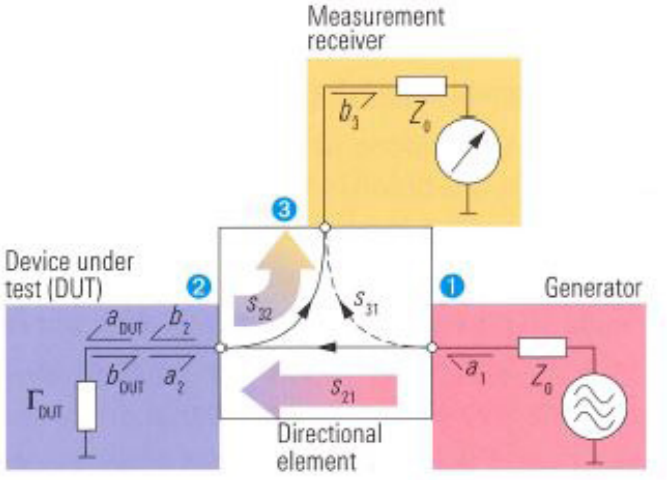


Fig. 2.2.1 Measurement circuit with directional element.

In figure 2.2.1, the two main signal directions of the directional element are shown in color. The wave a_1 produced by the generator is forwarded to port (2) with transmission coefficient s_{21}, where it leaves the element as wave b_2. In the case of a one-port DUT, the wave b_{DUT} arises from the wave a_{DUT} due to reflection with the reflection coefficient Γ_{DUT}.

$$b_{DUT} = \Gamma_{DUT} \; a_{DUT} \qquad (2.2\text{-}1)$$

From the viewpoint of the DUT, the wave b_2 corresponds to the incident wave a_{DUT} and the wave a_2 corresponds to the reflected wave b_{DUT}. Formula (2.2-1) can thus be expressed using quantities a_2 and b_2:

$$a_2 = \Gamma_{DUT} \; b_2 \qquad (2.2\text{-}2)$$

Finally wave a_2 reaches port (3) with the coupling coefficient s_{32}. At this port the measurement receiver is located. Ideally, S-parameters s_{21} and s_{32} would both have a value of 1. The signal path that leads directly from port (1) to port (3) is undesired. Accordingly, we would like to obtain the best possible isolation in which s_{31}=0. Reflection at port (2) back to the DUT also has an unwanted effect. In the ideal case, we would like this reflection to disappear. Moreover, if we assume that generator wave a_1 is constant, then wave quantity b_3 is directly proportional to reflection coefficient Γ_{DUT} of the DUT. In the real world, however, the ideal assumptions made above are not valid. We will remove them one-by-one in the following subsections.

Constancy of the A Wave

In practice, it is possible to maintain the generator wave a) at an approximately constant level, e.g. within a limit of a 0 dBm ± 0.3 dB. The remaining inaccuracy of the a) wave would directly affect the measurement result. To prevent this, we can determine the value of the a_1 wave using an additional receiver which we refer to as the reference receiver. To generate a reference channel signal a power splitter can be used (see next figure). Both output branches of the power splitter are symmetrical. It can be shown that these branches are directly coupled to each other. The signals exiting as waves a_1 and a'_1 are always the same, regardless to the

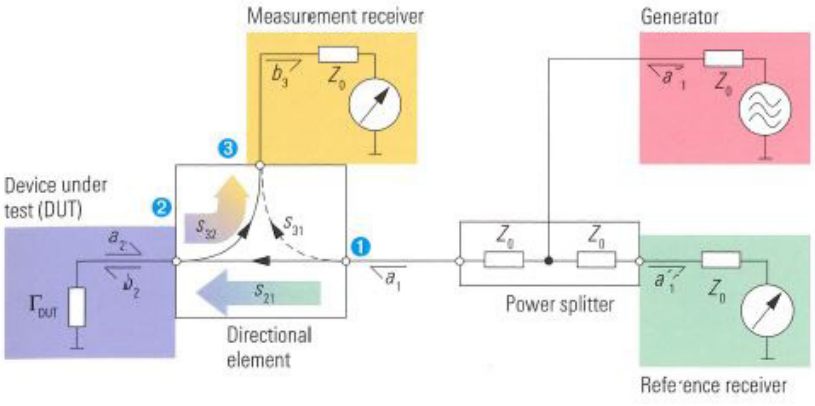


Fig. 2.2.2 Directional element with reference channel.

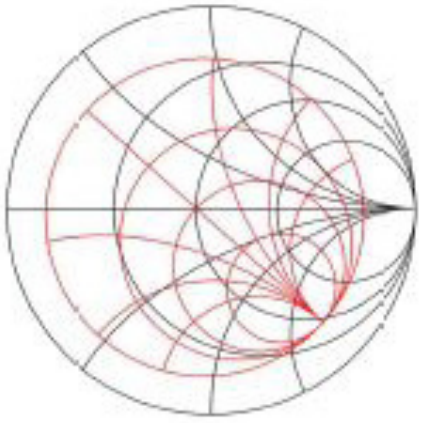


Fig. 2.2.3 Ideal and distorted Smith chart for formula (2.2-5).

possible mismatch on the output ports of the power splitter.[1] If the DUT is connected to one branch, via a directional element, the wave quantity a'_1 can be used instead of quantity a_1.[2] (Figure 2.2.2)

The reflection coefficient Γ_{DUT} is measured by the following ratio, which we refer to as the measured value M:

$$M = b_3/a'_1 \qquad (2.2\text{-}3)$$

Reflection Tracking

In the real world, transmission coefficient s_{21} and coupling coefficient s_{32} have values less than 1. They are multiplied to obtain the reflection tracking *R*.

$$R = s_{32}s_{21} \qquad (2.2\text{-}4)$$

For quantities M and R, we thus obtain the following equation:

$$M = R \cdot \Gamma_{DUT} \qquad (2.2\text{-}5)$$

To illustrate the effects of R, several significant points of the Smith chart have been inserted into formula (2.2-5) as value Γ_{DUT}. As a result the Smith chart in Fig. 2.2.3 is transformed into the red diagram which is compressed by magnitude IRI and rotated by angle arg(R).

Directivity

Crosstalk from port (1) to port (3) bypasses the measurement functionality. S-parameter S_{31} characterizes this crosstalk. To compare it to the desired behavior of the directional element, we introduce the ratio known as the directivity:

$$D = s_{31}/R \qquad (2.2\text{-}6)$$

The directivity vectorially adds on the quantity Γ_{DUT}. Accordingly, we must modify formula (2.2-5) as follows:

$$M = R\,(\Gamma_{DUT} + D) \qquad (2.2\text{-}7)$$

To assess the measurement uncertainty, we factor out the quantity Γ_{DUT} in the formula above and form the complex ratio $x = D/\Gamma_{DUT}$. Furthermore the product $R \cdot \Gamma_{DUT}$ is denoted as W.

$$M = R \cdot \Gamma_{DUT}\,(1 + x) = W(1 + x) \qquad (2.2\text{-}9)$$

1) This characteristic of the power splitter is essential for its function. Unlike the power splitter a power divider formed of three Zo/3 resistors is unsuitable for this use.

2) Along with the drawings of this primer a consistent background color scheme is used (e.g. the DUT's background color is always a light purple). You can take Fig. 2.2.2 as a representative example.

The expression (1 + x) characterizes the relative deviation of the measured quantity M, from value W. For a general discussion this relative deviation (1 + x) is shown in Fig. 2.2.4 as superposition of the vectors 1 and x. Any assessment of the phase of the complex quantity x requires vectorial system error correction. Without this correction, we must assume an arbitrary phase value for x, resulting in the dashed circle shown in Fig. 2.2.4. Any more accurate calculation of (1 + x) is not possible.

Based on the points in the circle, the two extreme values produced by addition of 1-|x| and 1 + |x| can be extracted. They are shown in Fig. 2.2.4 in blue and red, respectively. They designate the shortest and longest sum vectors, respectively. In RF test engineering, decibel values (dB) are commonly used in reference to magnitudes. The two extreme values can also be represented on a decibel scale.

$$20\,\lg(1 - |x|)\ \text{dB} \quad \text{and} \quad 20\lg(1 + |x|)\ \text{dB} \qquad (2.2\text{-}10)$$

A third point of interest in Fig. 2.2.4 is where the phase deviation produced by x reaches its maximum value:

$$\Delta\psi_{max} = \arcsin(x) \qquad (2.2\text{-}11)$$

The following table provides an evaluation of formulas (2.2-10) and (2.2-11) for various magnitudes of x. The quantity of each column in the table can be found in Fig. 2.2.4. All the values in the table are scaled in decibels (dB). To demonstrate the usage of this table, an example is provided as follows:

We assume a directivity D of -40 dB and a value W of -30 dB in equation (2.2-9), as well as an ideal reflection tracking resulting in W= Γ_{DUT}. To use table 2.2.1 we calculate the normalized value |x| = |D|/| Γ_{DUT} | meaning -40 dB - (-30 dB) = -10 dB in dB scale. According to the table we read of deviations |x+1| = 2.39 dB and |x-1| = -3.30 dB. Together with the value W = -30 dB we calculate the magnitude limits of the measured value M: Upper limit M = W + 2.39 dB = -27.61 dB and lower limit M = W - 3.30 dB = -33.3 dB.

For the limit case where x = 1, i.e. the case in which the directivity D and the reflection coefficient Γ_{DUT} to be measured are exactly equal, the measured value M is between $b_3/a_1 = 0$ and $b_3/a_1 = 2R^* D$ corresponding to values of -∞ and 6.02 dB. Accordingly, it is not possible to directly measure reflection coefficients less than D. On the other hand, when measuring medium and large reflection coefficients,

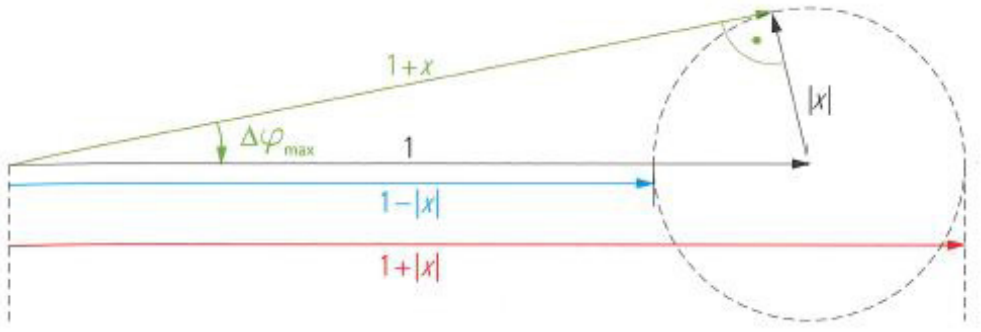


Fig. 2.2.4 Vectorial superposition of 1 and |x|.

$\lvert x\rvert$	$1+\lvert x\rvert$	$1-\lvert x\rvert$	$\Delta\varphi_{max}$
0 dB	6.02 dB	$-\infty$	90.00°
–1 dB	5.53 dB	–19.27 dB	63.03°
–2 dB	5.08 dB	–13.74 dB	52.59°
–3 dB	4.65 dB	–10.69 dB	45.07°
–4 dB	4.25 dB	–8.66 dB	39.12°
–5 dB	3.88 dB	–7.18 dB	34.22°
–6 dB	3.53 dB	–6.04 dB	30.08°
–7 dB	3.21 dB	–5.14 dB	26.53°
–8 dB	2.91 dB	–4.41 dB	23.46°
–9 dB	2.64 dB	–3.81 dB	20.78°
–10 dB	2.39 dB	–3.30 dB	18.43°
–11 dB	2.16 dB	–2.88 dB	16.37°
–12 dB	1.95 dB	–2.51 dB	14.55°
–13 dB	1.75 dB	–2.20 dB	12.94°
–14 dB	1.58 dB	–1.93 dB	11.51°
–15 dB	1.42 dB	–1.70 dB	10.24°
–16 dB	1.28 dB	–1.50 dB	9.12°
–17 dB	1.15 dB	–1.32 dB	8.12°
–18 dB	1.03 dB	–1.17 dB	7.23°
–19 dB	0.92 dB	–1.03 dB	6.44°
–20 dB	0.83 dB	–0.92 dB	5.74°
–21 dB	0.74 dB	–0.81 dB	5.11°
–22 dB	0.66 dB	–0.72 dB	4.56°
–23 dB	0.59 dB	–0.64 dB	4.06°
–24 dB	0.53 dB	–0.57 dB	3.62°
–25 dB	0.48 dB	–0.50 dB	3.22°
–30 dB	0.27 dB	–0.28 dB	1.81°
–35 dB	0.15 dB	–0.16 dB	1.02°
–40 dB	0.09 dB	–0.09 dB	0.57°
–45 dB	0.05 dB	–0.05 dB	0.32°
–50 dB	0.03 dB	–0.03 dB	0.18°

Table 2.2.1 Estimate of the measurement uncertainty for superposition of vectorial quantities.

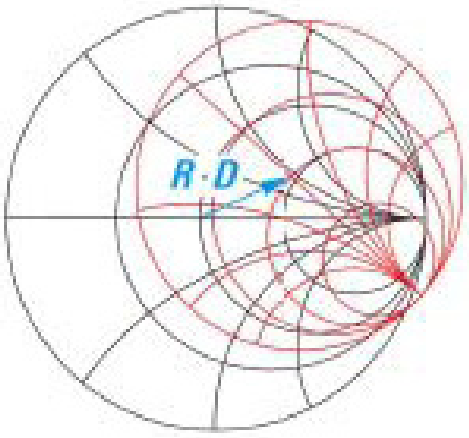


Fig. 2.2.5 Ideal and distorted Smith chart for formula (2.2-7).

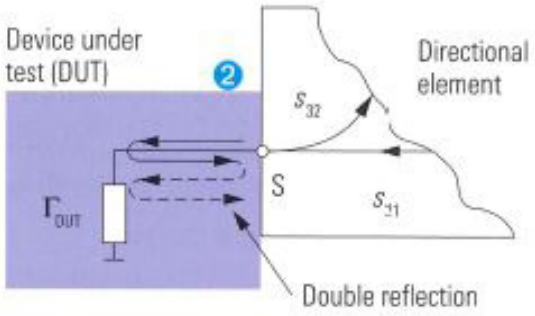


Fig. 2.2.6 Multiple reflection at the test port.

the influence of the directivity is negligible.

Like the reflection tracking, the directivity can directly be shown in the Smith chart. Recalling formula (2.2-7), addition of D to the value Γ_{DUT} corresponds to a shift of the red chart shown in Fig. 2.2.3 by the vector R*D.

Test Port Match and Multiple Reflections

Besides the DUT, it is also possible to assign a reflection coefficient to the test port. The term we use for this is test port match S. In practice, we have to assume a test port match S ≠ O. As a simplification, we first assume that the remaining components in the network analyzer are ideal. The test port match is then determined solely by the directional element, i.e. its scattering parameter s_{22}.

The wave b^{DUT} (= a_2) that has been reflected by the DUT is not fully absorbed by the test port (port (2)). This means that part of this wave is reflected back to the DUT. Between the test port and the DUT, multiple reflections occur. They are shown in Fig. 2.2.6 as a snaking arrow. Let's analyze this phenomenon in more detail. From the first reflection at the DUT we obtain the contribution b_2 Γ_{DUT} to a_2. Part of this wave is reflected by the test port

with the reflection coefficient S. It travels again to the DUT where it makes a contribution $b_2 \Gamma_{DUT} S \Gamma_{DUT}$ to the a_2 wave. After this double reflection, we can generally stop our consideration of this phenomenon. We now add up the contributions as follows:

$$a_2 = b_2 \Gamma_{DUT} + b_2 \Gamma_{DUT} S \Gamma_{DUT} \qquad (2.2\text{-}12)$$

After factoring out $b_2 \Gamma_{DUT}$ we obtain a formula with a structure that is similar to formula (2.2-9):

$$a_2 = b_2 \Gamma_{DUT} (1 + \Gamma_{DUT} S) \qquad (2.2\text{-}13)$$

We can examine the measurement uncertainty introduced by the test port match in a similar manner to formulas (2.2-10) to (2.2-11) or Table 2.2.1 using (1 + x) as (1 + Γ_{DUT} S), If we want to take into consideration reflections that go beyond the double reflection, we can use the following formula. It holds assuming IS· Γ_{DUT} I < 1, which is generally the case since S «1.

$$a = (\Gamma_{DUT}/1\text{-}S\times\Gamma_{DUT}) \bullet b_2 \qquad (2.2\text{-}14)$$

Summary

The non-ideal properties of the test setup that we have considered so far are related primarily to the directional element.

Its influence on the measured value $M = b_3/a_1$ can be combined to the following formula:

$$M = R(D+\Gamma_{DUT}/1\text{-}S\times\Gamma_{DUT}) \qquad (2.2\text{-}15)$$

If we only take into account double reflections between the DUT and the test port (which is generally sufficient), then we can simplify this formula as follows:

$$M = R(D+\Gamma_{DUT}(1+S\times\Gamma_{DUT})) \qquad (2.2\text{-}16)$$

The reflection tracking R results in a relative measurement error, that is independent by the value Γ_{DUT}. It can be corrected easily using a constant (complex) factor. The directivity D and the test port match S are causing a measurement error, which depends on the reflection coefficient Γ_{DUT} or in other words depends on the measured value M. A complex system error correction has to be used to compensate for this error. If we do not carry out this correction, the two quantities D and S will influence the measurement uncertainty. Here, the influence to the measurement quantity M is a function of the value I Γ_{DUT} I. The following relationships hold:

- The directivity limits the measurement accuracy for small values of the reflection coefficients magnitude I Γ_{DUT} I
- The test port match determines the limit for large magnitudes I Γ_{DUT} I ·

If we assume ideal conditions apart from S and D, we can deduce the resulting measurement uncertainty from Fig. 2.2.8.

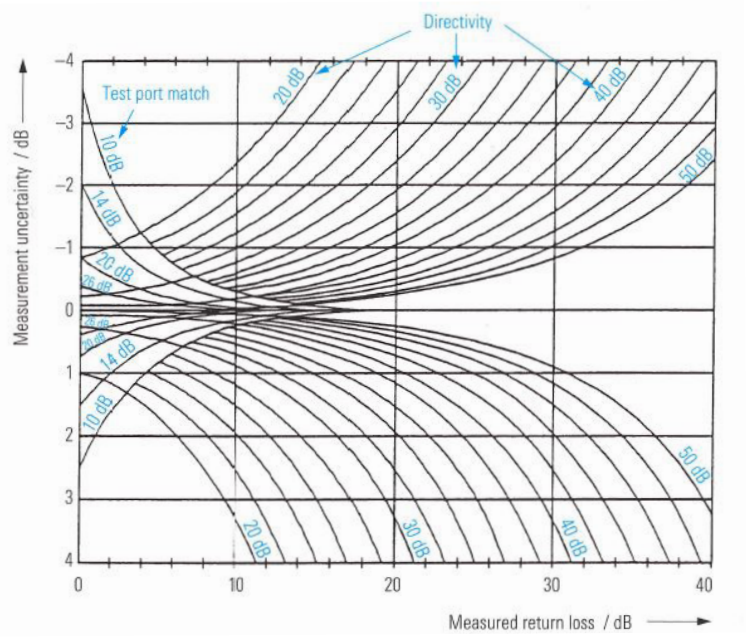


Fig. 2.2.8 Measurement uncertainty as a function of the directivity, test port match and measured return loss.

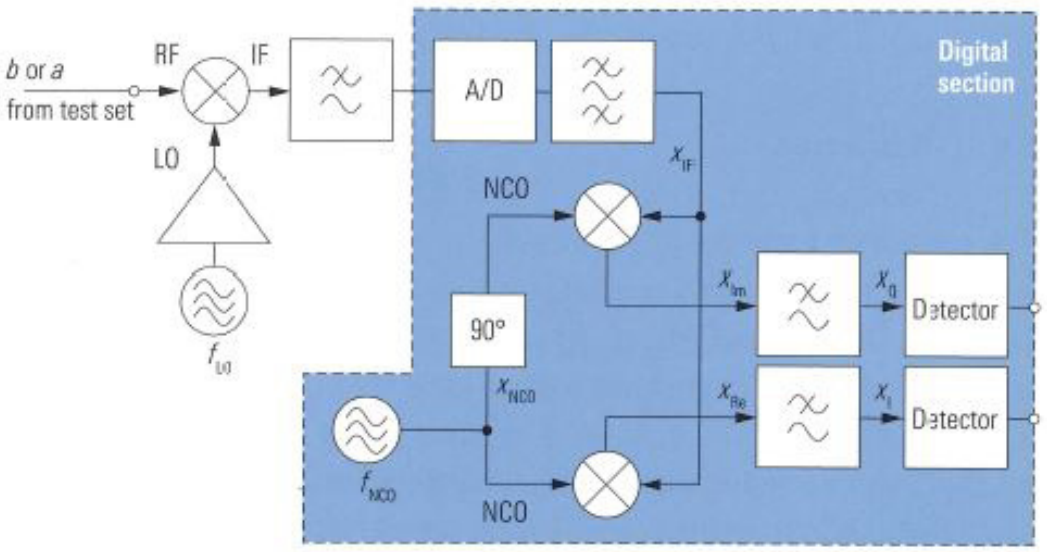


Fig. 2.4.1 Transmission measurement using a heterodyne receiver.

Generator

The generator produces the sinusoidal stimulus signal. At its heart is an electronically tunable oscillator with a relatively wide tuning range. To ensure the required frequency stability and spectral purity, this oscillator is embedded in a phase locked loop (PLL). Fixed frequency reference oscillators have very good spectral purity and long-term stability. Here, we can distinguish between temperature-compensated crystal oscillators (TCXO) and oven controlled temperature crystal oscillators (OCXO). The latter type of oscillator is generally available as an instrument option. A PLL links the frequency and phase of the tunable oscillator to the reference oscillator. This helps to improve the spectral purity of the tunable oscillator in the vicinity of the operating frequency. Using suitable techniques in the PLL, it is possible to modify the frequency ratio between the reference oscillator and the tunable oscillator in order to generate the frequencies needed for the stimulus signal. Implementation of the required tuning range with the tunable oscillator requires several switchable voltage-controlled oscillators (VCOs). As an alternative, a Yttrium-Iron-Garnet (YIG) oscillator can be used which has a very wide tuning range, e.g. 2 GHz to 20 GHz. Since YIG oscillators are tuned using an external magnetic field, they exhibit hysteresis effects which make fast, precise tuning more difficult. They also tend to be relatively unwieldy due to their massive field coils. Low frequencies are usually generated using frequency division or mixing.

Reference and Measurement Receiver

The heterodyne principle (Greek: hetero = different) involves having a local oscillator frequency f_{LO} that is different from the received frequency f_{RF}. This means that the measured signal is converted to an intermediate frequency $f_{IF} = |f_{RF} - f_{LO}|$ (see left side of Fig. 2.4.1). The magnitude and phase information in the measured signal is retained. Initial filtering is used at the intermediate frequency stage which helps to keep a large share of the received broadband noise out of the following signal processing chain. It also serves as an anti-aliasing filter for the analog/digital converter.

The analog/digital converters used in modern network analyzers generally have a resolution of at least 14 bits. Using additional techniques like dithering, it is possible to further increase their effective resolution. The analog mixer must generally be seen as the component limiting the dynamic range. If the level is too high, it produces nonlinear distortions. However, the linear range of the mixer is also not suitable for arbitrarily low-amplitude signals since the noise prevents measurement of very low-amplitude signals. Using a switchable amplifier, we can specifically target the levels in the analog section of

the receivers and optimize them for the current RF input level. The required amplification is determined during a fast preliminary measurement. This technique is known as automatic gain control (AGC).

By choosing a suitable value for the local oscillator frequency f_{LO}, we can convert any RF frequency within in the receiver's range to a fixed intermediate frequency. This simplifies the subsequent IF processing which is handled digitally in modern instruments. A possible implementation of the digital IF processing stages can be seen on the right side of Fig. 2.4.1. To achieve even better selectivity, additional filtering is performed as part of the digital signal processing (DSP). A numerically controlled oscillator (NCO) generates a sinusoidal signal which is used to mix the IF signal down to the frequency f = 0. Two digital multipliers are used in this procedure known as I/Q demodulation. One of the multipliers is operated with an NCO signal that is phase-shifted by 90°. We assume that the signals $X_{IF}(t)$ and $x_{NCO}(t)$ are represented by the following formulas:

$$x_{IF}(t) = A_{IF} \times \cos(2\pi f_{IF} t + \varphi_{IF}) \qquad (2.4\text{-}1)$$

$$x_{NCO}(t) = A_{NCO} \times \cos(2\pi f_{NCO} t) \qquad (2.4\text{-}2)$$

With $f_{NCO}=f_{IF}$ and the trigonometric identities[1] we can calculate the signals $x_{Re}(t)$ and $x_{Im}(t)$:

$$x_{Re} = 1/2 A_{IF} A_{NCO} [\cos(\varphi_{IF}) + \cos(4\pi f_{NCO} t + \varphi_{IF})] \qquad (2.4\text{-}3)$$

$$x_{Im} = 1/2 A_{IF} A_{NCO} [\sin(\varphi_{IF}) - \sin(4\pi f_{NCO} t + \varphi_{IF})] \qquad (2.6\text{-}4)$$

The lowpass filtering suppresses the frequency components where f ≠0. This produces the DC signals x_1,x_Q corresponding to the real and imaginary parts of the complex phasor $x_{IF}(f)$. In the field of communications engineering, they are known as the inphase and quadrature components.

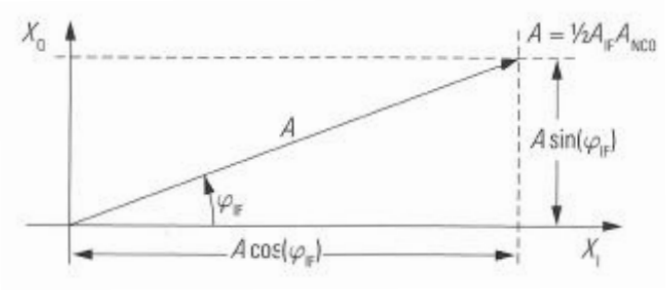


Fig. 2.4.2 Illustration of the quadrature and inphase components.

Using the heterodyne technique, a phase shift occurs when mixing down the RF signal that is dependent on the phase of the LO and NCO signals. Since the same LO and NCO signals are used for all reference and measurement receivers, this phase shift cancels out during computation of the S-parameters. However, a fixed frequency reference between the generator, LO and NCO signals is required which can be achieved using one of the following synchronization techniques. For the sake of simplicity, a transmission measurement s_{21} without any test set is assumed for the following discussion.

In the implementation shown in Fig.2.4.3a, the LO and RF oscillators are linked via a phase locked loop (PLL) to a common crystal-stabilized frequency reference. This technique allows an arbitrary frequency offset between the generator and receiver frequency, which is limited only by the instrument's frequency range. Using this technique, the generator and receiver frequency can also be swept using different step sizes and in opposite directions if required. However, this implementation increases costs, due to the need for additional hardware.

Another possibility is shown in Fig. 2.4.3b. There, the local oscillator is connected to the generator via a PLL. The IF frequency is used as the control variable. Since it is necessary to measure the a1 wave at the respective

1) $\cos\alpha\cos\beta = 1/2[\cos(\sigma-\beta) + \cos(\alpha+\beta)]$
$\cos\alpha\sin-\beta = 1/2[\sin(\sigma-\beta) + \cos(\alpha+\beta)]$

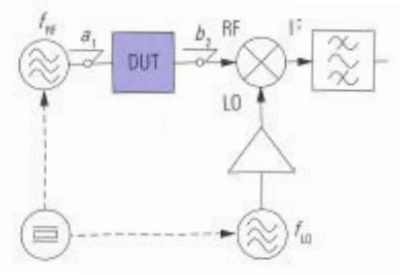


Fig. 2.4.3a Common reference.

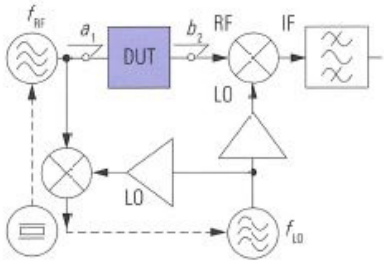


Fig. 2.4.3b Locking of the LO to the RF.

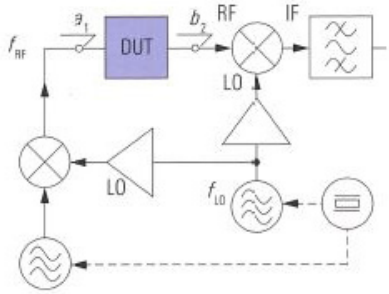


Fig.2.4.3c Generation of the stimulus signal from the LO.

active port anyway in order to measure the S-parameters, it can be used to lock the local oscillator's frequency to the RF. In measurements that involve frequency conversion, we set the receiver and the generator to different frequencies so that it is no longer possible to directly receive the generator signal and the control variable is no longer available. In most cases, an external reference mixer is then required to generate a suitable control variable.

In Fig. 2.4.3c, the generator signal is produced from the LO signal using an auxiliary oscillator. To ensure that the frequencies fRF and fLO are stable, the auxiliary oscillator must be synchronized to the common reference oscillator. Compared to the implementation in Fig.2.4.3b, a frequency offset is possible, but it is normally limited to a maximum offset of approx. 10 MHz to 100 MHz.

With equipment operating at higher frequencies, it is generally very expensive to provide the local oscillator. By using harmonic mixing, a simpler design is enabled. Instead of the usual LO frequency $f_{LO} = f_{RF} \pm f_{IF}$, all that the mixer requires is, e.g. the frequency $f_{LO}=1/3(f_{RF} \pm f_{IF})$ so that the local oscillator no longer has to cover the upper stimulus frequency range. Of course, the conversion loss of harmonic mixers is higher which means we must pay for this benefit with a loss of dynamic range.

Measurement Accuracy and Calibration

Any measurement result is subject to a measurement uncertainty that characterizes the expected statistical deviation of the measured values from their true value.

We can distinguish two types of measurement uncertainties:

- Measurement uncertainties of type A [IS93], which are caused by random measurement errors. It is possible to statistically describe such errors, but they cannot be systematically corrected.

- Measurement uncertainties of type B [IS93], which are caused by systematic measurement errors. These errors occur in a reproducible manner and can be systematically corrected using computational techniques. However full correction is impossible, due to superimposed random fluctuations in the measurement result.

To correct systematic measurement errors as fully as possible, the errors and the measured quantity must be known vectorially. Since scalar network analyzers only record the magnitude of the measured quantities, systematic measurement errors cannot be corrected by these type of analyzers. As a consequence these errors make a sizable contribution to the total measurement uncertainty in this case. It is possible to determine this contribution using the maximum and minimum values 1 + |x| and 1-|x| as described in Fig. 2.2.4. Even if a scalar network analyzer is basically adequate for measuring the magnitude of the S-parameters, a vector network analyzer will generally provide significantly better measurement accuracy when used for the same measurement task (assuming suitable correction is performed). The remainder of chapter 3 will not consider scalar network analyzers since they are outside the scope of this primer.

Vector network analyzers usually offer several techniques for correcting systematic measurement errors. Assuming we have selected a suitable technique for correcting systematic measurement errors, reduction of random measurement errors becomes a key task when making extremely precise measurements. Modern network analyzers are designed to minimize random influence factors. Of course this cannot replace the proper use of the relevant equipment and accessories. The most important things to look out for in this area are described in section 3.1.

Reduction of Random Measurement Errors

Thermal Drift

A warmup phase should be observed even in equipment with good thermal stability to ensure that the equipment is operated in thermal equilibrium. The warmup time for analyzers are normally specified in the relevant data sheet along with the warmup time (if any) required for calibration equipment. Once the equipment has warmed up, an environment with a stable temperature helps to keep temperature fluctuations small. When making measurements, avoid unnecessarily touching the device under test (DUT).

The repeatability describes the correlation between successive measurements done over a short period of time under the same conditions (same measured quantity, same instrument, same instrument settings, same measurement procedure, same DUT ...). Achieving high repeatability requires use of suitable connectors and test port cables. Fig. 3.1.1 shows what can be expected under favorable conditions using the example of a PC3.5 connection. Here, a one-port DUT (short standard with a PC3.5 female connector) was connected directly to the network analyzer for a total of nine times and the measured value for s_{11} was recorded every time. Using trace mathematics, the measurement results were normalized to the saved measured values for the first contact. Fig. 3.1.1 shows the relative error for the individual measurements in dB. Using a short standard for this measurement provides an insight into stability of contact impedance and parasitic reflections

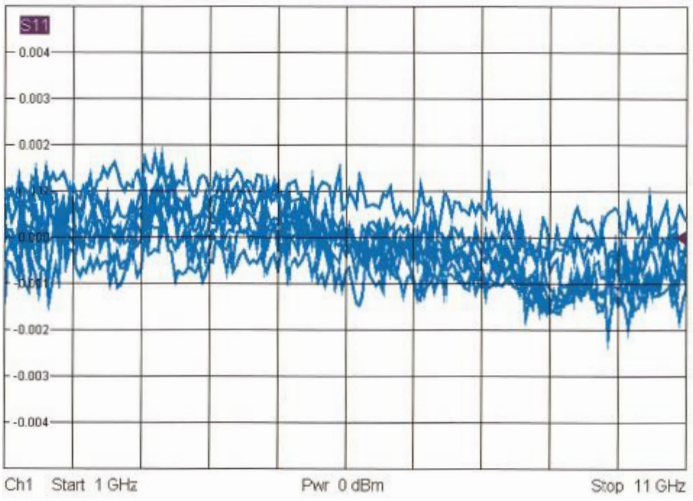


Fig. 3.1.1 Repeatability for a PC3.5 connection during a reflection measurement.

caused by the connectors. Displaying the phase instead of the magnitude would reveal the repeatability of the connection's electrical length.

Besides the connection assemblies, the test port cables that are used have a significant influence on the repeatability. Here, the quality can vary significantly. We can get an initial impression of the phase and amplitude stability of a set of test port cables by connecting both cables to a network analyzer and connecting their open ends using a through standard. The measured quantities are the phase and amplitude of the transmission coefficient s_{21}. We save the result of the first measurement and use it for normalization using trace mathematics Data/Mem. Now, we move the cables and observe the change in the displayed results. Fig. 3.1.2a shows the measurement results for two consumer quality cables each with a length of 1 m (made of cable material RG400). For comparison,

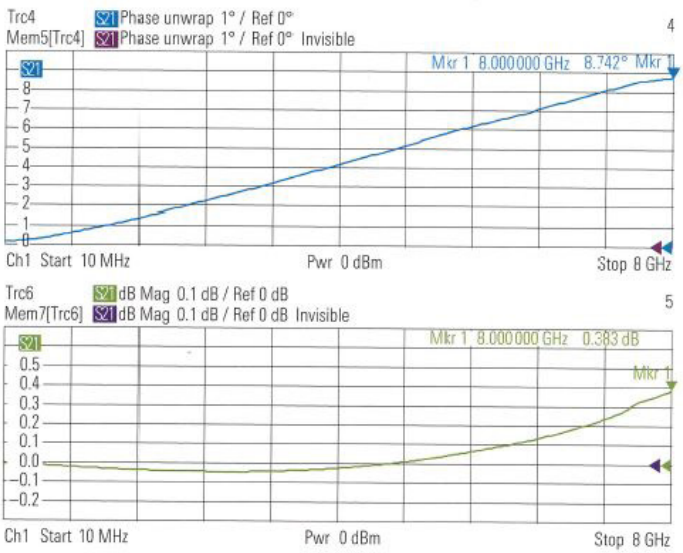


Fig. 3.1.2a Deviation due to bending on two RG400 cables with a length of 1 m each.

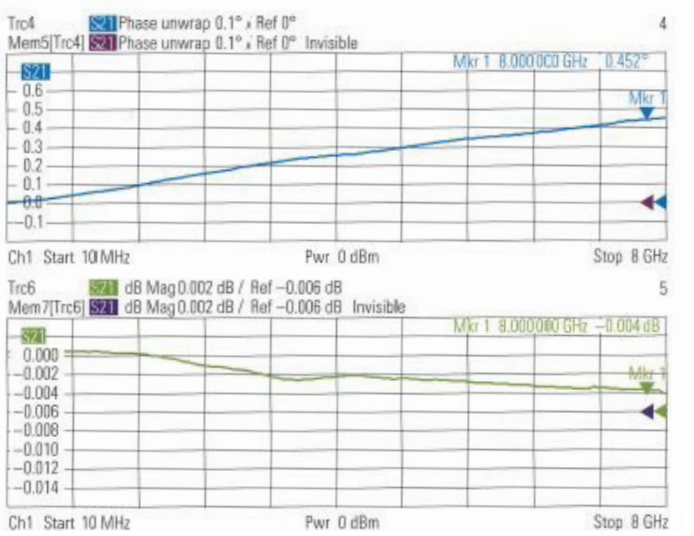


Fig.3.1.2b Deviation due to bending on a high-performance test port cable set with a length of 2 x 1 m.

Fig. 3.1.2b illustrates the behavior of a set of high-performance test port cables with a comparable length. In both cases, the same bending radius and bending angle was chosen. Note the scale difference with a factor of 10 (phase) and 100 (amplitude) between the displayed results!

Besides the quality of the equipment, proper handling and care of connection assemblies is important:

- A connector's interfaces should be kept clean and free of dirt. Clean connectors are essential for good RF performance. Never use water, acids or abrasives. Using a cotton swab moistened with isopropyl alcohol, you can do a relatively good job of cleaning the connectors. However, do not saturate the swab with alcohol; verify that no cotton remains in the connector after cleaning. As alternatives, pure low-pressure compressed air or nitrogen can be used for cleaning. Be sure to heed the applicable safety guidelines for using and storing the materials listed above (e.g. protective goggles, labels on bottles, fire hazards)!

- Tighten the connecting nut using a torque wrench. During the tightening process, only the nut should be tightened. Do not rotate the connector. Rotating the connector will cause unnecessary stress to the inner and outer contacts of the connectors, which can lead to excessive wear. Do not over torque the connectors; it may damage them.

- Using a suitable pin depth gage, it is important to regularly check the offset of the inner conductor with respect to the reference plane. This measurement is important particularly prior to initial usage of newly acquired test instruments, cables and accessories to avoid damage to existing equipment.

Noise

The thermal noise which is superimposed on the measured values involves another random measurement uncertainty. Fig. 2.8.5 illustrates this in a qualitative way. It shows the relationship between the IF bandwidth and the noise level. The following discussion will provide a quantitative description of the same:

At room temperature (290 K), the noise power density of thermal noise is equal to $4x10^{-21}$ W/Hz, which corresponds to a noise density level of -174 dBm (Hz). If we were to use an ideal rectangular filter with a 1 Hz bandwidth as our IF filter and our analyzer would not produce any internal noise, a noise level of -174 dBm would be superimposed on the measured signal. In practice, we must also consider the internal noise of the test instrument. The displayed noise level is increased by the noise figure (NF) of the instrument, which is typically specified in dB. By keeping the step attenuator set to as little attenuation as possible, the noise figure can be minimized. In special applications, further improvements can be achieved by feeding in the measurement signals via the direct receiver inputs and by using a preamplifier. The IF bandwidth B_{IF} is generally greater than 1 Hz. The IF filter that we use does not have a rectangular transmission characteristic. Due to this different transmission characteristic, the noise bandwidth of the IF filter is always somewhat greater than the 3 dB bandwidth that is used to characterize the filter setting. The ratio of the two bandwidths yields the shape factor $S_F \geq 1$. The IF filters that can be selected in a network analyzer generally have an approximately Gaussian transmission characteristic with a shape that is identical for all of the filters and which is scaled only in terms of the bandwidth. Accordingly, in the following computation, we can use a constant shape factor. From the noise figure (NF) in dB, the shape factor $S_F \geq 1$ and the IF bandwidth B_{IF} in Hz, we can calculate the noise level L_N as follows:

$$L_N = -174\text{dBm} + NF + 10\lg(S_F)\text{dB} + 10\lg[B_{IF}/\text{Hz}]\text{dB} \quad (3.1\text{-}1)$$

Fig. 3.1.3 shows the noise level for IF bandwidths of 10 Hz, 1 kHz and 100 kHz. As expected, the noise level increases by 20 dB if we increase the IF bandwidth by a factor of 100.

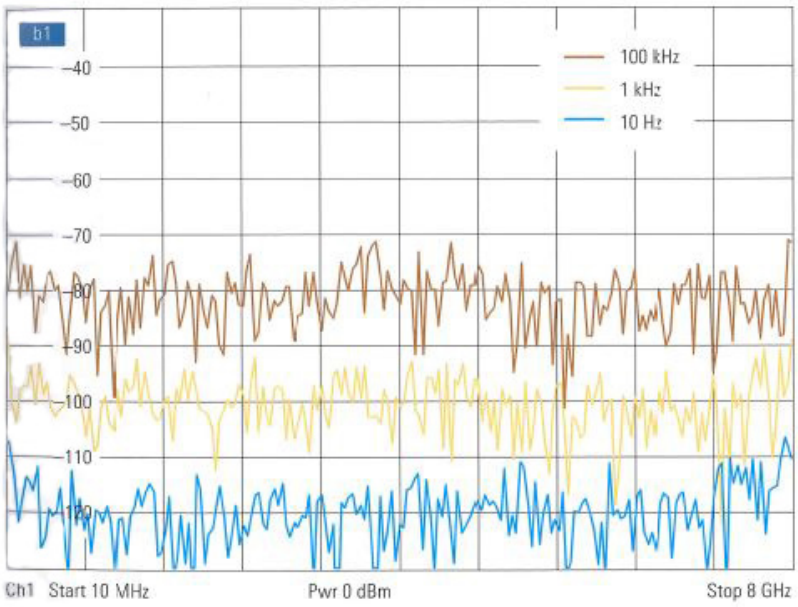


Fig.3.1.3 Noise level for different IF bandwidths.

Unfortunately, any change in the IF bandwidth also influences the sweep time. For small IF bandwidths, these two quantities are inversely proportional. In other words, doubling the IF bandwidth cuts the sweep time in half. For larger IF bandwidths, other settling constants (e.g. for the level and phase lock loops) are predominant. In the measurements discussed here, for example, the required sweep times were as follows: 10 s at 10 Hz, 0.1 s at 1 kHz and 4.5 ms at 100 kHz.

Correction of Systematic Measurement Errors

In principle, we can distinguish between two different types of systematic measurement errors: nonlinear errors and linear errors.

Nonlinear Influences

If we operate the measurement or reference receiver in the vicinity of its upper power limit, compression effects will occur. This is due to the mixers used in RF signal processing. If the compression effects occurred to the same extent in the measurement and reference channel, they would mutually compensate one another when computing the S-parameters. In practice, however, the amplitudes of the measurement and the reference channel are usually very different. The necessary level range of the reference receiver is known a priori. It is dictated by the adjustment range of the test port output power. When it comes to designing a network analyzer the linear operating range of the reference receiver can be optimized to these needs. The linear operating range of the measurement receiver can only be optimized for typical DUTs. In case of active DUTs with a high output power, the receiver can be driven into compression. Typical measurement uncertainties that occur as a result of this compression can be seen in the left part of Fig. 3.2.1.

Since the signal-to-noise ratio decreases at low signal levels, the linear range cannot be exploited for arbitrarily low-amplitude signals (Fig. 3.2.1). Accordingly, for precision measurements, it is necessary to select a level that

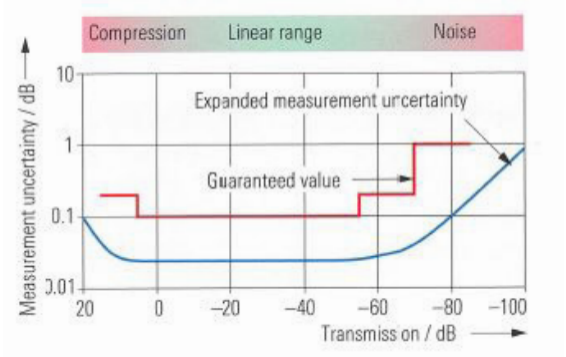


Fig. 3.2.1 Bath-tub curve

circumvents the compression effects described above while still ensuring a good signal-to-noise ratio. For reflection measurements and transmission measurements on passive components, a test port output level of -10 dBm generally represents a good compromise. When working with high gain DUTs, it may be necessary to further reduce the source power.

Linear Influences

As seen in section 2.2.6, a one-port network analyzer can be separated into an error network[1] and an ideal network analyzer. The same approach can also be extended for use with an N-port network analyzer. (Figure 3.2.2)

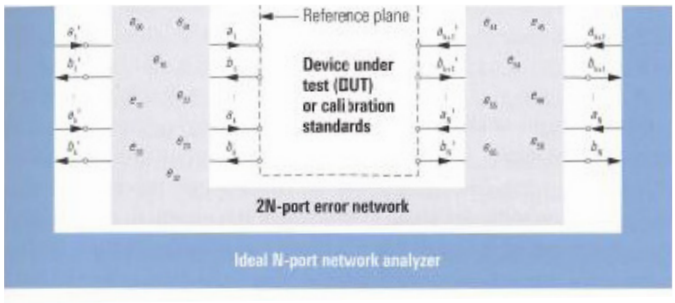


Fig. 3.2.2 Separation between an error network and an ideal network analyzer.

The parameters of the error network can be denoted as error terms[2] e_{ik}. Most error terms can also be directly interpreted as raw system data. System error correction involves mathematical compensation of the error network. The systematic measurement errors that remain after correction are expressed by the effective system data. They are dependent on the accuracy of the error terms e_{ik}. The stability of the system error correction is limited by random measurement errors caused by temperature drift, noise and so on. The table 3.2.1 provides a comparison between typical raw and effective system data for a network analyzer.

System data	Raw system data	Effective system data
Reflection tracking	≤2 dB	≤0.04 dB
Directivity	≥29 dB	≥46 dB
Source match	≥22 dB	≥39 dB
Transmission tracking	≤2 dB	≤0,06 dB
Isolation	≥130 dB	≥130 dB
Load match	≥22 dB	≥44 dB

Table 3.2.1 Comparison of typical raw and effective system data.

A procedure known as calibration is used to determine the error terms eik. Using the test setup (network analyzer with test port cables and possibly a test fixture), measurements are made sequentially using several calibration standards. These are one- and two-port networks with known properties. The calibration technique determines which of the properties of the standards must be known. Since it is impossible to manufacture ideal calibration standards (e.g, an ideal short where Γ = -1), the inherent deviations of the standards are provided to the network analyzer in the form of characteristic data. Once the calibration procedure is completed, the analyzer computes the error terms e_{ik}. For this the analyzer uses the values it measured during the calibration process and the characteristic data belonging to the standards. In the processing chain in Fig. 2.7.2 the steps described in this paragraph appear in green. Using the error terms e_{ik} it is possible to correct raw measured values during subsequent measurements and calculate the S-parameters for the DUT. However, corrected display of individual wave quantities is not possible using this technique since it does not correct the absolute magnitude and absolute phase. When displaying measured values as ratios of wave quantities, e.g. b_2/a_1, b_2/b_1 and

1) In the technical literature, the term "linear error model" can also be found.

2) Also known as correction data in the technical literature.

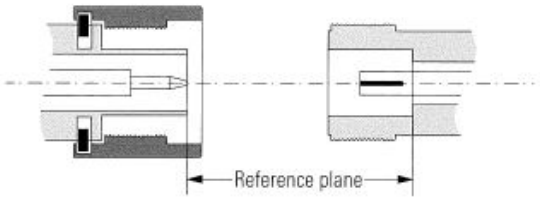


Fig. 3.2.3 Location of the reference plane in the N-type connector.

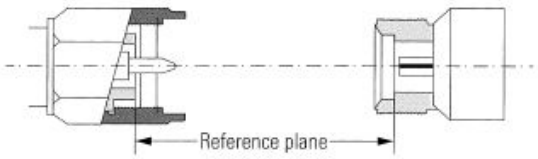


Fig. 3.2.4 Location of the reference plane in the connector types pe3.S, 2.4 mm and 1.8S mm.

so on, most test instruments do not perform any system error correction.

The physical interface between the error network and the DUT is known as the reference plane. The corrected measured S-parameters are referred to this plane as a matter of principle. When using coaxial calibration standards, the reference plane is given by the mating plane of the outer conductor (see Fig. 3.2.3 and Fig. 3.2.4).

Calibration Standards

The calibration process requires special one and two-port devices. Due to inherent manufacturing constraints, their properties will necessarily diverge from ideal standards (ideal open where $\Gamma = 1$, ideal short where $\Gamma = -1$ and so on). For this reason, the actual properties are gathered in the form of the characteristic data. The process of measuring these values is known as characterization. It must be performed in accordance with generally accepted principles so that the characteristic data are traceable to the primary standards of the National Metrological Institutes such as the PTB (Physikalisch-Technische Bundesanstalt) in Germany, the NPL (National Physical Laboratory) in Great Britain and the NIST (National Institute of Standards and Technology) in the United States. It is important to have the characterization verified at regular intervals by an accredited measurement laboratory. The characteristic data are usually included with a calibration kit. They are provided in digital format (e.g. diskette, memory stick or magnetic tape) and as a measurement report.

Description of the standards using special coefficients is a proven approach which will be considered in greater detail in the following sections. The primary benefit of this description is that it is very compact. Even across a wide frequency range from, say, DC to 40 GHz, we only need a

Fig. 3.3.1
3.5 mm Calibration Kit

maximum of seven coefficients per standard. In addition, it is now common to describe the standards also using complex S-parameters. They can be saved using the Touchstone" file format, for example. This eliminates the need to extract the coefficients. A loss of accuracy which can occur due to extraction can be avoided. Of course, the S-parameter description involves a large amount of data which must be provided on a digital storage medium of some sort. (Fig. 3.3.1)

To make the calibration process as fast and straightforward as possible, most manufacturers of network analyzers also offer automatic calibration equipment. For the user, this eliminates the time-consuming and error-prone process of switching manually between different calibration standards. Automation is particularly advantageous in production areas. Since automatic calibration equipment has the characteristic data saved internally, there is no need to transfer the data using a separate storage medium. This eliminates the risk of confusing different storage media in this process.

Practical Hints for Calibration

As a basic rule, a suitable calibration kit should be available for each connector type that is in use. Since the calibration standards are subject to wear their properties should be verified regularly. Alternatively their common condition can be checked when verifying the measurement uncertainty of the calibrated network analyzer. This form of verification is necessary anyway if a measurement uncertainty has to be guaranteed together with the results. Moreover, calibration should be repeated at regular intervals. The actual time interval will depend on the required measurement accuracy, the temperature stability of the environment, the quality of the cables that are used and the overall test setup. Note that modifying the channel settings that are the physical conditions underlying the measurement (e.g. start frequency, number of measurement points, ...) can make the calibration invalid or can involve an interpolation of the calibration data. By nature any interpolation leads to a reduced accuracy. In contrast to this a change in trace settings will not influence the accuracy of the calibration. The following table

Fig. 3.3.2 Calibration unit with four ports

Action \ Influence	No influence	Interpolation	Loss of validity
Modification of the test port output level	X[1]		
Modification of the IF bandwidth	X[1]		
Loosening and retightening of the connections	X[2]		
Modification of the trace format	X		
Change in the measured quantity to S, Y, X parameters or μ_1, μ_2, k coefficients	X		
Modification of the scaling of the axes	X		
Reduction in the frequency range to be displayed or increase in the number of measurement points		X	
Switchover to logarithmic sweep		X	
Exchange or extension of test port cables			X
Increase in the frequency range to be displayed			X
Change to wave quantities or wave ratios or usage of frequency conversion capabilities			X
Switching the step attenuator	X[3]		X[3]

Table 3.4.1 Influence of various actions on the calibration.

Calibration technique	OSM	TOM	TRM	TRL	TNA	UOSM	TOSM	TOM-X
Error model	3-term	7-term	7-term	7-term	7-term	7-term	12-term	15-term
Suitable for transmission measurements		✓	✓	✓	✓	✓	✓	✓
No band limitation due to singularities	✓	✓	✓		✓	✓	✓	✓
Implicit error verification		✓						
Partially unknown standards			✓	✓	✓	✓		
Consideration of DUT-dependent crosstalk								✓
Usage of standards with a different gender		✓	✓[1]	✓[1]	✓[1]	✓	✓	✓
Suitable for non-insertable DUTs						✓		
Possible usage of a sliding match	✓	✓	✓			✓	✓	
Well-suited for on-wafer measurements			✓	✓	✓	✓		✓
Effective directivity attained	+	+	+	+++	+	+	+	+
Number of receivers in N-port network analyzer	$N+1$	$2N$	$2N$	$2N$	$2N$	$2N$	$N+1$	$2N$
Minimum number of calibration standards	3	3	3	3	3	4	4	5
Contacts[2] in two-port analyzer	3	6	6	6	6	8	8	10

Table 3.4.2 Summary of properties for different error models and calibration techniques.

should help you to properly assess the relevant influence factors.

Notes for Table 3.4.1:

1) Assuming that the network analyzer and OUT show a linear behavior.
2) Depending on the repeatability of the connectors.
3) Switching a receiver step attenuator always results in a loss of validity. In the case of a generator step attenuator, we must distinguish between the two implementations.

The five error models (5-term, 7-term, 10-term, 12-term and 15-term model) and the calibration techniques that are based on them are summarized in the following table. The name of each technique is derived from the calibration standards that it uses.

Notes for Table 3.4.2:

1) Assuming the standards produce symmetrical reflections.
2) The number of "contacts" is used to assess the amount of work involved in the calibration procedure. By contact, we mean setting up an electrical connection. For example, mounting a one-port standard requires one contact. Mounting a two-port standard requires two contacts.

Linear Measurements

This section covers some typical measurements that fall under the category of linear measurements. The DUTs that are used are normally part of the standard equipment in a laboratory or at a test shop, meaning the measurements can be handled without any additional expense. The results described here were produced using network analyzers currently available from Rohde & Schwarz. The information contained here is based on the user interface provided by these instruments, but it can be adapted easily to other network analyzers.

Performing a TOM Calibration

To make precision measurements, the vector network analyzer must be calibrated first. The S-parameters are then referred to a defined location ("reference plane"). It can be moved using the corresponding operating functions provided by the analyzer. Once the calibration process is completed, the DUT is connected.

Calibration is a procedure that must be repeated regularly, making a specific work sequence highly advantageous. For most newcomers, calibration looks like a barrier to qualified measurements. To help them overcome this barrier, sections 4.1 and 4.2 are arranged in the form of a step by step procedure.

Test Setup:

- Vector network analyzer
- Two test port cables with PC3.5 connectors (male)
- Calibration kit PC3.5 system
- Torque wrench

Procedure:

1. Connect the test port cables to the analyzer.

2. To make highly precise measurements, please note the network analyzer's warm-up time (e.g., given in the data sheet).

3. First, plan the channel settings such as the start and stop frequency, number of points, sweep type, test port output level and measurement bandwidth. An uncalibrated trial measurement made during the warm-up phase can help in properly selecting the parameters.

4. Make sure that the calibration kit contains the proper connector type. You should use the same connector type that is used by the DUT. However, there are certain exceptions: In case of a DUT with SMA connectors, use a PC3.5 calibration kit since the SMA-type connector is not suitable for building highly precise calibration kits. It has the same reference planes as the PC3.5-type connector. Make sure that the calibration kit is suitable for the measured frequency range.

Measurement Tip:

If cables are necessary for the test setup, they should be installed on the network analyzer prior to the calibration process. As a result, their influence will be taken into account in the calibration process and automatically compensated during system error correction. The cables should have the best possible phase and attenuation stability and should not be unnecessarily long. We recommend using test port cables that are specified for the analyzer.

5. The calibration kit includes a storage medium containing characteristic data for the calibration standards. Import the data into the analyzer.

6. Begin a two-port TOM calibration. Select the connector type for the test ports (in this case, PC3.5(m)).

7. If you have installed the relevant characteristic data during a prior calibration process, it is sufficient to just select the proper data set. In this case note that different calibration kits with the same connector type will be installed on the network analyzer. Make sure you select the characteristic data you previously installed for the connector type PC3.5 (m).

8. Now, connect the calibration standards one after another to the test ports and perform a calibration measurement for each connected standard. To avoid corrupting the measurement, do not move the test cable and standard during this measurement. The necessary sweep time is dependent on the number of points and the selected measurement bandwidth.

9. Once you have measured all of the relevant standards, the network analyzer can determine the correction data using the apply button. The instrument indicates that it is calibrated once the computations are completed e.g. by means of a "Cal" label displayed in the diagram area.

Measurement Tip:

The term test port refers to the connectors of the test port cables or adapters where calibration is performed and which are later connected to the DUT.1f we are not using any test port cables or adapters, then the ports of the analyzer serve as our test ports. The gender is female (f), male (m) or sexless and refers to the inner conductor. If a connector is configured as a jack (socket), it is female (f).

Measurement Tip:

Usage of a calibration unit simplifies the calibration process and speeds it up considerably. Once the calibration unit has been connected and the technique selected, a single keystroke is sufficient to start the calibration. The transfer of the characteristic data and execution of the calibration procedure are both automated.

Measurement Tip:

Verification of the calibration by measurement of standards previously involved in the calibration process is not possible. Calibration kits normally contain through, open and match (T, 0, M) standards as well as a short (S) standard. Since the latter is not used in the TOM calibration, it can be used as a verification standard. An ideal short standard would result in a frequency independent point z = 0 + 0j in the Smith chart. The short standard in the calibration kit is not ideal. The most significant nonideality is its electrical length that is not equal to 0, e.g. 5 mm. This results in a frequency dependence for the phase like the one measured in Fig. 4.1.1 with a value of e.g. 96° at 8 GHz. Verification using only a single standard offers only limited insight.

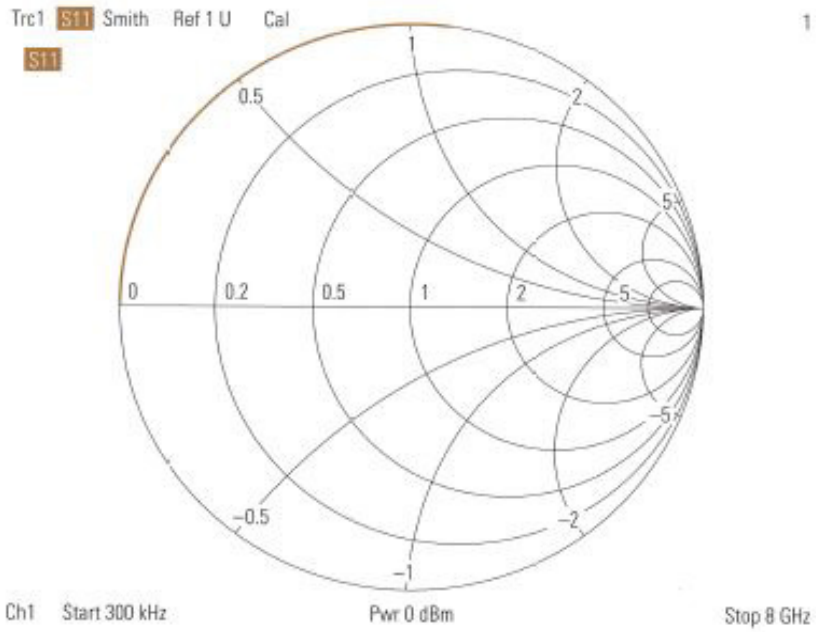


Fig. 4.1.1 Measurement of a short using the calibrated network analyzer.

Performing a TNA Calibration

The TNA calibration technique is particularly well-suited for measurements using a test fixture. The primary benefit of this calibration technique lies in the minimal requirements that are placed on the properties of the standards.

Test Setup:

- Vector network analyzer
- Two test port cables, connected to test fixture
- Through (T) standard, arranged on a substrate
- Attenuator (A) standard, arranged on a substrate
- Symmetrical network (N) standard[1]

Procedure:

1. First, a new calibration kit must be created in the network analyzer. In some cases, it is necessary to first define a connector type as well. Assign a name, e.g. "Fixture". As the connector gender, choose "sexless". You can now assign a new calibration kit to the connector type "Fixture", e.g. with the name "TNAFix".

2. Now create the TNA standards for the calibration kit as described below. Make sure that the connector type "Fixture" is selected.
   a. Define the through (T) standard by entering its exact electrical length and its losses.
   b. The symmetrical network (N) is specified as an approximate open.

Measurement Tip:

Test fixtures generally have a mechanically rugged design, but their inner conductors and the often brittle substrates do necessitate careful handling.

1) In some cases an extra substrate for the symmetrical network is not necessary, leaving the test fixture open without locking it will then serve as a symmetrical network.

c. The attenuator (A) standard does not require any specifications.

3. Set up the electrical connections between the test fixture and the network analyzer.

4. As described in section 4.1, make all of the channel settings and be sure to heed the warm-up time for the instrument.

5. Begin a two-port calibration and use the TNA calibration procedure. Select the calibration kit "TNAFix" that you created.

6. Carefully insert the attenuator (A) into the test fixture. Lock the fixture. Make sure that the inner conductor is correctly positioned. Perform the calibration measurement for the attenuator (A). Unlock the test fixture and remove the standard.

7. Carefully insert the symmetrical network (N) standard (formed in this case by the substrate without a transmission line) into the test fixture. Proceed as described under item 6.

8. Carefully insert the through (T) into the test fixture and proceed as described under item 6.

9. If a short is available, you can use it to verify the results.

Measurement of the Reflection Coefficient and the SWR

Reflection coefficient measurements are sometimes made on one-port devices. One-port DUTs only have a single pair of terminals so that a one-port calibration (OSM calibration procedure) is sufficient. When it is necessary to determine a reflection coefficient on a two-port device, twoport calibration is required. Both ports of the DUT must be connected to the network analyzer. As a general rule, an N-port DUT requires an N-port calibration.

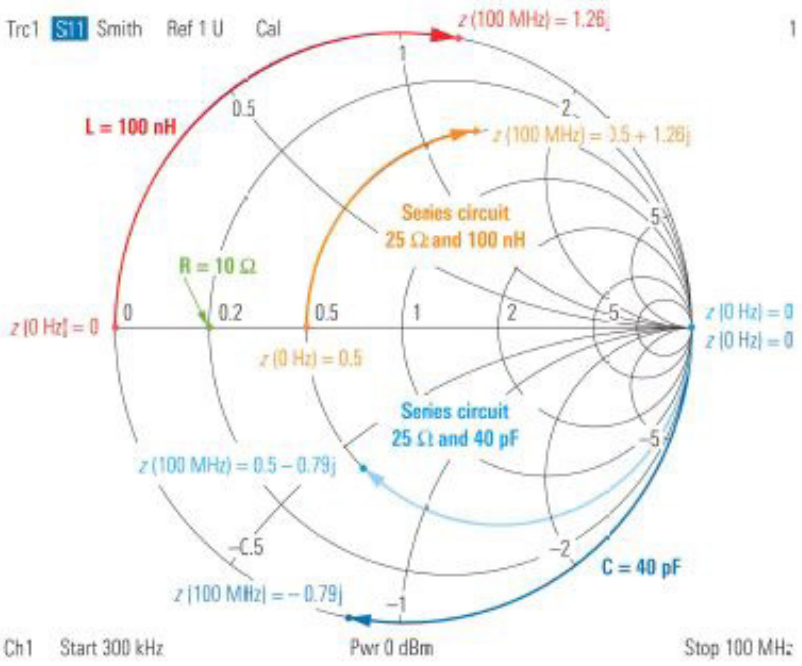


Fig. 4.3.1 Measurement of typical one-port components (reference impedance 50 Ω).

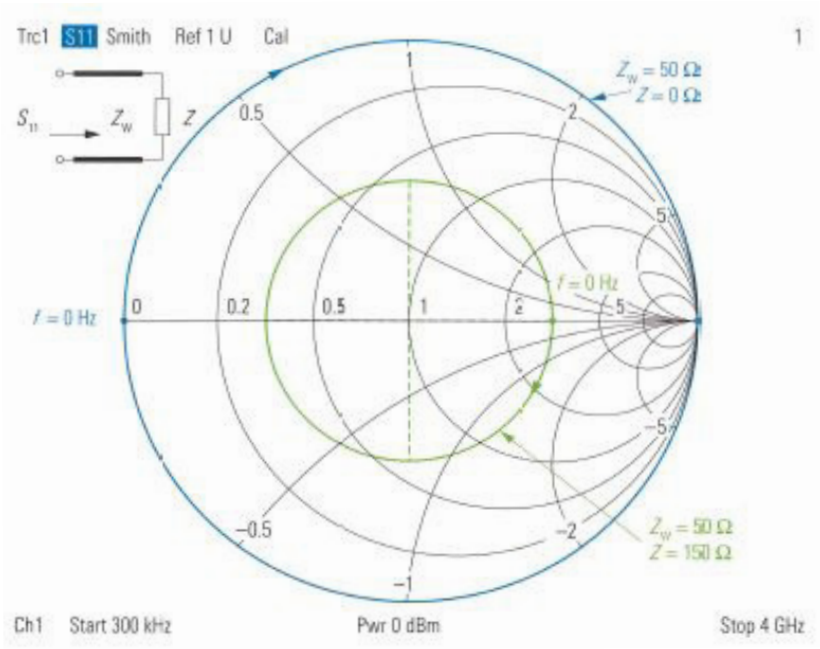


Fig.4.3.2 Measurement on different transmission line sections (reference impedance 50 Ω).

The Smith chart is most common known for designing simple matching networks. All the parameters necessary for these networks can directly be read off using graphical design techniques. A Smith chart is also suitable for identifying components. In the following two Smith charts, reflection coefficient traces for typical one-port devices are shown. These charts have been normalized to the reference impedance Z_0 = 50 Ω. The traces are generated as a result of the frequency sweep performed by the vector network analyzer. They should not be confused with the transformation paths used to determine a matching network.

A resistor with a value of 10Ω leads to the frequency-independent point z = 0.2 shown in green in Fig. 4.3.1. For f = 0 Hz, the inductance L = 100 nH behaves like a short. At increasing frequencies, the imaginary part of its impedance grows (red curve in Fig. 4.3.1). At the stop frequency of 100 MHz, it reaches the point z = 1.26j. If we connect a 25 Ω resistor in series with the inductance, we obtain the trace shown in orange in Fig. 4.3.1. At the stop frequency, the series circuit reaches the point z = 0.5 + 1.26j. The capacitor with C = 40 pF behaves like an open at f = 0 Hz. The same applies to the series circuit consisting of a capacitor and a resistor. At increasing frequencies, the imaginary part of the capacitor impedance assumes finite values. It has a negative sign. The frequency-dependent capacitor impedance (or the impedance of the series circuit consisting of a capacitor and resistor) is shown in Fig. 4.3.1 as a dark-blue (or light-blue) trace. Particularly at higher frequencies, losses as well as parasitic resonance effects in the components lead to divergence from the traces shown in Fig. 4.3.1.

Fig. 4.3.2 shows the input reflection coefficients for various transmission line configurations. A homogenous, approximately lossless transmission line with a characteristic impedance Z_c and load impedance Z is assumed. In the first case, a transmission line with Z_c= 50 Ω is terminated with a short Z = 0 Ω. The trace starts for 0 Hz at z = 0. With increasing frequency it runs along the outer circle of the Smith chart in clockwise direction (blue trace in Fig. 4.3.2). The center point of the circle lies at the normalized impedance z = Z_c /50 Ω = 1. If we terminate the same transmission line with Z = 150 Ω, the trace for 0 Hz will begin at the point z = 3. Similar to the previous case, we again note a circular curve around the point z = 1 with the same direction of rotation (green trace in Fig. 4.3.2). In case of lossy transmission lines, a spiral-shaped curve arises, and the trace tends with increasing frequency

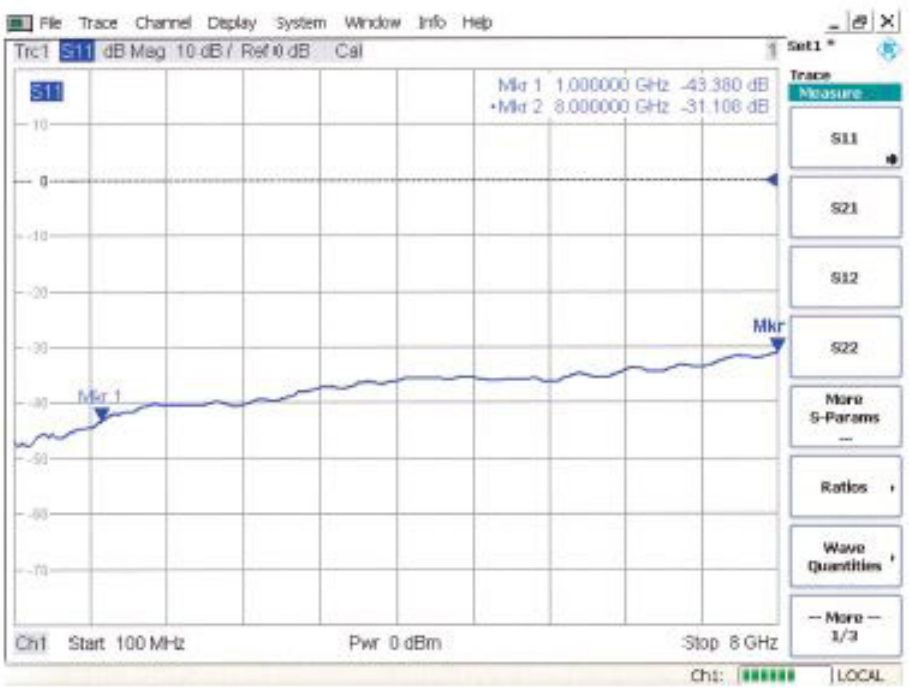


Fig. 4.3.3 Check of the match of a high-performance attenuator pad.

towards the point s_{11} = 0 (corresponding to z = 1 + 0j).

To check the match, of a high-performance attenuator pad, the magnitude of the reflection coefficient will be sufficient in most cases. Moreover, the Smith chart is not suitable for small reflection coefficients due to its linear scale. It is advantageous to format the reflection coefficient s_{11} as a decibel (dB) value and display it in a Cartesian diagram. The converted reflection coefficient is termed reflection. A typical example is the reflection a_{s11} calculated from the reflection factor s_{11}:

$a_{s11}=20lg|s_{11}|dB$ (4.3-1)

Due to the incident wave a and the reflected wave b, a superposition pattern forms along a RF transmission line. We can distinguish between three different patterns:

- If the transmission line is terminated with $|\Gamma| = 1$,

Measurement Tip:

Well matched DUTs have a reflection of about -20 dB or lower. Values starting around -30 dB are generally considered to indicate a good match. Verification of reflections below -40 dB is technically very challenging.

Measurement Tip:

Instead of an S-parameter (e.g. s_{11}), it is also possible to display the corresponding ratio (e.g. b_1/ a_1). Note, however, that network analyzers generally do not perform system error correction for ratios of wave quantities.

Resistance Z in case Z is real and a reference impedance $Z_L = 50\ \Omega$ is used	Magnitude of the reflection coefficient $\lvert s_{11} \rvert$	Reflection according to formula (4.3-1)	Standing wave ratio SWR according to formula (4.3-2)
0.0 Ω or ∞	100.000 %	0 dB	∞
8.5 Ω or 292.4 Ω	70.795 %	−3 dB	5.848
16.6 Ω or 150.5 Ω	50.119 %	−6 dB	3.010
19.1 Ω or 130.7 Ω	44.668 %	−7 dB	2.615
21.5 Ω or 116.1 Ω	39.811 %	−8 dB	2.323
23.8 Ω or 105.0 Ω	35.481 %	−9 dB	2.100
26.0 Ω or 96.2 Ω	31.623 %	−10 dB	1.925
28.0 Ω or 89.2 Ω	28.184 %	−11 dB	1.785
29.9 Ω or 83.5 Ω	25.119 %	−12 dB	1.671
31.7 Ω or 78.8 Ω	22.387 %	−13 dB	1.577
33.4 Ω or 74.9 Ω	19.953 %	−14 dB	1.499
34.9 Ω or 71.6 Ω	17.783 %	−15 dB	1.433
36.3 Ω or 68.8 Ω	15.849 %	−16 dB	1.377
37.6 Ω or 66.4 Ω	14.125 %	−17 dB	1.329
38.8 Ω or 64.4 Ω	12.589 %	−18 dB	1.288
39.9 Ω or 62.6 Ω	11.220 %	−19 dB	1.253
40.9 Ω or 61.1 Ω	10.000 %	−20 dB	1.222
41.8 Ω or 59.8 Ω	8.913 %	−21 dB	1.196
42.6 Ω or 58.6 Ω	7.943 %	−22 dB	1.173
43.4 Ω or 57.6 Ω	7.079 %	−23 dB	1.152
44.1 Ω or 56.7 Ω	6.310 %	−24 dB	1.135
44.7 Ω or 56.0 Ω	5.623 %	−25 dB	1.119
46.9 Ω or 53.3 Ω	3.162 %	−30 dB	1.065
48.3 Ω or 51.8 Ω	1.778 %	−35 dB	1.036
49.0 Ω or 51.0 Ω	1.000 %	−40 dB	1.020
50.0 Ω	0.000 %	−∞	1.000

Table 4.3.1 Conversion of characteristic one-port quantities.

standing-wave pattern will arise. No energy will be transported along the transmission line. The envelope of the oscillation has fixed nodes where $U_{min}=0$.

- If the line is terminated with $|\Gamma| = 0$, no superposition occurs since b = O. Only traveling waves are formed, there are no standing-waves present. The envelope does not have any node points. It has the same voltage $U_{min} = U_{max}$ at every location.

- During normal operation, a portion of the energy is reflected at the load. Using a voltage probe, it is possible to measure the voltages U_{min} and U_{max} on the envelope. (Figure 4.3.4)

The standing wave ratio or voltage standing wave ratio (SWR or VSWR) is the ratio of the maxima Umax to the minima U_{min}. However, measurement with a voltage

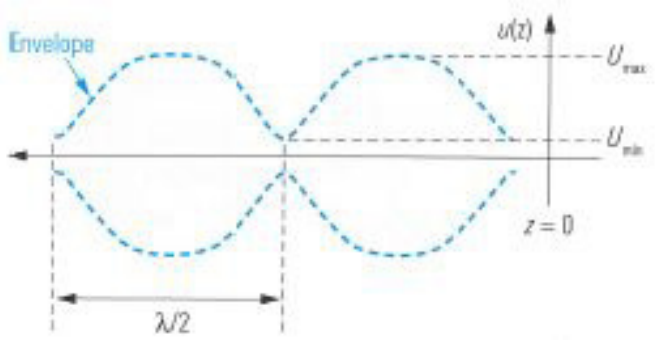


Fig. 4.3.4 Superposition of the incident and reflected waves.

probe is very time consuming and relatively inaccurate. On the other hand, a network analyzer computes the standing wave ratio, e.g. at test port 1, from the magnitude $|s_{11}|$ of the reflection coefficient.

$$SWR = U_{max} / U_{min} = 1+|s_{11}| / 1-|s_{11}| \qquad (4.3\text{-}2)$$

To display the standing wave ratio, just choose the SWR format on the network analyzer. The following conversion table shows the relationship between the four quantities Z, S_{11}, a_{s11} and *SWR*.

Measurement of the Transmission Coefficient

When measuring the transmission coefficient, it is first necessary to perform a calibration, e.g. as described in section 4.1 or 4.2. For a rough estimation of the magnitude, we can do without previous calibration. In this case the trace typically has superimposed ripple and for the test port cables, a loss is also expected (e.g, 1 dB to 3 dB for 1 m cables at a frequency of 8 GHz).

For a passive DUT, the magnitude of the transmission coefficient will have a value in the range 1 (ideal through) to 0 (no connection). Transmission coefficients are usually displayed in dB-magnitude format.

$$a_{s21}=20lg|s_{21}|dB \qquad (4.4\text{-}1)$$

The advantage of a logarithmic scale is that it offers the possibility to read off very small transmission coefficients such as $s_{21} = 10^{-8}$ corresponding to a_{s21} = -160 dB with great accuracy. At the same time, this scale allows display of large values such as s_{21} = 100 corresponding to a_{s21}= 40 dB in the same diagram. The logarithmic scale is thus better suited than a linear axis. Fig. 4.4.1 illustrates this using the example of a high pass filter. In the left part of the figure, a dB scale is used and in the right part, a linear scale. A transmission of a_{s21}= 0 dB would correspond to an ideal passband. Due to losses in the filter, Fig. 4.4.1 shows an insertion loss of 4.4 dB. In the stopband range of the filter, its ports are decoupled. There is no connection. This corresponds to a transmission a_{s21} = -∞. Effectively, the ports of the filter are not fully decoupled so that we can expect a finite value for a_{s21}. The noise floor of the receivers and the isolation of the test ports in the analyzer limit our ability to verify this value. The limit depends on the properties of the network analyzer and its settings; in Fig. 4.4.1 it is approx.-130 dB.

In some cases, it is also necessary to display the phase of the transmission coefficient. In this case, system error correction is a must. The phase values are plotted in de-

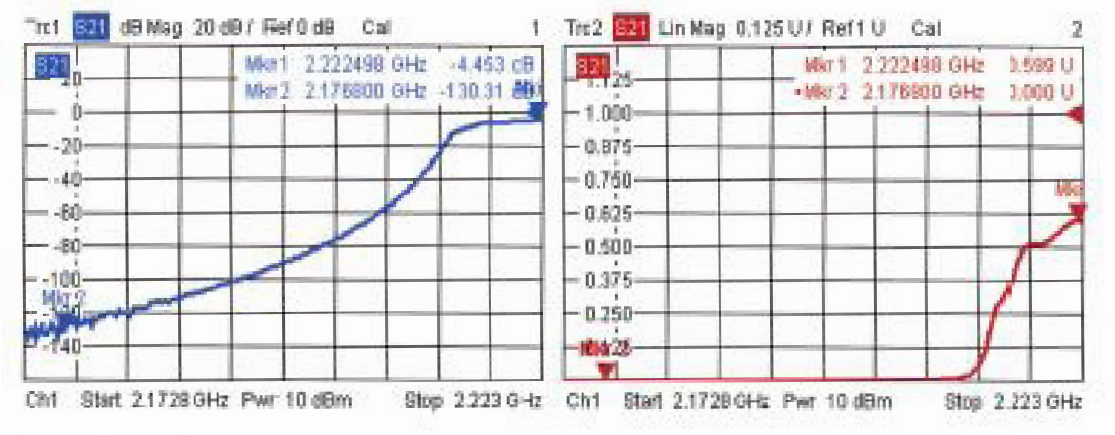


Fig. 4.4.1 Measurement of the transmission of a highpass filter.

grees vs. frequency. The phase values are normally confined to the range -180° to 180°. If you need to investigate the phase in close proximity to these limits, you can have the phase displayed continuously using unwrapped phase format.

Measurement of the Group Delay

The group delay τ_G is calculated from a transmission coefficient, e.g. s_{21}.

The group delay τ_G (f_0) corresponds to the slope of the phase arg(s_{21}(f_0)) at the frequency f_0. This slope is scaled by the factor -1/(2π) or -1/360°, yielding the physical unit of the group delay to be the second (s).

τ_G (f_0) = -1/360° d/df arg(s_{21}(f_0)) (4.5-1)

A two-port network is free of linear distortion in the frequency range f_{min} to f_{max} if it meets the following requirements in this frequency range:

Constant group delay τ_G (f) (see Fig. 5.4.1b) and

Constant magnitude of the transmission coefficient |s_{21}(f)|.

Under these circumstances, the group delay τ_G is a measure of how long it takes the modulation components of a signal with a carrier frequency $f_{min} < f < f_{max}$ propagate through the network. These components can be observed in the time-domain as the envelope curve of the modulated signal.

The curves shown in Figs.4.5.1a/b are based on algebraic expressions like $s_{21}(f)=0.5*f_{CF}^2/(f_{CF}^2 - f^2+jBf)$ which can be directly inserted in to formula (4.5.1) and the derivative (differential quotient d/df) can be done by means of

$$\tau_G(f_0) \approx -\frac{1}{360°}\frac{\arg(s_{21}(f_0 + f_d/2)) - \arg(s_{21}(f_0 - f_d/2))}{f_d}$$

(4.5-2)

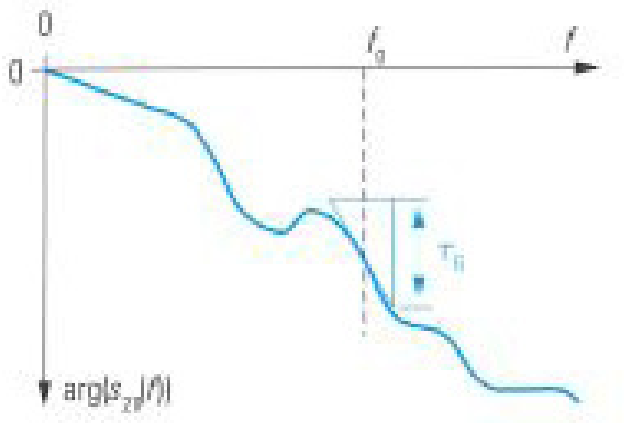


Fig.4.5.1a Definition of the group delay.

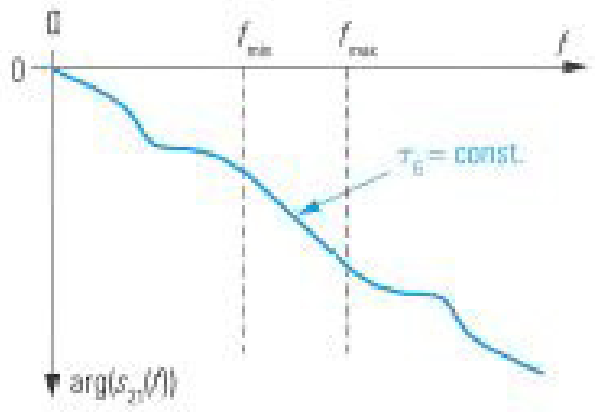


Fig.4.5.1b Interval f_{min}, f_{max} where τ_G = constant.

algebra. In contrast to this a network analyzer measures S-parameters over a discrete frequency axis. This frequency axis has a step size Δf, like the one denoted in Fig. 4.5.2. Therefore the derivative d/df must be approximated numerically by a difference quotient.

For greater flexibility it is useful to distinguish between Δf and a frequency step size f_d = κΔf that is used for calculating the difference quotient. It is known as the aperture. The factor κ is called aperture stepwidth.

In the following example, a bandpass filter was measured. Fig. 4.5.3 shows the related phase and magnitude

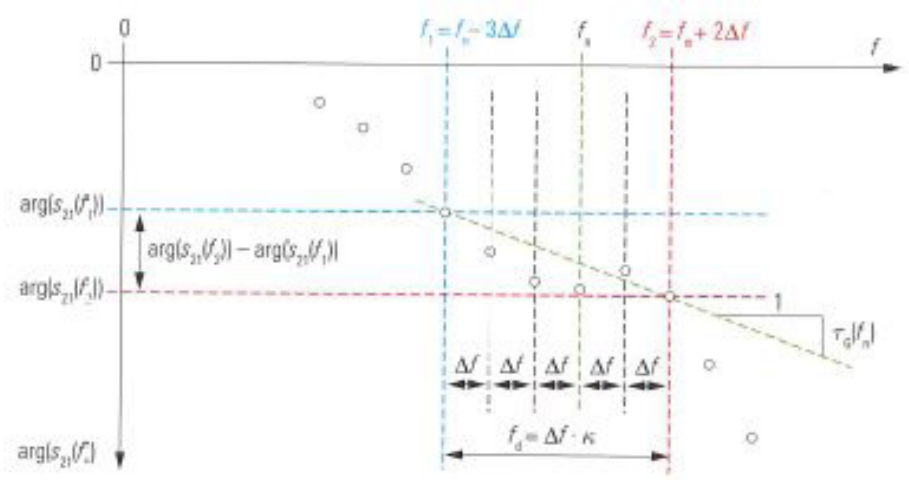


Fig. 4.5.2 Group delay on the network analyzer for aperture stepwidth K = 5.

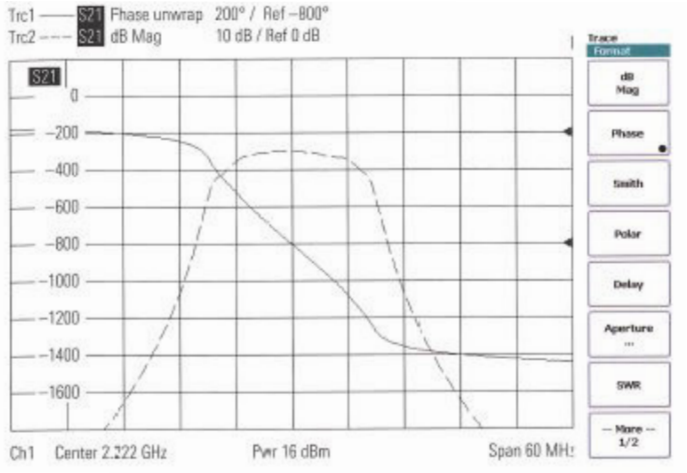


Fig. 4.5.3 Phase and magnitude curve for a bandpass filter.

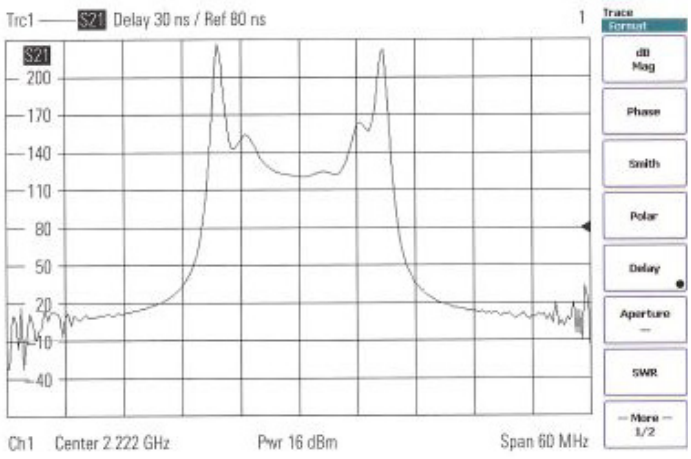


Fig. 4.5.4 Group delay for Fig. 4.5.3 with aperture stepwidth 1 (too small).

Measurement Tip:

The selected aperture stepwidth K = f_d / Δf has an influence on the calculated group delay curve τ_G (f). A value which is too large results in a loss of details, while a value that is too small will overemphasize the influence of the noise that is superimposed on the measured value. Unfortunately, there is no general rule for selecting the aperture stepwidth. The necessary value must be determined empirically.

curve for the transmission coefficient s_{21}. In Figs. 4.5.4 to 4.5.6, the aperture stepwidth K was varied. In the present example, an aperture stepwidth of K = 10 represents a good choice. Documentation of a group delay measurement should include the aperture f_d that is used. The noise that is inherent present in any measurement extends to the measured group delay. For example, by reducing the IF bandwidth, it is possible to reduce the noise superimposed on the S-parameter and thus reduce the noise on the group delay trace.

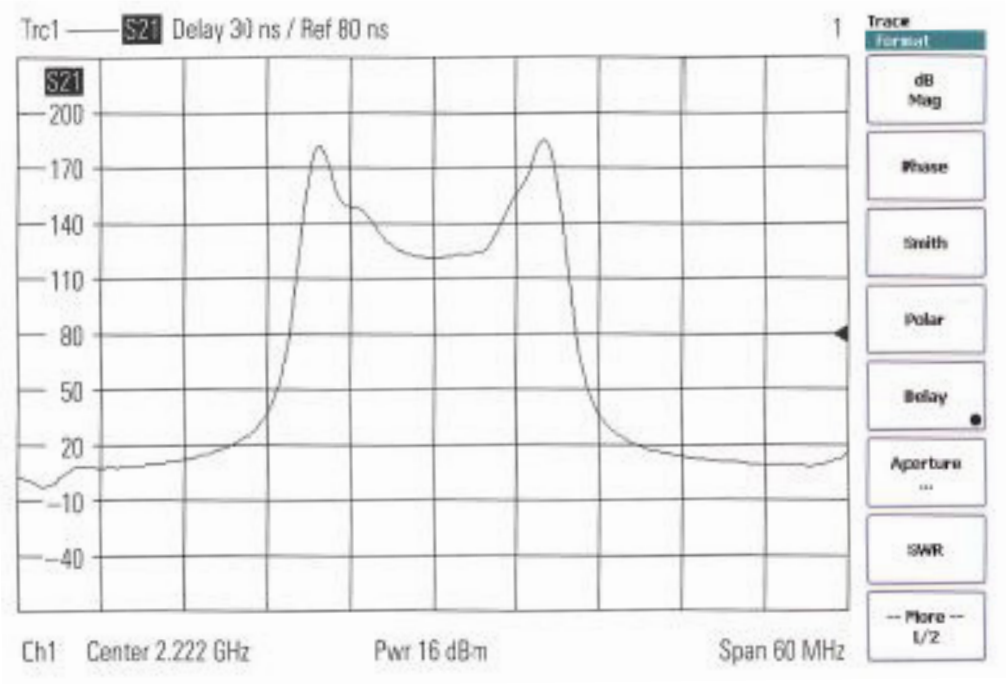


Fig.4.5.5 Group delay for Fig. 4.5.3 with aperture stepwidth 10 (ideal).

Fig.4.5.6 Group delay for Fig. 4.5.3 with aperture stepwidth 100 (too large).

Time-Domain Measurements

Time-Domain Analysis

In time-domain analysis, the measured quantity is presented as a function of time. In case of homogeneous propagation conditions, the time axis is equivalent to a distance axis. This type of analysis has significant benefits in certain applications, including typical examples such as:

- Examination of faults in transmission lines (e.g. distance-to-fault)
- RF imaging for nondestructive evaluation
- Measurement of the humidity inside of a material specimen
- Localization of cross-sectional discontinuities of a capillary
- Measurement of the triple transit signal on surface acoustic wave filters
- Separation of echo from the wanted signal in case of multi path propagation
- Moving the reference plane across unknown irregularities
- Calibration optimization using time-domain measurements

This section discusses the fundamental concepts behind time-domain analysis and shows its relation to frequency-domain analysis. Usage of oscilloscopes and network analyzers for time-domain measurements is described. Like in all previous chapters, DUTs are assumed to be linear and time invariant systems (LTI system). Because this assumption is essential for this chapter, we should give it a closer look along with some other preconditions:

A linear network can be completely described by a linear frequency domain characterization in accordance with section 1.2, formula (1.2-1) or formula (1.2-7). Note that the quantity Γ (or the S-matrix) must be independent of the stimulus power at all frequencies *f*.

In the case of time invariance, the response of the network does not depend on the particular time at which the stimulus occurs. For example, if the stimulus is delayed and occurs at the time $t = t_d$ instead of $t = 0$, then the response will only differ by a time offset td. The shape of the response will remain unchanged compared to that obtained for a stimulus at $t = 0$. Drift effects or unexpected influence quantities can disturb the time invariance of a network.

Our discussion can be limited to networks consisting of physical components.[1] These networks always fulfill two additional conditions:

- They are causal, meaning that their response never precedes their stimulus.[2]
- The time-domain signals all have real values i.e. current i(t), voltage v(t), wave quantities a(t) and b(t) and all of the time-domain quantities derived from them.

Impulse and Step Response

Networks can directly be analyzed and described in time domain. A Dirac impulse δ(t) is a typical stimulus function. It has all of its energy concentrated in an infinitely narrow time interval around $t = 0$. This energy is given by the overall integral and is equal to 1.

$$\delta(t) = \begin{cases} \infty & \text{for } t = 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{with} \int_{-\infty}^{\infty} \delta(t)\,\mathrm{d}t = 1 \qquad (5.1\text{-}1)$$

1) Typical physical components include resistors, inductors, capacitors, transformers, diodes, transistors, amplifiers and so on. On the other hand, it is also possible to model non-physical components using computer simulation techniques. One example would be a two-port network that suppresses all spectral components where $f < 0$. This network exhibits a non-causal, complex impulse response.

2) The response of stable networks consisting of physical components is always subject to (at least minimal) time delay due to the finite signal propagation speed.

The response of a linear network to a Dirac impulse δ(t) is known as the impulse response h(t). The impulse response is an important quantity for use in characterizing linear networks in the time domain.

Theoretically, it is possible to decompose any arbitrary stimulus a(t) into an infinite series of equidistant Dirac impulses that are weighted by the corresponding function values of a(t). Each Dirac impulse generates an impulse response that has the same weight and same time offset as its corresponding Dirac impulse. These similar shaped copies of the impulse response have an infinite overlap. Summing up all portions belonging to one particular time t yields the response b(t). The mathematically exact characterization of these relationships is given by the following convolution.[1]

$$b(t) = a(t) * h(t) = \int_{-\infty}^{\infty} a(\tau)h(t - \tau)d\tau \qquad (5.1\text{-}2)$$

The generation of an ideal Dirac impulse is impossible, even its approximation is problematic. It is much simpler to generate a unit step[2] σ(t). It is preferred for test setups. It has a step with a height of 1 at the time t = 0. For t > 0, its value remains at 1. The response that is generated using a unit step stimulus is known as the step response ϑ(t).

The step response ϑ(t) can be calculated by integrating the impulse response h(t) with respect to time

$$\vartheta(t) = \int_{0}^{t} h(\tau)d\tau \qquad (5.1\text{-}3)$$

Vice versa, we can calculate the impulse response h(t) by taking the derivative of the step response ϑ(t) with respect to time.

$$h(t) = d/dt\vartheta(t) \qquad (5.1\text{-}4)$$

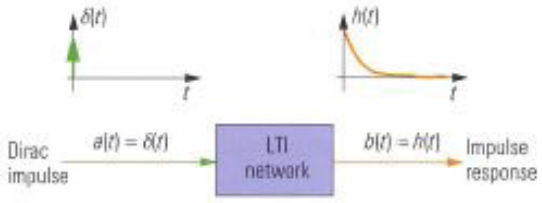


Fig. 5.1.1 Linear time-invariant network with Dirac impulse as stimulus.

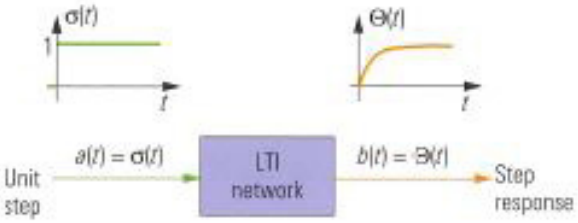


Fig. 5.1.2 Linear time-invariant network with an unit step as stimulus.

Time-Domain Analysis of Linear RF Networks

The wave quantities of a one-port device can be categorized in terms of stimulus and response. The stimulus a(t) characterizes the behavior of the incident wave vs. time. The response b(t) characterizes the behavior of the reflected wave vs. time. The following reflection quantities can be defined:

$\Gamma_h(t)$ as the impulse response from a(t) = δ(t); b(t) = $\Gamma_h(t)$ (5.1-5)

The impulse response $\Gamma_h(t)$ describes the rate of change of the impedance characteristics over time lover distance. It is particularly useful for localizing irregularities and

1) The notation b(t) = a(t) *h(t) has been introduced and should not be confused with the product a(t) * h(t)!

2) Also called Heaviside function in mathematical literature.

discontinuities along a transmission line.

Γ_ϑ (t) as the step response from a(t) = σ(t); b(t) = Γ_ϑ (t) (5.1-6)

It is recommended to use the step response Γ_ϑ (t) if the impedance characteristics of the DUT are of interest. The sign and size of the response vs. time indicate whether the DUT is resistive, inductive or capacitive.

Analysis of a network based on the quantities Γ_ϑ (t) and Γ_h (t) is known as time domain reflectometry (TDR). In theory, any arbitrary measured quantity such as the impedance Z, the admittance Y or the S-parameters can be represented in time domain as an impulse response or step response. The following discussion is limited to the reflection coefficient Γ since it is the most commonly used of these quantities.

Time Domain Measurement Example

Now, we will have a look at some typical measurements that are done using the time-domain transformation. They can be performed with any network analyzer designed to handle the time domain transformation. This feature is usually available as an option. The instrument should have an upper frequency limit of at least 4 GHz otherwise the time/distance resolution will not be sufficient for some of the examples.

Distance-to-Fault Measurement and Gating

Description

This example can be reproduced using simple equipment that should be available at any test station. The aim of the measurement is to locate an irregularity (short) in a transmission line. Building upon this, measurements are then made on a "healthy" section of the transmission line using a time gate.

Test Setup:

- Network analyzer $f_{max} \geq 4$ GHz
- Cable 1 with SMA connectors, l_1 = 48.5 cm[1]
- Cable 2 with SMA connectors, l_2 = 102 cm[1]
- Calibration kit, PC3.5 system
- SMA T-junction (see Fig. 5.2.3)
- Through (if the test ports at the analyzer are of type PC3.5)
- Adapter N to SMA (if the test ports at the analyzer are of type N).

Part 1: Determination of the Cable Properties

To perform cable measurements with the reference to a mechanical distance axis, it is necessary to know the propagation speed of electromagnetic waves in the cable. We determine this speed in a reference measurement done on a cable of the same type.

1. Make the following settings on the network analyzer:
   a. Stop frequency: f_{Stop}= 8 GHz (4 GHz if necessary)
   b. Start frequency: f_{Start} = 20 MHz (10 MHz if f_{Stop} = 4 GHz)
   c. Number of points: N = 400 points
   d. Test port output power: -10 dBm
   e. Measurement bandwidth: 1 kHz
   f. Measured quantity: s_{11}
   g. Format: Real
2. Connect the trough or the adapter (N to SMA) to test port 1. It should remain on the network analyzer during all the subsequent work steps.
3. Select the time domain transformation, type lowpass impulse (see Fig. 5.2.1).
4. All of the following measurements are one-port measurements so it is sufficient to perform a OSM calibration at test port 1.
5. Select cable 1 and measure its mechanical length l_{mech}. You should orient yourself towards the reference planes of the two connectors (see Fig. 3.2.4). Here: $l_{mech} = l_1$ = 1.02 m
6. Connect the cable to test port 1. Leave the other end of the cable open (do not install an open standard there).

1) A slightly different length is also possible. Both transmission lines should be coaxial cables and be made of the same material (e.g. RG400).

7. Configure the time axis as follows:
   a. Start time: t_{start} = -2 ns
   b. Stop time: t_{stop} = 18 ns
8. The measurement result you obtain should be similar to that shown in Fig. 5.2.1. Use a marker (e.g. automatic maximum search) to measure the delay t_p up to the first main pulse (here: t_p = 9.679 ns). Calculate the velocity factor v_p / c_0 for the current cable type.

$$\frac{v_p}{c_0} = \frac{2l_{mech}}{t_p c_0} = \frac{2 \cdot 1.02\ \mathrm{m}}{9.679 \cdot 10^{-9}\,\mathrm{s} \cdot 2.99792458 \cdot 10^{8}\,\mathrm{m/s}} = 0.703 \qquad (5.2\text{-}1)$$

9. Enter the calculated velocity factor on the network analyzer (see Fig. 5.2.2) and switch to distance display.

10. The displayed marker value (Fig. 5.2.2) should correspond to the mechanical length l_{mech} measured in step 5 of part 1.

Measurement Tip:

For a time-domain transformation in lowpass mode, a harmonic grid is required. If you have not used the settings as described at step 1 above, it is possible that the actual grid is not a harmonic one. In this case you have to modify the grid to meet the requirements of a harmonic grid. The analyzer used here would inform you about the conflict and when using the button lowpass settings offer you some possibilities to adapt the grid. However, the previous calibration that was done before modifying the grid might become invalid. For this reason, we recommend that you perform the calibration after lowpass mode has been configured (like here).

Measurement Tip:

If we want to display the trace with respect to the mechanical length, we must either enter the velocity factor v_p/c_0 or the effective relative permittivity $\varepsilon_{r,eff} = (c_0/v_p)^2$ at the network analyzer. During a reflection measurement, the signal first traverses the distance *d* from the test port to the irregularity. Next, it returns via the same path. The measured delay τ_p is thus given as $\tau_p = 2\ d/v_p$. When displaying reflection quantities with respect to a distance axis, most analyzers take into account the relationship $d = t*v_p/2$, whereas they compute $d = t*v_p$ for transmission quantities.

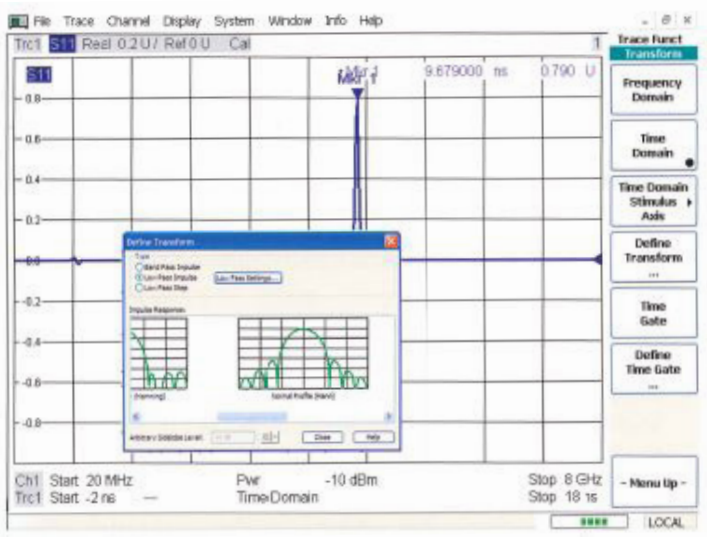


Fig. 5.2.1 Delay for an open transmission line.

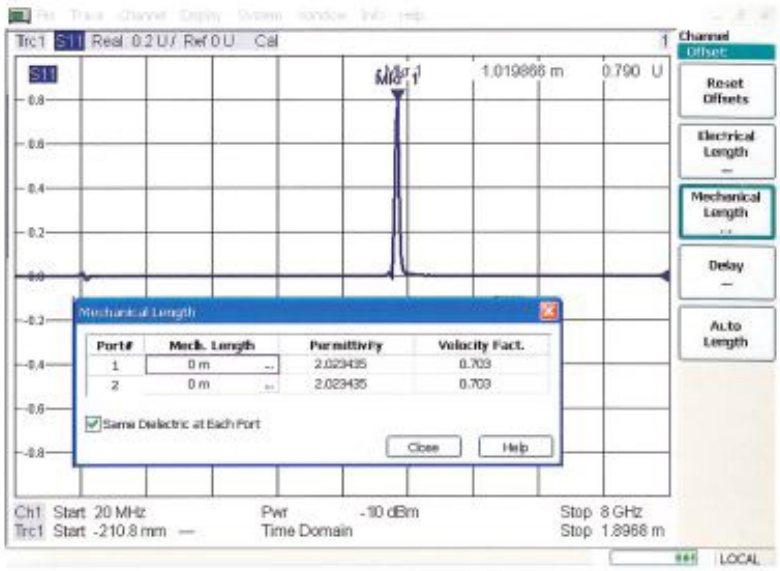


Fig.5.2.2 Verification of the length measurement.

Part 2: Locating and Masking Irregularities

A short-circuited T-junction (see Fig. 5.2.3) is used as our irregularity. It is located between two transmission lines with mechanical lengths l_1 and l_2.

Fig. 5.2.3 Test setup: Transmission line with a shortcircuited T-junction.

1. In part 1, step 1, we provided a well suited configuration for the network analyzer. Check if this configuration is also acceptable with the test setup shown in Fig. 5.2.3.
   a. Check the stop frequency
   $\Delta t = 1/(2 \cdot 8\ \text{GHz}) = 62.5$ ps, i.e. $\Delta d = v_p\, \Delta/2 =$ 0.66 cm
   $\Delta t = 1/(2 \cdot 4\ \text{GHz}) = 125$ ps, i.e. $\Delta d = v_p\, \Delta/2 =$ 1.31 cm

Measurement Tip:

The time resolution Δt for the transformation in lowpass mode is given by Δt≈1/(2fstop). Based on this relation and the velocity of propagation vp , we can calculate the distance resolution of the reflection measurement as Δd = vp Δt/2≈ vp /(4 fstop).

Distance resolution here 0.66 cn at f_{stop} = 8 GHz or 1.31 cm at f_{stop} = 4 GHz

b. Check the frequency step size

Frequency range: 0 Hz to 8 GHz

400 measurement points from 20 MHz to 8 GHz → N = 401

$T = 1/\Delta f = (N\text{-}I) / f_{stop} = 400/8\ \text{GHz} = 50$ ns

Ambiguity range: -25 ns to 25 ns or ±5.274 m at f_{stop} = 8 GHz or -50 ns to 50 ns or ±10.548 m at f_{stop} = 4 GHz.

The ambiguity range is thus greater than the total length $l_1 + l_2$ of the cables.

c. Check of the stop length 1.8968 m
The stop length is greater than the total length $l_1 + l_2$.

2. Assemble the DUT as shown in Fig. 5.2.3 and connect it to test port 1 of the network analyzer.
3. Determine the distance to the irregularity (first main pulse) and to the end of the transmission line (second main pulse); see Fig. 5.2.4.
4. Consider whether you could optimize the time domain transformation in the present case (Fig. 5.2.4) using a different window function.
   a. In the present case, no improvement is possible using a different window since neither "case a" nor "case b" applies.
5. Define a time gate that encompasses the "healthy" section of the transmission line including its open end (here: 12 ns to 14.8 ns), Select a normal gate as the shape of the time gate. For the definition of the time-domain transformation, select the rectangle window.
6. Install the match at the end of the transmission line. Change the trace setting from distance display back to time-domain display.

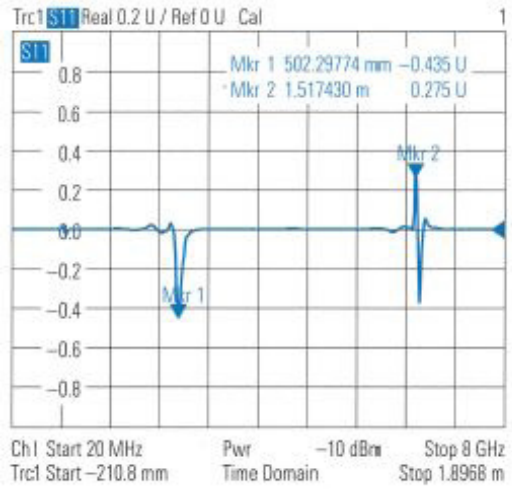


Fig. 5.2.4 Searching for the irregularity and the open end of the transformation line.

Measurement Tip:

The discrete Fourier transform used in the vector network analyzer provides unambiguous results only in the interval -T/2 to T/2 where T = 1/Δf The spectrum repeats itself periodically outside of this interval. Based on the velocity of propagation vp, we obtain the ambiguity range in terms of distances as ±T* vp /4 for reflection measurements.

Measurement Tip:

Selecting a Hann window usually represents a good compromise between the pulse width of the window and the suppression of side lobes. However, a different window can be better in the following cases:

Case a: Two pulses with very similar values are very closely spaced and cannot be distinguished due to the pulse width of the Hann window. Here, the rectangle window represents a better choice.

Case b: A second pulse is present at a larger distance from a first pulse, but the second pulse has a significantly lower level. This weak pulse can be masked by the side lobes of the Hann window, which have a minimum suppression of 32 dB. In this case it is best to switch to the Dolph-Chebyshev window with variable sidelobe suppression of, say, 80 dB.

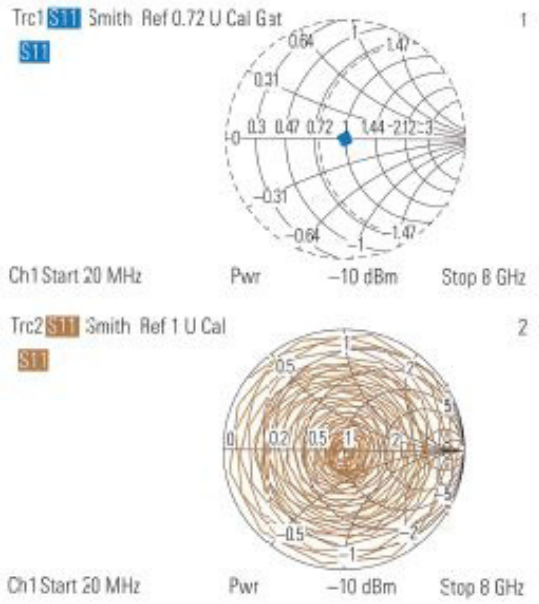


Fig. 5.2.5 Complex reflection coefficient for Fig. 5.2.3 with and without the time gate.

7. Activate the time gate and display the reflection coefficient as a function of the frequency in the Smith chart. It could be necessary to use a reference value of 0.72 to zoom into the center of the chart. Compare the result with the trace measured directly in the frequency domain. Your measurement result might look like shown in Fig. 5.2.5.

Conclusion

One of the most common measuring tasks in RF engineering involves analysis of circuits (networks) and network analysis, using a Vector Network Analyzer (VNA) is among the most essential of RF/microwave measurement approaches. Circuits that can be analyzed using network analyzers range from simple devices such as filters and amplifiers to complex modules used in communications satellites. As a measurement instrument a network analyzer is a versatile, but also one of the most complex pieces of precision test equipment and therefore great care has to be taken in the measurement setup and calibration procedures.

For more information about Vector Network Analyzers, please visit http://www.rohde-schwarz.com.

About Rohde & Schwarz

Rohde & Schwarz is an independent group of companies specializing in electronics. It is a leading supplier of solutions in the fields of test and measurement, broadcasting, radiomonitoring and radiolocation, as well as secure communications.

Established more than 75 years ago, Rohde & Schwarz has a global presence and a dedicated service network in over 70 countries. Company headquarters are in Munich, Germany.

Customer Support

- North America | 1 888 837 8772
  customer.support@rsa.rohde-schwarz.com
- Europe, Africa, Middle East | +49 89 4129 123 45
  customersupport.asia@rohde-schwarz.com
- Latin America | +1 410 910 7988
  customersupport.la@rohde-schwarz.com
- Asia/Pacific | +65 65 13 04 88
  customersupport.asia@rohde-schwarz.com

www.rohde-schwarz.us
www.rohde-schwarz-scopes.com